Survey Society of the Swedish Statistical Association

Judging the Quality of Web Panel Surveys

Methods and Metrics

November 2015

Contents

Preface

Since the beginning of this century, market and opinion research has to an increasing degree been carried out using web panels. Since web panel surveys are characterized by other possibilities and problems than traditional sample surveys, both producers and clients are confronted with difficulties when it comes to judging the quality of survey results. Traditional measures, such as confidence intervals, are no longer relevant, and there is a lack of established tools for judging quality.

In 2009, the Survey Society of the Swedish Statistical Association commissioned a task force to produce a report on quality issues in web panel surveys and how to judge their quality. This is the report of the task force. The report addresses all of the actors involved with web panels: panel providers, researchers, clients and other users of results from web panel surveys.

The members of the task force were:
Gösta Forsman, adjunct professor emeritus, Linköping University (chairman)
Karin Dahmström[1], senior lecturer, Stockholm University
Mats Nyfjäll, Fil.lic. senior consultant, Statisticon AB
Åke Wissing, senior consultant, Åke Wissing & Co.
Jan Wretman, professor emeritus, Stockholm University

In addition, Henrik Kronberg, Norstat, has served as an adviser, and Bengt Larsson from the Swedish ISO Committee has helped to coordinate definitions with those in ISO Standards 26362 and 20252.

This report is an English translation of a Swedish report (Survey Society of the Swedish Statistical Association (2014)). A comparative overview of English and Swedish terminology is given in Appendix 3.

[1] Karin Dahmström died in June 2012.

Short presentation of the report

The purpose of the report

The purpose of this report is to give recommendations for a declaration of the quality of web panel surveys. In practice it mostly concerns market and opinion surveys carried out by commercial research companies. There are already a few international standards which are generally accepted by these companies. Two important documents in this area are *ISO 26362: Access panels in market, opinion and social research – Vocabulary and service requirements* and *ESOMAR 28 questions to help buyers of online samples*; see the reference list. The present report aims at supplementing these documents with recommendations concentrating on quality judgment, specifically with Swedish conditions in mind.

From a statistical-methodological point of view there is an important difference between panels recruited by probability sampling and self recruited panels. For probability sampling there is a long established scientific theory, while with self-recruiting the scientific ground is more shaky. In reality one can find surveys of varying quality of both kinds. The present report should not be seen as a contribution to the debate for and against the use of self-recruiting. We will only discuss various types of information that should be presented. Most numerical measures presented in this report can be used irrespective of recruitment procedure, so one could say that both cases of panel recruitment will be covered.

We will suggest what kind of information should be made available when survey results are presented. Transparency is a key concept in this context. However, we will not try to give simple rules of thumb on which conclusions could be drawn from this information. Instead we hope that the information will give rise to discussion and increased awareness about the uncertainty that is always present concerning this kind of surveys.

Organization of the report

Chapters 1-5 will discuss basic concepts and terminology with web panels and web panel surveys. The problems of drawing conclusions from sample surveys, especially web panel surveys, are discussed. Chapters 6-7 are the main part of the report. There, the recommendations of the task force are presented on what should be included in a quality declaration of a web panel survey. Both verbal information and numerical measures are suggested. Finally, a short review of numerical measures is given in two appendices.

What should be included in a quality declaration?

A quality declaration should contain verbal information (see the ISO and ESOMAR documents mentioned above) as well as numerical measures of a descriptive nature.

The *verbal information* is important, but we give no exact pattern for how it should be formulated. The most important things, however, could be summarized in the following three points (for more details, see Chapter 7).

- How recruitment to the panel was carried out.
- How the sample for the specific survey was drawn from the panel.
- How weighting, if any, was done when estimates were calculated.

The *numerical measures* can be summarized in the following nine points (for definitions, see Chapter 6).

A. Measures describing participation:

- Cumulative participation rate (composed of recruitment rate, profile rate and response rate in the specific survey).

B. Measures describing the panel itself:

- Panel size at a certain date.
- Proportion newly recruited.
- Response burden.
- Attrition rate.
- Dominance.

C. Measures describing the specific survey:

- Conditioning.
- Absorption rate.
- Break-off rate.

Note that the participation rates (under A above) can only be calculated with probability sampling. The remaining measures, however, can be calculated with panels recruited by probability sampling as well as self-recruited panels.

1 Introduction

Since the beginning of this century, web panel surveys have increasingly been used by commercial companies doing market and opinion research. In official statistics and in academic research, however, these types of surveys seem not to be extensively used.

The commercial companies often emphasize that the big advantage with web panel surveys is that they are both rapid and cost effective, because the cost of panel recruitment is shared by several surveys and the respondents are prepared to participate. The fact that profile data are available for the panel members facilitates drawing selected samples from the panel, which means that selection can easily be restricted to special target groups of people which could otherwise not be identified in advance.

Statisticians with traditional academic training, however, will find problems with web panel surveys. The fact that these surveys often have a large non-response, makes it difficult to make statements about the quality of survey results using traditional probability based theory. Further complications of the same kind will arise when the panel is obtained by voluntary (non-probabilistic) enrollment.

In the present report, several descriptive measures will be suggested, and their usefulness as a basis for quality judgment will be discussed. All of these measures can be used when the panel recruitment is made using probability sampling. Many of them can also be used when panels are self-recruited. Only typically Swedish conditions will be considered, which means high internet coverage and access to the Swedish population register.

Among recent publications discussing web panels and web panel surveys we mention the following (for more references, see the reference list):

- ”ISO 26362: Access Panels in Market, Opinion and Social Research – Vocabulary and Service Requirements” (ISO (2009))
- ”AAPOR Report on Online Panels” (AAPOR (2010)).
- ”Computing Response Metrics for Online Panels” (Callegaro and DiSogra (2008)).
- “Computing Response Rates for Probability-based Web Panels” (DiSogra and Callegaro (2009)).
- ESOMAR 28 Questions to Help Buyers of Online Samples. (ESOMAR (2012))

2 Web panels and web panel surveys

2.1 Introduction

A *web panel* is a data base containing data on people who have declared that they are willing to be respondents in future surveys if they are selected. Web panels presently used by Swedish research companies vary in size from a few thousand people up to a hundred thousand or more.

The most important information on a member of the panel data base is the e-mail address, but other background information is also desirable, which may be useful for selection in future surveys. Background information of a more general kind could be, for example, age, gender, income, education, and home address. Usually more detailed and specific information on the panel member is also desired.

A panel is not static, but is changing all the time insofar as some members are leaving the panel while new members are being added. Also, the values of the background variables may change for people remaining in the panel.

The panel is constructed for the purpose of being used as a sampling frame in a number of future surveys, where data will be collected via the internet. Thus, from the panel (or from a subset of the panel) various samples of people will be drawn in future surveys. A specific survey of this kind is called a *web panel survey*. The prefix "web" in this context means that data are collected via the internet. (Other terms, used in the same meaning as "web panel" are "online panel" and "internet panel". The term "access panel" is also seen, but it is a wider concept, including panels where survey data are not necessarily collected via the internet.)

2.2 Panel recruiting

Recruiting to a panel can be done in different ways. In the following examples a few typical ways are outlined.

EXAMPLE 2.1: *Direct recruiting with probability sampling*.
A probability sample of people is drawn from a suitable sampling frame. Each selected person is asked via mail or telephone if he or she would like to become a member of a web panel. Those who accept this invitation have to register on a special web site and at the same time provide some personal data. People who register in this way will become members of the panel, provided that some eligibility requirements are fulfilled; see below. ■

EXAMPLE 2.2: *Indirect recruiting in connection with some other probability sample survey*.
In a traditional survey (using telephone interview, mail questionnaire, or face to face interview), primarily done for another purpose, the respondent is finally asked if he or she would like to become a member of a web panel. Usually, this recruitment procedure is repeated in a number of different surveys. Those who declare willingness to become panel

members will also become panel members, provided that some eligibility requirements are fulfilled; see below. This type of recruitment is often done continuously in connection with repeated omnibus surveys made by the research company. ■

EXAMPLE 2.3: *Self-recruited panel*.
People are invited (for example, via pop-ups) to declare, on their own initiative, if they would like to become members of a panel. The notification, together with some personal data, is to be done on a special web site. Those who register in this way will become panel members, provided that some eligibility requirements are fulfilled; see below. With this procedure, the panel members cannot be considered selected at random from a given well-defined sampling frame. This is an example of non-probability sampling with all the problems associated therewith. ■

EXAMPLE 2.4: *Mixed recruiting*.
Sometimes a panel may be obtained by merging several smaller panels recruited in various ways. It may also happen that a research company buys a sample, ready to use for a specific web panel survey, from another company (a "panel hotel") which has lots of different panels at their disposal. Such a ready-made sample may consist of people from different panels of varying origin and recruited in various ways. ISO 26362 requires that the research company shall disclose to the client, upon request, the types of sources used to recruit the panel, as well as the proportion of the total panel represented by each type of source. ■

When a person has declared willingness to become a panel member, the research company will usually require a *confirmation* from the candidate as a control measure. This is called a *double opt-in*. For a confirmed candidate to ultimately be accepted as a panel member, some further eligibility criteria must be fulfilled, the most important one being that the candidate has a valid e-mail address. Sometimes there may also be special requirements with respect to, for example, age, gender and education.

For people fulfilling the eligibility criteria, one usually wants some further information which may be useful for planning future samples from the panel. Such information, called *profile data*, is obtained via a special *profile survey*, where those who fulfill the eligibility criteria will have to answer a number of questions, usually via the internet. Those who answer these questions in the profile survey will finally be accepted as panel members. The access to profile data will make it possible in future surveys to restrict sample selection to special target groups.

2.3 Sampling from the panel for a specific survey

Sampling for a specific web panel survey is done using the panel as a sampling frame. It is important that the sampling procedure is carefully described. ISO 26362 declares that the methods used when sampling from the panel should be reported to the client, or be made otherwise available.

Sometimes sampling from the panel is done in such a way that all panel members (or all members of the actual target group) will have an equal probability of being selected. Sometimes varying sampling fractions are used for different subgroups of the panel, to ensure that the sample will have the desired proportions of people from different profile groups. People who have recently taken part in a survey will often be put in quarantine for some time, since it is usually preferable for panel members not to take part in surveys that are too close together.

When sampling for a specific survey, it sometimes happens that one is interested only in a special target group, for example, a special age group, or people who are using a special product. In this context the terms "selecting" and "screening" are used to denote various ways to reach the actual subgroup of the original panel. *Selecting* means that the target group can be identified and marked off in advance in the panel, using profile data, so that selection can be done directly from the target group. *Screening* is used when the target group cannot be identified in advance. Then filtering questions have to be used in the survey to find people belonging to the target group. In the first round one has to draw a large sample in order to end up with the desired number of people from the target group.

When calculating estimates from a panel survey, a weighting procedure is often used in the hope of adjusting for bias in selection. ISO 20252 (Section 4.5.1.6) says: "If a weighting process is used, it shall be appropriately described together with the weighting variables applied. The source and date of weighting target data shall be provided. The weighted and unweighted sample structures shall be documented."

3 Drawing conclusions from a web panel survey

3.1 The main problem

A web panel survey aims at describing and drawing conclusions about some target population of people, more or less clearly defined. For example, one may want to find out what percentage of the population members hold a special opinion on a market product.

The estimates are to be based on data obtained from respondents in the current survey. These respondents are selected from the web panel, and this panel is in turn recruited from a population which hopefully corresponds to some degree with the desired target population. The question then arises: *How much can we trust the survey results*? *With what degree of certainty can we draw conclusions from the survey data about the target population*?

When trying to judge the quality of the survey, there are several factors which must be taken into consideration. An important question is: what was the *sampling procedure*? Two important conceptual distinctions are probability vs. non-probability sampling and self-recruiting vs. not self-recruiting. Another important question is concerning *non-response*: How large is it, and in what way can it be expected to disturb the estimates? A third question has to do with *internet coverage*: Are there important groups of people who cannot be reached via the internet? A fourth question deals with *measurement procedure*. Do the questions catch what the client wants to know? Will the respondent understand and interpret the questions as it was intended? Are the questions designed so that the respondent will be encouraged to give truthful answers? In this report, interest is focused on sampling procedure and non-response, while problems concerning internet coverage and question wording will not be discussed.

3.2 "Representative" samples

A term sometimes met in the survey context is that a sample is "representative". A client may, for example, want a sample to be representative of a special target group. Or a research company may market itself saying that its panel is representative of the adult population of the country.

It should be remembered, however, that "representative" is a somewhat unclear concept, which can have different meanings. Those who say that a sample is "representative" probably mean that it in one way or another resembles the target population. They could, for example, mean some of the following:

- The distribution, with respect to certain background variables (gender, age, education etcetera) in the sample, is the same as the corresponding distribution in the whole population.

- All subgroups of interest in the population are represented in the sample.
- Every single member of the sample is a typical representative of the population.
- All people in the population have an equal chance of being selected.
- It is possible, with some weighting procedure, to obtain proper estimates of the population parameters of interest.

When a research company claims that their panel is "representative" of the whole country, they usually mean that the distribution of the panel members with respect to certain profile variables (age, gender, education and the like) is roughly the same as for the population in the whole country. It should be remembered, however, that this is not a guarantee that the same similarity holds also for the survey variables. Such a warning is especially justified when the panel is recruited using non-probability sampling with unknown and uncontrollable selection mechanisms.

By "representativity of a sample" we mean the extent to which a sample is "representative" of a population. Since "representative" is a term with various meanings, ISO 26362 (Section 4.4.1) states that "terms such as 'representative' shall be used only if tightly defined."

3.3 Probability and non-probability sampling

When it comes to sample surveys, a distinction is usually made between two classes of selection procedures: probability sampling and non-probability sampling. In this section we will discuss these two concepts with respect to sample surveys in general, not specifically web panel surveys. *Probability sampling* means that (1) there is a uniquely defined population from which a sample is to be drawn using a certain sampling frame, and (2) a sample is drawn from this sampling frame by means of a randomization mechanism controlled by us, so that each element in the sampling frame has a known nonzero probability of being selected. Any selection procedure that does not fulfill these requirements will be called a *non-probability sampling* procedure.

Well-known examples of probability sampling procedures are:

- *Simple random sampling*. The sample size is fixed in advance, and all possible subsets of that size from the population shall have equal probability of being selected.
- *Stratified random sampling*. The population is divided into groups, "strata", and a simple random sample is drawn independently from each stratum.

Examples of non-probability sampling procedures are:

- People can voluntarily register as respondents in a survey, *self-recruiting*.
- The interviewers have freedom to decide on their own who they shall interview. Often there may be certain restrictions implying that the sample is to contain a fixed number of people, decided in advance, from certain groups. This is called *quota sampling*.
- Experts select people they consider to be *typical representatives* of the population.

- Phone calls are made to randomly selected telephone numbers until a *predetermined number of interviews* are obtained. It is not registered how many calls are needed in order to get the desired number of respondents.
- *Snowball sampling* is sometimes used in order to get in contact with people from elusive populations. It may, for example, be expectant mothers or people using a special commodity. One starts by contacting a few people already known to belong to the relevant population. They are then asked to give information on other people who they know belong to the same population, who are then contacted. These new people are in turn asked to give the names of other people they know from the population and so on, until the desired number of respondents is obtained.

In survey practice there are usually some complications with probability sampling. In the ideal case, the sampling frame should correspond exactly to the target population of interest. In reality this is seldom the case. If people belonging to the target population are missing from the sampling frame, these people will have a zero probability of being selected, and one of the conditions for probability sampling is not satisfied. This is called *undercoverage*. If the frame contains people not belonging to the target population, there is *overcoverage*.

Another complication is that surveys are to a greater or lesser degree affected by *non-response*, that is, answers on one or more questions are missing from some people in the sample. A person for whom answers of all questions are missing is classified as *unit non-response*. A person for whom answers are obtained for at least one but not all questions is classified as *item non-response*. Non-response can occur for various reasons, but the effect is in any case that the final set of respondents on one or more questions is smaller than the original sample drawn. Since the mechanisms that determine if a person will answer or not are uncontrollable, one does not know the probability that a selected person will ultimately be a respondent or not. This means that the probability of each person in the population becoming a respondent is not known. Thus even if the original sample is drawn by probability sampling, the final set of respondents is not a probability sample in a strict sense.

The effect of non-response on the quality of the survey results depends on how much the non-respondents differ from the respondents with respect to the survey variables. If this difference is sharp, survey results may be biased. Therefore, it is important to determine, with as much certainty as possible, if the response propensity may covary with the survey variables. If the response propensity does not covary with the actual survey variables, the non-response need not necessarily cause bias, but will only have the effect that the uncertainty of estimates increases because the realized sample size is smaller than desired. If there is such covariation, however, non-response can give rise to bias in the survey results.

If non-coverage and non-response are less extensive, various ad hoc solutions are usually relied on, and the term "probability sampling" is still used. But when these sources of error are more considerable, serious investigators give up thinking of their data as obtained through probability sampling (even if the original sample was drawn with probability sampling). A problem here is that there is no generally accepted rule for how high the response rate must be for the sample to be considered as a random sample.

With non-probability sampling, it is not meaningful to talk about non-response and response rate. In this case one usually knows only how many repondents have really answered, and this number cannot be put in relation to how many would have answered under fictitious ideal conditions.

3.4 Probability sampling with web panel surveys

Can the selection procedure in web panel surveys be considered probability sampling? If we want to keep strictly to the definition, the answer must practically always be no, which will now be explained.

Sampling for a web panel survey is carried out in two phases. In the first phase a panel is obtained by sampling from the target population. In the second phase a subsample for the specific survey is drawn from the panel. Let us first look at the case where the panel is created through self-recruiting (as described in Example 2.3). Here non-probability sampling is used in the first phase, and even if the sample from the panel for the specific survey is drawn by probability sampling, the selection procedure as a whole must be described as non-probability sampling.

Now let us look at the case where probability sampling is used in both phases. In theory this is probability sampling, but in practice the non-response in both phases is usually so large that the final set of respondents cannot be considered to be obtained through probability sampling. The following fictitious (but not unrealistic) example will illustrate what we mean.

EXAMPLE 3.1: *Non-response in a web panel survey*.
Let the panel be recruited by indirect recruiting as described in Example 2.2 Thus, the respondents in an "original" survey (with probability sampling) are asked if they want to become members of a panel. The sequential loss of respondents in the different steps might be as follows.

Original survey: 50 percent answering.
Recruiting question: 50 percent of those answering want to become panel members.
Profile survey: 80 percent of those remaining are answering.
Specific web panel survey: 50 percent of those selected are answering

After the profile survey only twenty percent of the original sample remains ($0.5\times0.5\times0.8 = 0.2$). So the panel obtained can hardly be looked upon as a probability sample from the population of the original survey. Then there will be another fifty percent non-response in the specific web panel survey, which means that the respondents in this survey could to an even lesser extent be described as a probability sample from the target population. ■

There is also another possible problem with web panel surveys, namely, undercoverage. Members of the target population, who don't have access to internet, have neither any chance to become panel members, nor to be sampled for the specific survey. It is true that a large proportion of Swedish households have access to the internet, so this type of undercoverage is

not a major problem, but it may still be troublesome with some target groups, for example, elderly people.

3.5 Uncertainty of estimates

When estimates of population characteristics are based on data from a strict probability sample, one can rely on probability based statistical theory for judging uncertainty. Usually confidence intervals are calculated, which give margins of error for the estimates obtained.

With non-probability sampling, by contrast, it is not possible to calculate confidence intervals. Forming a judgment on how much one could rely on results from a survey with non-probability sampling must be done without theoretical support. Instead, one has to make a more subjective judgment, based on a description of the procedures used for sample selection and calculation of estimates, focusing on potential weaknesses.

Note that it is not claimed that surveys based on non-probability sampling should necessarily produce poorer results than probability sample surveys. What we say is only that with non-probability sampling there is no statistical theory at hand for measuring the uncertainty of the estimates. (The use of Bayesian model based quality discussion is outside the scope of this report.) However, it is still possible to discuss the quality of non-probability surveys. There is no established set of quality indicators available, but the discussion is often centered on the following three questions:

a) *Is there any reason to believe that the survey respondents are much different from the members of the target population with respect to the distribution of the study variable*? Could they be supposed to have consistently higher, or lower, values? The hope is that the sample should look like the population as much as possible in relevant respects.

b) *How are the respondents distributed with respect to certain background variables, as compared to the corresponding distribution in the target population*? It is then assumed that (1) some background variables (such as age, gender, and income) are available for the whole population, and (2) the study variable covaries with these known background variables. Large differences give rise to suspicion, while small differences do not necessarily guarantee that the estimate is good.

c) *Do we know anything about people's willingness to take part in the survey*? How many of those who were exposed to the invitation really did answer? How many phone calls were needed to get the desired number of respondents? Is there any reason to believe that the propensity to respond would covary with the study variable?

Depending on the answers to these and similar questions one could have more or less confidence in the survey results.

3.6 Judging the uncertainty of estimates from web panel surveys

In the rest of this report we will tacitly assume that both the panel itself and the sample from the panel have been obtained through probability sampling. Surveys carried out in this manner often reflect a high level of ambition from the researcher. Most of the quality indicators presented in this report are nevertheless possible to use in the case of non-probability sampling. Our task is to suggest how the quality of web panel surveys should be declared in both cases, and not to take up a definite position in the controversial question of probability vs. non-probability sampling.

Even if the intention was to obtain a panel through probability sampling, we still look at the subsequent web panel survey as based on non-probability sampling, because of the usually large number of missing answers in the whole series of activities that were necessary in order to obtain a panel, and finally a sample from the panel. We therefore recommend that research companies should not use probability based measures such as confidence intervals and the like. As a basis for quality judgment we would rather suggest answers to more practical questions on how the survey was carried out, for example, questions of the kind given in the following list.

- How was recruitment of the panel carried out?
- Which sampling frames were used?
- How many people were sampled?
- How many of those were willing to join a panel?
- What were the eligibility requirements?
- How many people remained after the eligibility check?
- How many of those remained after the profile survey?
- How many members are there in the panel at present?
- For how long have they been in the panel?
- In how many surveys have they participated in the last year?
- How many have left the panel for various reasons?
- How is replacement carried out?
- What is the policy to eliminate professionalizing and attrition?
- Can the system of incentives, if any, have affected the quality of the answers?
- How often are profile data updated?
- What is the structure of the panel with respect to profile variables such as age, gender, and income?
- Is there any reason to believe that the panel is in some respect markedly different from the target population?
- How was sampling carried out from the panel to the specific survey?
- What was the response rate in this survey?
- How were estimates calculated?
- What kind of weighting was used, if any?

These were just examples of questions to be answered in a quality declaration accompanying the results from a web panel survey. In practice, the questions would have to be further specified (which will also be done in the following chapters). Note that there are two general classes of questions, namely, questions relating to the panel per se, and questions relating to the specific survey based on that panel.

In certain cases it might be possible to evaluate a survey by comparing survey results with results from some reliable source, such as official statistics or other similar surveys of high quality. Since such possibilities are probably rare when it comes to market and opinion surveys, this type of evaluation will not be further discussed in this report.

In conclusion, what we think is needed to give an idea of the quality of the results from a web panel survey is:

- A verbal description of the procedures used in various stages of panel construction, sampling from the panel, data collection, and calculation of estimates. Special attention should be directed to potential weaknesses of these procedures.
- Numerical indicators which could help to give information on the points mentioned above.

The present report aims at giving more substance to these recommendations, especially as concerns numerical indicators.

4 What can web panels be used for?

In Chapters 2 and 3 it was seen that a scientific basis for generalizing results from web panel surveys of the general public to the whole population is often lacking. Therefore web panel surveys are in most cases less suitable if the purpose is to estimate levels in the population with high precision. Instead, web panel surveys might be considered in the following situations:

- For non-recurrent surveys where the objective is to investigate if certain properties are present in the population, or to generate ideas in general, not focusing on exact levels. There is no theoretical basis for calculating measures of uncertainty. Under this heading we find various types of qualitative surveys, and surveys where people are interviewed in the street.
- When high accuracy is desired for estimates of population levels, a web panel survey might still be used if it can be verified empirically that sampling from the current panel gives reliable estimates. This requires that you have accumulated knowledge of how useful the current panel is for measuring various survey variables. To get such knowledge it is necessary to compare results from the panel with results from traditional probability based surveys or from censuses. Even in this case it is true that numerical measures of uncertainty cannot be calculated.
- For repeated surveys where the same questionnaire and procedures are used each time, and where it is hoped (optimistically) that a systematic error, if any, will have the same size and direction each time. Interest should then be focused on trends, not levels. The assumption of "the same systematic error each time" should be confirmed in one way or another.
- Web panels might also be used in experiments where the panel, or a subgroup of the panel, is divided at random into groups which are given different "treatments". One might, for example, want to see how different versions of a questionnaire may affect the responses. The randomization will make it possible to compare the groups even if the panel itself is skewed in distribution as compared to the population.

It is assumed above that routines for panel maintenance are ambitious and survey procedures are well accomplished. To judge if this is the case one could make use of non-numerical information as well as the numerical indicators to be suggested in Chapter 6.

5 Concepts, terms and definitions

There are concepts and terms that are used in connection with web panel surveys. A few terms have been discussed in the previous chapters, and some of them will be used in Chapter 6 where numerical measures are defined. To make the presentation in Chapter 6 as concise as possible, some of the terms to be used there will now be defined once and for all. For those terms that are also used in ISO 26362, we will use the definitions given there. Terms defined in the present Chapter will be bold-faced in chapter 6.

When a person explicitly consents to becoming a panel member, this is called an **opt-in**. If the person confirms this consent later, for example, by participating in a profile survey, it is a called a **double opt-in**. The second consent is often given in connection with the profile survey. Note that there may be people in the panel that have not given member consent at all. This can happen, for example, if the panel provider has been given access to a register of customers or a register of members of an association.

For a person to become a **panel member** it is required that the person

- has been recruited from a documented source;
- has provided profile data;
- has given appropriate information for validation of identity;
- has given explicit consent to participate in surveys according to the terms and conditions of panel membership.

An **active panel member** is a panel member who

- has participated in at least one survey, if requested;
- has updated his/her profile data or has registered to join the panel within the last 12 months.

By the **size** of a panel we mean the number of active members; see Section 6.3.1.

When a person has been a panel member for some time, it may happen that his/her response behavior has changed. This can manifest itself in various ways. It could mean that the member changes his/her attitudes and values in order to be a "clever respondent", that he/she becomes more observant, giving answers that he/she would not have given otherwise, or becomes more of an expert. It could also mean that he/she becomes careless, giving answers not properly thought out. Irrespective of how it manifests itself, longtime membership in the panel may cause a panel member to behave differently from how he/she would have done if he/she had been new in the panel. This change of behavior is called **conditioning**. The special type of conditioning where the respondent becomes more of an expert is called **professionalizing**. A panel member who becomes careless is called **inattentive**. A panel member who gives deliberately untruthful answers is said to be **fraudulent**.

The composition of a panel is changing all the time, for various reasons. ISO 26362 says that "panel members shall be given a straightforward method for being removed from the access panel if they choose". The term for leaving the panel is **attrition**. There are different kinds of attrition:

- voluntary attrition, that is, the panel member leaves the panel at his/her own request;
- involuntary attrition, often due to inactivity from the panel member;
- ineligibility.

With voluntary attrition, the panel member shall, according to ISO 26362, "not be selected for future research studies within the access panel unless a new acceptance is obtained from the panel member in future recruitments or research". The most common example of involuntary attrition is when the panel member no longer fulfills the requirement for being an active panel member.

Recruitment to the panel can take place on one single occasion or, more commonly, on repeated occasions. People recruited on the same occasion are said to belong to the same **cohort** of panel members.

6 Numerical measures for judging quality of web panel surveys

6.1 Introduction and reading guidelines

In this chapter we will suggest various numerical measures to be used for judging the quality of web panel surveys. They all shed light on how the panel is maintained or how the survey has been managed in various ways. Note that for web panel surveys we cannot give confidence intervals or similar measures to indicate the uncertainty of results. We refrain from calling our measures "quality measures", since they are not supposed to measure quality in an exact way. Their main purpose is to draw attention to quality aspects of the panel and the survey. Seen together, they will be important for judging the usefulness of survey results. We also need some non-numerical (verbal) information to get a more complete picture of how much we can rely on survey results; see Chapter 7.

We recommend that research companies make these measures easily available and we also recommend that clients ask for them. It should be seen as a warning signal for possible shortcomings of the panel or survey if several of these measures are not presented.

The definitions in Chapters 6 and 7 are the same as in ISO 26362.

6.2 Cumulative participation rate and its components

Cumulative participation rate is an important measure which has similarities with *response rate* in a conventional (non-panel) sample survey with probability sampling. It can be useful for giving a hint on potential non-response error in probability based web panel surveys. The measure has three components, called *recruitment rate*, *profile rate* and *specific participation rate*. These components will be defined in Subsections 6.2.1 – 6.2.3, and a comprehensive measure, *cumulative participation rate*, will be defined in Subsection 6.2.4. It is assumed throughout that the panel is recruited via probability sampling, and that the sample for the specific survey is drawn with probability sampling from the panel.

6.2.1 Recruitment rate

Case a) Direct recruiting with probability sampling on one single occasion.

We assume that a probability sample is drawn from a target population with the sole purpose of recruiting panel members. Each person in the sample is asked if he/she would like to become a panel member. Figure 1 below will illustrate these things step by step. The following notation will be used:

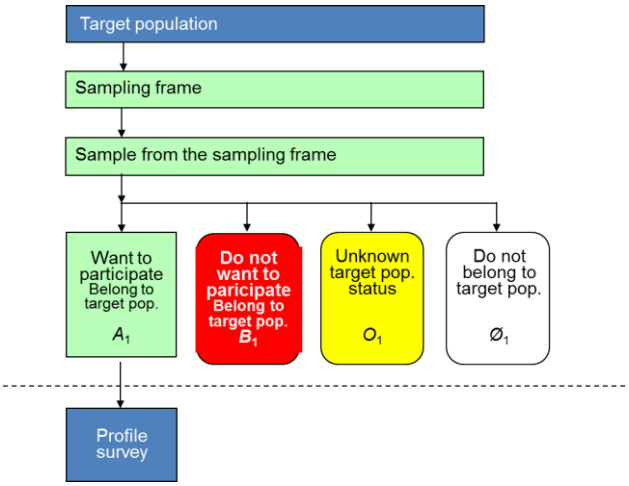
A_1 = the number of respondents who belong to the target population and want to participate in the panel.

B_1 = the number of respondents who belong to the target population but do not want to participate in the panel.

O_1 = the number of people in the sample with unknown target population status.

$\emptyset_1$ = the number of people in the sample who do not belong to the target population.

Figure 1. Direct recruiting with probability sampling on one single occasion



The *recruitment rate*, *RECR*, is defined as the proportion of people in the target population part of the sample who want to participate in the panel, formally,

$$RECR = \frac{A_1}{A_1 + B_1 + u_1 \times O_1} \qquad (1a)$$

where u_1 is an estimate of the proportion of people that belong to the target population among the O_1. (As a crude estimate one could use $u_1 = (A_1 + B_1)/(A_1 + B_1 + \emptyset_1)$, but it is also possible that information from other sources can be used.)

Those who give consent to become panel members will then go on to a profile survey.

EXAMPLE 6.1. *Recruitment rate*.
Say that a probability sample of 1 000 people is drawn from a sampling frame, and that they are asked on the phone if they want to become members of a panel. Among these 1 000 people,

A_1 = 300 (want to become panel members)
B_1 = 400 (do not want to become panel members)
O_1 = 250 (no contact, hence unknown target population status)
$\emptyset_1$ = 50 (do not belong to the target population)

If the crude value

$$u_1 = \frac{A_1 + B_1}{A_1 + B_1 + \emptyset_1} = \frac{300 + 400}{300 + 400 + 50} = 0.933$$

is used, the recruitment rate will be

$$RECR = \frac{A_1}{A_1 + B_1 + u_1 \times O_1} = \frac{300}{300 + 400 + 0.933 \times 250} = 0.321$$

that is, the recruitment rate is 32 percent. ■

Case b) Direct recruiting with probability sampling on several occasions.

In this case a sort of summary recruitment rate could be obtained by taking an average of *RECR* values from the different occasions.

Case c) Indirect recruiting with probability sampling on one single occasion.

The recruiting question here emerges as a question among other questions in another more comprehensive survey, for which the target population resembles the target population of the intended panel. We assume for simplicity that these two target populations are identical. Questions are asked which make it possible to find out if a person belongs to the target population or not. In the recruiting question, the respondent has to answer if he/she would like to become a panel member. Figure 2 below will show what happens, step by step. Analogous to *Case a* above, let

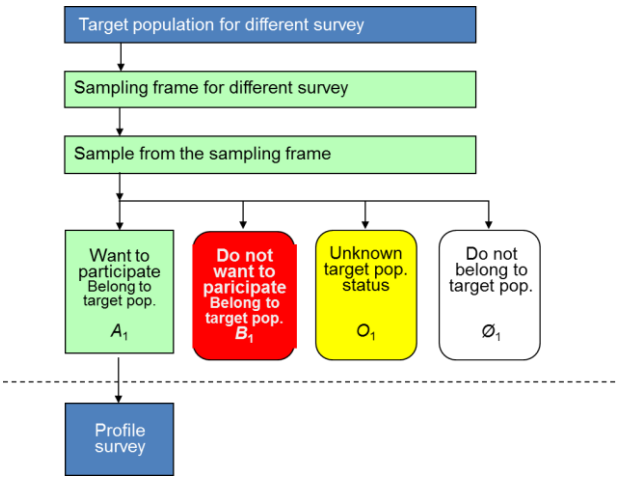
A_1 = the number of respondents who belong to the target population and want to participate in the panel.

B_1 = the number of respondents who belong to the target population but do not want to participate in the panel.

O_1 = the number of people in the sample with unknown target population status.

$Ø_1$ = the number of people in the sample who do not belong to the target population.

Figure 2. Indirect recruiting with probability sampling on one single occasion



The *recruitment rate, RECR*, is defined in the same way as in *Case a*:

$$RECR = \frac{A_1}{A_1 + B_1 + u_1 \times O_1} \qquad (1a)$$

where u_1 is an estimate of the proportion of people that belong to the target population among the O_1.

Case d) Indirect recruiting with probability sampling on several occasions.

It is unusual that indirect recruiting for a panel takes place on only one single occasion. Usually several surveys are involved. Then a sort of summary recruitment rate could be obtained as an average of *RECR* values from the different occasions.

Often, indirect recruiting is done via a series of omnibus surveys, based on probability samples, where the last question each time is the recruiting question. If substitution for non-response is used in the omnibus survey, then the substitutes should not be considered to belong to the probability sample, but be classified as non-respondents.

6.2.2 Profile rate

After the recruitment phase, a profile survey is done via the web, directed at those who have given their consent to become panel members. The purpose of the profile survey is to check the eligibility of the candidates, and to collect information that can be used for forming strata or identifying target groups in future surveys.

In the profile survey, the candidate is given a chance to confirm that he/she is willing to be a panel member, which is called **double opt-in**. It is not certain that all of those who previously agreed to become panel members will also agree a second time. The situation is shown in Figure 3 below, where

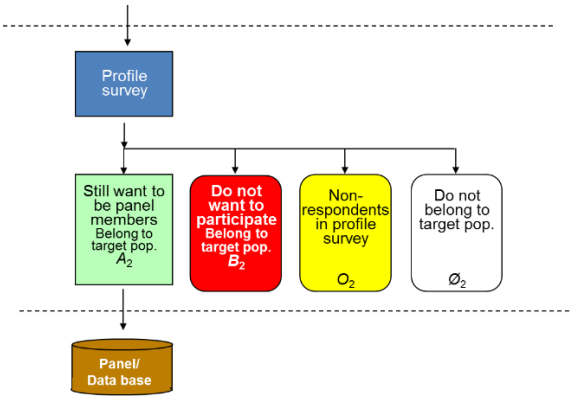
A_2 = the number of respondents in the profile survey who belong to the target population and still want to become panel members.

B_2 = the number of respondents in the profile survey who belong to the target population but no longer want to become panel members.

O_2 = the number of non-respondents in the profile survey.

$Ø_2$ = the number of respondents in the profile survey who turn out not to belong to the target population.

Figure 3. Profile survey



The *profile rate*, *PROR*, is defined, analogous to earlier measures, as

$$PROR = \frac{A_2}{A_2 + B_2 + u_2 \times O_2} \qquad (1b)$$

where u_2 is an estimate of the proportion of people that belong to the target population among the O_2.

EXAMPLE 6.2. *Profile rate.*
In Example 6.1, 300 people gave their first consent to be a panel member. Say that in the subsequent profile survey, they were distributed as follows:

A_2 = 190 (still want to become panel members)
B_2 = 20 (do no longer want to be panel members)
O_2 = 90 (non-respondents in the profile survey)
$\emptyset_2$ = 0 (identified in the profile survey as not belonging to the target population)

The profile rate, according to (1b), is then calculated as

$$PROR = \frac{A_2}{A_2 + B_2 + u_2 \times O_2} = \frac{190}{100 + 20 + 1.0 \times 90} = 0.633$$

that is, the profile rate is 63 percent. ■

6.2.3 Participation rate in a specific survey

When the profile survey is finished and the panel is established, panel members will be utilized for various specific surveys. Normally, a probability sample is drawn from the panel or from a subset of the panel, corresponding to the specific target population. This is like a conventional sample survey and the situation is shown in Figure 4 below, where

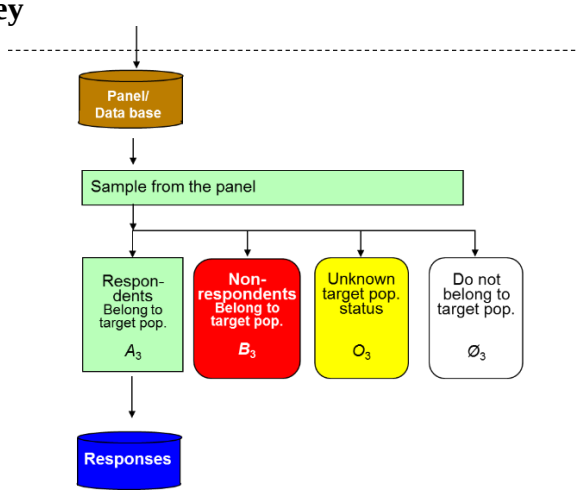
A_3 = the number of respondents in the specific survey who belong to the target population.

B_3 = the number of non-respondents in the specific survey who belong to the target population.

O_3 = the number of people in the specific survey with unknown target population status.

$Ø_3$ = the number of people in the specific survey who do not belong to the target population.

Figure 4. Specific survey



The *participation rate*, *PARR*, is defined, analogous to earlier measures, as

$$PARR = \frac{A_3}{A_3 + B_3 + u_3 \times O_3} \qquad (1c)$$

where the constant u_3 is an estimate of the proportion of people belonging to the target population among the O_3

The participation rate reflects the interest of panel members to take part in the survey, as well as the ability of the research company to encourage panel members to co-operate.

When considerable non-response occurs in a specific survey, it is often advisable to analyze it further. This can be done in various ways. A usual procedure is to compare participation rates for different groups of people. The grouping can be done using basic profile variables, such as gender, age and income. Other panel information could also be used that is connected with, for example, consumption, behavior or attitudes. Possible examples are what type of car a person owns (if any), which newspaper a person subscribes to (if any), how often a person goes to the cinema. A criterion for choosing such grouping variables is that they should covary with the response propensity in the specific survey. If response rates differ considerably between groups, this is an indication that non-response may give rise to biased survey results.

EXAMPLE 6.3 P*articipation rate.*
Say that in a specific survey with a random sample of 800 people, we find after completing data collection that

A_3 = 450 (number of respondents who belong to the target population)
B_3 = 340 (number of non-respondents who belong to the target population)
O_3 = 10 (number of people with unknown target population status)
$Ø_3$ = 0 (number of people who do not belong to the target population)

If the definition of the target population is based on profile variables only, then the panel provider knows the status of all panel members. In this case it might be that $Ø_3 = 0$.

The participation rate, as defined in (1*c*), is found to be

$$PARR = \frac{A_3}{A_3 + B_3 + u_3 \times O_3} = \frac{450}{450 + 340 + 1.0 \times 10} = 0.562$$

That is, the participation rate in the specific survey is about 56 percent. ■

6.2.4 Cumulative participation rate

Once the three measures *RECR*, *PROR* and *PARR* are defined, we define the *cumulative participation rate*, *CUMPR*, as follows

$$CUMPR = RECR \times PROR \times PARR \qquad (1d)$$

The cumulative participation rate takes all of the steps in the process into consideration from panel recruitment to the specific survey. In this respect, the measure has some similarity with response rate in a conventional (non-panel) sample survey. This similarity allows the measure (1 – *CUMPR*) to be used for a crude comparison of non-response and the ensuing potential non-response error with other surveys, for example, with data collection by phone or by mail.

Note that *CUMPR* can only be calculated when probability based procedures have been used for panel recruiting and sampling for the specific survey.

EXAMPLE 6.4 *Cumulative participation rate*.
In Examples 6.1, 6.2 and 6.3 the recruitment rate (*RECR*), the profile rate (*PROR*) and the specific participation rate (*PARR*) were calculated. Now the cumulative participation rate, defined by (1d) above, is obtained by

$$CUMPR = RECR \times PROR \times PARR = 0.321 \times 0.633 \times 0.562 = 0.114$$

That is, the cumulative participation rate is about 11 percent. Note that this measure is tied to the specific survey. ■

6.3 Descriptive measures related to the panel

In this section we present a few numerical measures which are related to the panel and its maintenance. A common thread is that none of them can be interpreted as directly measuring the quality of survey results. There is often a complex relationship between the measures and the quality of the panel, and this relationship can also vary from case to case. But the measures may hopefully give rise to discussion about quality issues. The very ability to

present these measures could also be seen as indicating a high survey culture and awareness on the part of the panel provider.

6.3.1 Panel size

The panel size is part of the basic information about the panel, not a measure of its quality. It will be seen as a component in several definitions below. The panel size should be given with reference to a certain point in time, since the number of panel members varies over time. The measure of *panel size*, *PS*, that we recommend is the following:

PS = number of **active panel members**	(2)

In the sequel, we will use the notation PS_t with a subscript "*t*" to denote the panel size at a specific point in time (often a special date in year *t*). We advise against other measures of panel size, for example, measures which include members who are non-active, or have not given an **opt-in**.

If the panel provider has several panels or a panel where the members are recruited in various ways, the measure *PS* can be given for each panel and/or for each way of recruiting. One way to do this could be to present panel size separately for the recruiting procedures discussed in Sections 2.2 and 6.2, that is, direct recruiting with probability sampling, indirect recruiting in connection with some other probability sample survey, and self-recruiting.

Note: The terms "panel member" and "active panel member" were defined in Chapter 5 and should strictly speaking be boldfaced every time they appear in the sequel. But since these terms will be used so many times we have decided, for typographical reasons, not to boldface them.

EXAMPLE 6.5 *Panel size*.
Say that the total database of the panel, on December 31 in the year *t*, contains 120 000 e-mail addresses. Of those, 100 000 are defined as active members. The panel size, as defined in (2) above, will then be PS_t = 100 000 active members. Say that on December 31 the year before, there were 90 000 active members in the panel. Then PS_{t-1} = 90 000. ■

6.3.2 New recruitment rate

The purpose is to find the proportion of active panel members who have been in the panel for one year or less. To find this, we have to know the panel size at time *t*, and also the following quantity:

ACT_t = number of active panel members at time *t*, who have been panel members for one year or less.

The *new recruitment rate*, NRR_t, is then defined as

$$NRR_t = \frac{ACT_t}{PS_t} \quad (3)$$

A value of, for example, 0.1 means that 10 percent of the active panel members were recruited during the last one-year period. The proportion of newly recruited members is not by itself an indicator of panel quality, but it may be of interest for the interpretation of other measures. When a panel is being built up, it is self-evident that the proportion of newly recruited members will be high.

A similar renewal rate can also be calculated for a specific survey, telling us how many of the respondents are newly recruited. This measurement for the survey can then be compared to the corresponding value for the whole panel.

EXAMPLE 6.6. *New recruitment rate*.
Say that on December 31, in the year *t*, the panel has PS_t = 100 000 active members. Let the number of active panel members who, at this point in time, have been in the panel for one year or less be ACT_t = 20 000. Then the new recruitment rate is

$$NRR_t = \frac{ACT_t}{PS_t} = \frac{20\,000}{100\,000} = 0.20$$

That is, the renewal rate is 20 percent. ■

6.3.3 Response burden

The purpose is to find out how many surveys panel members are engaged for during a certain period of time. We suggest that a calendar year is an appropriate period of time, and that calculations should be based on data from the preceding calendar year. But other periods of time are also possible.

Two different measures of response burden in a panel will be given. We consider the panel during a certain calendar year, say year *t*. In order to calculate the measures we need the following quantities (where the first one was already defined in Section 6.3.1):

PS_t = number of active panel members on December 31, year *t*
INV_t = number of invitations sent to the PS_t active panel members during year *t*
$COMQ_t$ = number of completed questionnaires received from the PS_t active panel members during year *t*

The following two measures, RB_1 and RB_2, of *response burden* are suggested:

$$RB_1 = \frac{INV_t}{PS_t} \qquad (4a)$$

$$RB_2 = \frac{COMQ_t}{PS_t} \qquad (4b)$$

These measures can be interpreted as follows:

RB_1 = average number of invitations per panel member during year *t* for those who were active on December 31, year *t*.

RB_2 = average number of completed questionnaires per panel member during year *t* for those who were active on December 31, year *t*.

A value of RB_1 in the interval 10 – 12 means that an average panel member has received approximately one invitation per month during the year, and a value of 50 – 55 means approximately one invitation per week. The second measure, RB_2, can be interpreted in an analogous way. (Note that the value of RB_2 must always be less than or equal to RB_1.)

The two measures RB_1 and RB_2 could be interpreted separately, but also in combination with each other. What should be considered "high" or "low" values is not obvious, but has to be discussed from case to case. A "high" value of RB_1 indicates that panel members are heavily exploited by the panel provider. If, in that case, RB_2 also takes a high value, it indicates that the panel members are willing to co-operate when they are selected. This can be seen as satisfactory, but one has to remember that if a panel member is completing many questionnaires per year, there is also a risk of **professionalizing** (see Section 6.4.1). A low value of RB_2 as compared to RB_1 may also give rise to suspicion. It indicates that there is a problem of non-response in specific surveys.

Note that the measures of response burden do not necessarily have to be calculated referring to December 31 of a certain year. The response burden could be calculated for a period of twelve months back from any other date, for example from April 30, year *t*. In the definition of RB_1, then PS_t should be the number of active panel members on April 30, year *t*, and INV_t should be the number of invitations sent to them during the period from May 1, year *t*–1 to April 30, year *t*. (In the definition of RB_2, $COMQ_t$ would have to be defined analogously.) In principle, the measures of response burden could be calculated for any day of the year. We recommend, however, that December 31 should be used. If the panel provider has a rapidly changing panel with many new recruits and/or many **attritions**, it might be appropriate for internal follow-up to calculate renewal rates on a monthly instead of an annual basis.

EXAMPLE 6.7. *Response burden*.
Say that on December 31, year *t*, the panel has PS_t = 100 000 active members. Also, say that INV_t = 900 000 invitations have been sent to these panel members during year *t*, and that $COMQ_t$ = 450 000 of the questionnaires are completely answered (by some definition given by the panel provider). The two measures (4.a) and (4.b) will then be

$$RB_1 = \frac{INV_t}{PS_t} = \frac{900\,000}{100\,000} = 9.0$$

$$RB_2 = \frac{COMQ_t}{PS_t} = \frac{450\,000}{100\,000} = 4.5$$

The average number of invitations per panel member during year *t* is 9.0, and the average number of completed questionnaires per panel member during year *t* is 4.5. These measures refer to panel members who were active on December 31, year *t*. ■

6.3.4 Attrition rate: Leaving the panel

The purpose is to find out how usual it is for panel members to leave the panel. A measure, called "**attrition** rate", will be suggested below. We will need a reference period, and our recommendation is to use the calendar year as a reference period. We first define the *average panel size* during year *t* as

$$AVPS_t = \frac{1}{2}(PS_t + PS_{t-1})$$

where

PS_t = number of active panel members December 31, year *t*

PS_{t-1} = number of active panel members December 31, year t-1

The *attrition rate*, *ATTR*, referring to active panel members, is defined as

$$ATTR = \frac{\text{Number of active members leaving the panel during year } t}{AVPS_t} \qquad (5)$$

A general problem with a high **attrition** rate is that the "representativity" of the panel composition may be called to question, if some groups of panel members are leaving the panel to a greater extent than other groups. This may also lead to an extra cost for the panel provider, in addition to already planned costs of filling up the panel. This is especially true when panel recruitment is done with probability sampling.

A high **attrition** rate could indicate that the panel members think the surveys are, for example, too comprehensive, too frequent, or have an unsatisfactory design. There may also be discontentment with the system of incentives. Many people are members of several panels, and make comparisons among these. Recent panel members and old members may be affected in different ways.

A high **attrition** rate may also cause problems in longitudinal studies, where changes over time are studied for one and the same group of panel members. If one or several **cohorts** of panel members are studied, an **attrition** rate, *ATTR*, can be calculated separately for each **cohort**.

As mentioned in Chapter 5, there are several reasons for **attrition**. If the panel provider wants a more detailed picture, **attrition** rates can be calculated for various kinds of **attrition**.

EXAMPLE 6.8. *Attrition rate*.
For the panel in Example 6.5, the average panel size is $AVPS_t = \frac{1}{2}\left(100\,000 + 90\,000\right) = 95\,000$. Say that during year *t*, 6 000 active panel members have left the panel for various reasons. The **attrition** rate, as defined by (5) above, is

$$ATTR = \frac{6\,000}{95\,000} = 0.063$$

That is, the **attrition** rate is 6.3 percent. ■

6.3.5 Dominance: A few people take part in a large number of surveys

It has been found in several studies that there may sometimes be a small group of people who take part in an unusually large number of surveys. They are said to be "overparticipating". If these special panel members are not "representative" of the target population, there is a risk that they could have an undesired effect on the survey results. This is the problem of *dominance*.

One way to tackle the problem could be to have quarantine rules, for example, that a person who has participated in a survey on a particular subject should not be selected for a similar survey until after a specified time, say three months. Other types of rules could be that a panel member may only be selected for a limited number of surveys per time period (for example, not more than one per month), or is only allowed to stay in the panel for a limited time (for example, at most two years).

We suggest a descriptive measure that can shed some light on possible dominance in a panel. Say that we are interested in the panel during the year *t*. In order to calculate the measure we need the following quantities:

$COMQ_t$ = the number of completed questionnaires during year *t* from active panel members (the same quantity as in Section 6.3.3)

CQ_{20} = the number of completed questionnaires during year *t* from the 20 percent most active panel members

By "active panel members" during year *t*, we mean those who have been active during all or part of year *t*, and who are still active on December 31, year *t*. Members who have been active during the year but have left the panel before December 31 should not be included.

To understand what is meant by the "20 percent most active panel members", one could think as follows: Let the panel members be sorted and listed by the number of questionnaires they have completed during the year, from most to least. Then the 20 percent most active panel members could be identified as the 20 percent who are first on this list.

A *dominance rate*, DOM_{20}, that can indicate if dominance is present, is defined as

$$DOM_{20} = \frac{CQ_{20}}{COMQ_t} \qquad (6)$$

This measure is interpreted as the proportion of all completed questionnaires during a year which have been delivered by the 20 percent most active panel members. The measure can be the starting-point for a discussion on whether there is a problem of dominance. There is no definite limit for when one could start talking about dominance.

Sometimes one might prefer to have this type of measure calculated on a monthly rather than annual basis. Calculations could then be made on the same principles, after appropriate redefinition of the quantities involved. One could also think of calculating a corresponding measure for the 10 or 30 percent most active panel members, or some other percentage.

Note that the dominance measure is to some degree affected by how long the panel members have been in the panel. If the renewal rate (Section 6.3.2) is high, there are fewer panel members who will be able to dominate the survey results, because a panel member who has only been in the panel for a short time has not had time to receive as many invitations as one who has been there for a long time. In order to interpret the value of the dominance measure one would need some knowledge about the percentage of recently recruited panel members. For two panels with about the same renewal rate it is easier to compare the dominance measures than for two panels with very different renewal rates. The connection between the renewal rate and the dominance rate, and how it affects the interpretation, is complicated. Explicit advice as to what is normal or what is good or bad is hard to give.

EXAMPLE 6.9. *Dominance rate*.
Let the panel size on December 31, year *t*, be PS_t = 100 000 (as in Example 6.5). Let the number of completed questionnaires from those who were active on that date be $COMQ_t$ = 450 000 (as in Example 6.7). To calculate the dominance rate (6) we need CQ_{20}, the number of completed questionnaires during the year *t* from the 20 percent most active panel members, that is, from the 20 000 most active panel members. Say that CQ_{20} = 275 000. Then the value of the dominance measure is

$$DOM_{20} = \frac{CQ_{20}}{COMQ_t} = \frac{275\,000}{450\,000} = 0.611$$

Thus, the 20 percent most active panel members account for 61 percent of all completed questionnaires during the year. ■

6.4 Descriptive measures related to a specific survey

In specific surveys, the subject field and the questions can vary a lot. Some of the measures below will be defined not for the survey as a whole, but for a single question. It is up to the panel provider, perhaps in consultation with the client, to decide which questions should be analyzed this way.

6.4.1 Conditioning: Do respondents change their behavior as they get more experienced?

When a person has been in a panel for a long time, it can sometimes happen that his/her response behavior changes in an undesirable way. The repeated exposure to various survey questions may bring about changes in the habits, attitudes and values of the panel member. A term for this phenomenon is **conditioning**. It can involve two things, namely, that the panel member (1) becomes **inattentive** or **fraudulent**, or (2) shows signs of **professionalizing**, for example, becomes more sensible and reflective when expressing opinions. For further discussion of these terms, see Chapter 5.

We will suggest three different measures of conditioning. The purpose is to find out if there are differences in response patterns between respondents who have participated in many surveys as compared to those who have participated in only a few surveys. The first two measures, $COND_1$ and $COND_2$, aim at checking for **inattentive** or **fraudulent** respondents, while the third measure, $COND_3$, is directed towards **professionalizing**. All the three measures have to be interpreted with great caution.

Note that ISO 26362 says that the panel provider shall implement procedures to identify and remove **fraudulent** and **inattentive** panel members. This calls for an analysis of the response pattern of individual panel members. The three measures to be presented here are not intended to be a basis for such decisions concerning individuals. Instead, they are aimed at finding out more generally (without pointing out individual respondents) if there are signs that indicate the possible occurrence of **inattentive**, **fraudulent** or **professionalized** respondents in the specific survey.

For each of the three measures, the respondents have to be divided into two or more groups based upon how many surveys they have participated in before the current one. For example, the grouping could be

- Group I: people who have participated in at most 5 earlier surveys
- Group II: people who have participated in 6 to 20 earlier surveys
- Group III: people who have participated in 21 or more earlier surveys

The number of "earlier surveys" is from the date when the respondent became a member of the panel. It is assumed that one has already decided on what should be an adequate definition of "participate in a survey". Note that the grouping above is only an *example* of how a grouping could be made. In the rest of this section, however, this special grouping will

be used as an illustration throughout. All measures presented below are supposed to be calculated for each of the three groups.

The first measure, $COND_1$, is to be calculated for a single question (or separately for each one of a number of single questions). Important questions are those with a relatively high non-response. For each one of the three groups, $COND_1$ is defined as

$$COND_1 = \frac{\text{number of people in the group not responding to the actual question}}{\text{total number of people in the group}} \qquad (7a)$$

A value of $COND_1$ is obtained for each of the three groups. If, for example, Group III has a considerably higher value than the other two groups, it could be seen as indicating that conditioning might be present, in this case in the form of **inattentive** sample members not exerting themselves very much when it comes to answering a question.

To calculate the second measure, $COND_2$, one has to know how long it took each one of the respondents to complete the questionnaire. For each group, $COND_2$ is defined as

$$COND_2 = \text{the median time in the group for completing the questionnaire} \qquad (7b)$$

The median time is the time it took for the person in the middle, when the respondents are sorted by how much time they needed, from the lowest to the highest. A value of $COND_2$ is obtained for each of the three groups. If, for example, Group III has a considerably lower value of $COND_2$, that is, a *lower* median value, this could possibly be interpreted as indicating a problem with **inattentive**, perhaps also **fraudulent** sample members. The interpretation is that there may be people who have participated in many surveys and have a tendency to hasten through the questionnaire in a less serious way. (If, on the contrary, this group has a *higher* median value, a conclusion might be that the people in the group with a lot of experience devote more care to complete the questionnaire than people in the other groups do.)

The two measures considered so far, $COND_1$ and $COND_2$, are meant to indicate if there are **inattentive** (possibly also **fraudulent**) people in the sample. The third measure, $COND_3$, is thought of as a starting-point for discussing the possible existence of **professionalizing**. It aims to find out if the three groups differ considerably with respect to the answers they give. The measure $COND_3$ is to be calculated for a single question (or separately for each one of a number of single questions).

To calculate $COND_3$, one important question (or a few questions) within the subject area of the survey is to be chosen, perhaps in consultation with the client. For the chosen question, a dichotomization of the response alternatives has to be made, for instance, by letting one of the *extreme* response alternatives be one category while the remaining response alternatives together are the second category. For example, if there are five response alternatives, from "very good" to "very bad", then "very good" may be one category, while the other four alternatives together are the second category. This example will be used as an illustration in

the following definition of the measure $COND_3$. The measure $COND_3$ is defined for each group as

$$COND_3 = \frac{\text{number of "very good" answers to the question in the group}}{\text{total number of people answering the question in the group}} \qquad (7c)$$

Thus, a value of $COND_3$ is obtained for each of the three groups. If there is no **professionalizing**, the proportion "very good" should be approximately the same in all three groups. If Group III differs markedly from the other two groups, this may be seen as an indication of some **professionalizing**, in the sense that people with a lot of experience of surveys have developed another attitude concerning the subject matter of the question. Of course, there are no sharp boundaries, and the measure $COND_3$, as well as the two earlier measures, has to be interpreted with great caution. It is also important that caution be exercised when the groups are small, for example, when one of the three groups is small, with less than 50 people.

EXAMPLE 6.10. *Conditioning*.
In this example we assume that there is a specific survey with a sample of 800 people selected from the panel with probability sampling, for example, simple random sampling or stratified random sampling. The sampling frame consists of the whole panel or a subset of it. Say that we finally have 450 respondents belonging to the target population (same as in Example 6.3).

A special survey question is chosen in order to examine the possible presence of **conditioning** due to extensive experience with surveys. The 450 respondents are divided into three groups based on how many surveys they have taken part in before the current one. Say that we have:

- Group I: 250 people who have participated in at most 5 earlier surveys
- Group II: 150 people who have participated in 6 – 20 earlier surveys
- Group III: 50 people who have participated in 21 or more earlier surveys

We further assume that Question No. 8 has a high item non-response. Therefore we think it is appropriate for our analysis. We examine how many members of the three groups have not answered this question. Say that we find:

- Group I: 40 people did not answer Question 8
- Group II: 27 people did not answer Question 8
- Group III: 13 people did not answer Question 8

The first **conditioning** measure, $COND_1$, is calculated for the three groups as follows

$$\text{Group I: } COND_1 = \frac{40}{250} = 0.16$$

$$\text{Group II: } COND_1 = \frac{27}{150} = 0.18$$

$$\text{Group III: } COND_1 = \frac{13}{50} = 0.26$$

that is, 16, 18, and 26 percent, respectively. We find that Group III has a higher value than Groups I and II. Such a pattern may emerge if some of those with extensive survey experience have a tendency to become careless and refrain from answering. Of course we have to investigate it further before we can say anything with certainty.

In order to calculate the second **conditioning** measure, $COND_2$, we have to know how long it took each respondent to complete the questionnaire. Then the median time is calculated for the three groups. Say that we get

Group I: $COND_2$ = 6.1 minutes
Group II: $COND_2$ = 5.8 minutes
Group III: $COND_2$ = 4.2 minutes

It seems that people in Group III (with the most survey experience) have answered the question in less time than people in the other groups. The groups are small, however.

For the third **conditioning** measure, $COND_3$, which can indicate professionalizing, we again have to choose a survey question to be studied. Say that we choose an attitude question with five response alternatives from "very good" to "very bad". We study the number of "very good" in each group and find

- Group I: 130 people answered "very good"
- Group II: 70 people answered "very good"
- Group III: 18 people answered "very good"

The third **conditioning** measure, $COND_3$, is then calculated for each group as

$$\text{Group I: } COND_3 = \frac{130}{250} = 0.520$$

$$\text{Group II: } COND_3 = \frac{70}{150} = 0.467$$

$$\text{Group III: } COND_3 = \frac{18}{50} = 0.360$$

that is, 52, 47, and 36 percent, respectively. It is seen that Group III has a lower percentage than the other two groups, which should give rise to deeper analysis to see if extensive survey experience may cause respondents to change their attitudes. ■

6.4.2 Absorption rate: E-mail deliverability

It sometimes happens that an invitation does not reach the addressee. It could be because of "bounce back", due to an incorrect e-mail address, full mail box, network error, or some other reason. If the panel provider is striving to keep the data base up to date, there should only be a few panel members with incorrect e-mail addresses.

A measure directed towards the quality of the e-mail register of panel members is the *absorption rate*, *ABSR*, which is defined as

$$ABSR = \frac{INV - NDI}{INV} \tag{8}$$

where

INV = the number of invitations sent out in a specific survey
NDI = the number of invitations that did not reach the addressee

EXAMPLE 6.11. *Absorption rate*.
Say that 800 invitations were sent out, of which 10 turned out to be bounce backs. The absorption rate (8) is then obtained as

$$ABSR = \frac{INV - NDI}{INV} = \frac{800 - 10}{800} = 0.988$$

that is, an absorption rate of 99 percent. ■

6.4.3 Break-off rate: Opened but uncompleted questionnaires

When a panel member gets an invitation to take part in a survey, he/she can open the questionnaire, or refrain from doing so. Once the questionnaire is opened, the respondent can complete it, or leave it uncompleted. By "uncompleted" we mean that either no questions at all, or only a few questions are answered. An exact specification of what should be classified as a "completed" or an "uncompleted" questionnaire has to be decided on in every single survey, and it could vary from one survey to another. When we talk about "break-off", we mean that the questionnaire is opened but uncompleted.

With a given specification of break-off, the *break-off rate*, *BOR*, is defined as

$$BOR = \frac{ONQ}{ONQ + OCQ} \tag{9}$$

where

ONQ = the number of break-offs, that is, number of opened but uncompleted questionnaires
OCQ = the number of opened and completed questionnaires.

A questionnaire classified as break-off should not be used in any calculation of survey results. The break-off rate may indicate that there are problems with the questionnaire design (too long, or too boring), or it may indicate that there are technical problems that cause a person to stop answering.

EXAMPLE 6.12. *Break-off rate.*
As in Example 6.11, 800 invitations are sent out. 450 completed questionnaires are received, and there are 340 non-respondents (as well as 10 undeliverable). Of the non-respondents, 40 people had opened the questionnaire but not completed it. They could have answered just a few questions, or no questions at all. The break-off rate (9) is then obtained as

$$BOR = \frac{ONQ}{ONQ + OCQ} = \frac{40}{40 + 450} = 0.082$$

That is, the break-off rate is 8 percent. ■

7 Non-numerical information for judging the quality of web panel surveys

In the previous chapter various numerical measures were presented, which were considered useful for judging the quality of a web panel survey. In the present chapter we will briefly describe various kinds of non-numerical (verbal) information to supplement the numerical measures in Chapter 6 in order to give an idea of the quality. The presentation is in the form of a list of items that have to do with the panel itself, its maintenance, and the survey based on that panel. It should not be seen as a definitive list, given once and for all, but rather as an attempt to demonstrate what kind of information is useful in order to judge the quality of a web panel survey. The capability to give such information can be seen as reflecting the level of ambition of the survey company. The information should not be too detailed or technically advanced, but brief and concise, with information that is of importance for the client.

7.1 Information on the panel and its maintenance

It is important to know how the panel was obtained.

- Was recruitment to the panel based on probability sampling, or was it a non-probability procedure, for example, self-recruiting?
- Is it a mixed panel, made up of several parts obtained in different ways? In that case, what proportion of the mixed panel does each part stand for?
- If it is unknown how the panel was obtained, say so!
- If the panel was obtained using some form of probability sampling, describe it briefly (for example, population, sampling frame, coverage problems, sampling procedure).
- What target population is the panel meant to represent? Representative in what sense? Does one know anything about how successful one was in attempting to get a representative panel? Of special interest is the kind of representativity that is related to important survey variables.
- If the panel recruiting was not probability based, describe briefly how it was done. Is it possible in that case to say anything about the representativity of the panel?

Information about the panel maintenance is also of interest.

- How is replacement carried out? Will claims of representativity still be fulfilled?
- How are the profile data of panel members updated? How often is it done? Are external data sources used in addition to direct questions to the panel members?
- What are the quarantine rules? Is a panel member excluded from being selected for some period of time after participating in a survey? Are there limits on the number of surveys a panel member can participate in during a certain period of time?
- What are the rules concerning exclusion from the panel? Can a panel member be excluded as a consequence of inactivity? For how long is a panel member allowed to stay in the panel before he/she is automatically excluded?

7.2 Information on the specific web panel survey

We assume that probability sampling has been used. It is then of interest to have more detailed information on how sampling from the panel was done for the specific survey.

- From which subset of the panel was the sample drawn? Was the selection restricted to special parts of the panel, for example, special age groups, geographical regions, or other groups defined in terms of profile data?
- Which sampling method was used? Simple random sampling from the panel? Or stratified random sampling (with the panel divided into strata, and simple random sampling within every stratum)? In the case of stratified sampling, how were the strata defined? Was some other sampling design used?
- If some panel members were in quarantine when the sample was selected, can this fact have affected the representativity of the sample?
- What was the sample size? In the case of stratified sampling, what was the sample size from each stratum? Was the total sample size fixed in advance (which means that one did not know in advance how many respondents one would finally get)? Or was it the number of respondents that was fixed in advance (so that one had to go on with sampling until the desired number of respondents was obtained)?
- Which measures were taken in order to reduce non-response? Were reminders sent to those who did not answer? Were there any form of incentives to encourage people to answer?
- How long time was the survey open for the respondents to answer the questionnaire?

It is also of interest to know how estimates were calculated.

- Was some type of weighting used when estimates were calculated? If this was the case, describe the weighting procedure. What was the purpose of the weighting? Is it possible to give a formal expression to describe how the estimates were calculated?

Finally we point out that the quality of survey results also depends to a high degree on how the survey questions were designed. This is true for all kinds of surveys, and not specific to web panel surveys, so it is not discussed in this report.

References

AAPOR (2010). AAPOR Report on Online Panels. *Public Opinion Quarterly*, 74(4): 711-781.

Callegaro, M., and DiSogra, C. (2008). Computing Response Metrics for Online Panels. *Public Opinion Quarterly*, 72(5): 1008-1032.

DiSogra, C., and Callegaro, M. (2009). Computing Response Rates for Probability-based Web Panels. *Section on Survey Research Methods – JSM 2009*.

ESOMAR (2012). 28 Questions to Help Buyers of Online Samples. www.esomar.org

ISO (2006). ISO 20252:2006. Market, Opinion, and Social Research – Vocabulary and Service Requirements.

ISO (2009). ISO 26362:2009. Access Panels in Market, Opinion, and Social Research – Vocabulary and Service Requirements.

Survey Society of the Swedish Statistical Association (2005). Standard för bortfallsberäkning. (Standards for Calculation of Nonresponse Rate).

Survey Society of the Swedish Statistical Association (2014). Kvalitet i webbpanelundersökningar – Metoder och mått. (in Swedish)

Literature not mentioned in the text

AAPOR (2012). Understanding a "Credibility Interval" and How It Differs from the "Margin of Sampling Error" in a Public Opinion Poll. http://aapor.org/Understanding_a_credibility_interval_and_how_it_differs_from_the_margin_of_sampling_error_in_a_publi.htm#.UwomQdGYaUk

AAPOR (2013). Report of the AAPOR Task Force on Non-probability Sampling. www.aapor.org

Baker, R., Brick, J.M., Bates, N., Battaglia, M., Couper, M.P., Dever, J.A., Gile, K.J., and Tourangeau, R. (2013). Summary Report of the AAPOR Task Force on Non-probability Sampling. *Journal of Survey Statistics and Methodology*, 1(2): 90-105.

Bethlehem, J. (2008). Can We Make Official Statistics with Self-selection Web surveys? Proceedings of Statistics Canada Symposium 2008. Data Collection: Challenges, Achievements, and New Directions.

Bethlehem, J. and Cobben, F. (2013). Web Panels for Official Statistics? Invited paper, 59th ISI World Statistics Congress, Hong Kong, August, 2013.

Couper, M.P. (2001). The Promises and Perils of Web Surveys. In Westlake, A. (Ed.) *The Challenge of the Internet*. Association for Survey Computing, London. 35-56.

Couper. M.P., Kapteyn, A., Schonlau, M., and Winter, J. (2007). Noncoverage and Nonresponse in an Internet Survey. *Social Science Research*, 36:131-148.

ESOMAR/WAPOR (2009). Guide to Opinion Polls and Published Surveys. www.esomar.org

ESOMAR (2011). Guideline for Online Research. www.esomar.org

Grey Matter Research and Consulting (2009). Dirty Little Secrets of Online Panels. www.greymatterresearch.com

Grey Matter Research and Consulting (2012). More Dirty Little Secrets of Online Panel Research. www.greymatterresearch.com

ICC/ESOMAR (2007). International Code on Market and Social Research. www.esomar.org

Public Works and Government Services Canada (2008). The Advisory Panel on Online Public Opinion Survey Quality.

Public Works and Government Services Canada (2009). Standards for the Conduct of Government of Canada Public Opinion Research – Online Surveys.

Schonlau, M., Fricker, R., and Elliott, M. (2002). *Conducting Research Surveys via E-mail and the Web*. RAND, Santa Monica, CA.

Svensson, J. (2013). Web Panel Surveys – Can They Be Designed and Used in a Scientifically Sound Way? Invited paper, 59th ISI World Statistics Congress, Hong Kong, August, 2013.

Vonk, T., van Ossenbruggen, R., and Willems, P. (2006). "The Effects of Panel Recruitment and Management on Research Results." ESOMAR World Research Paper.

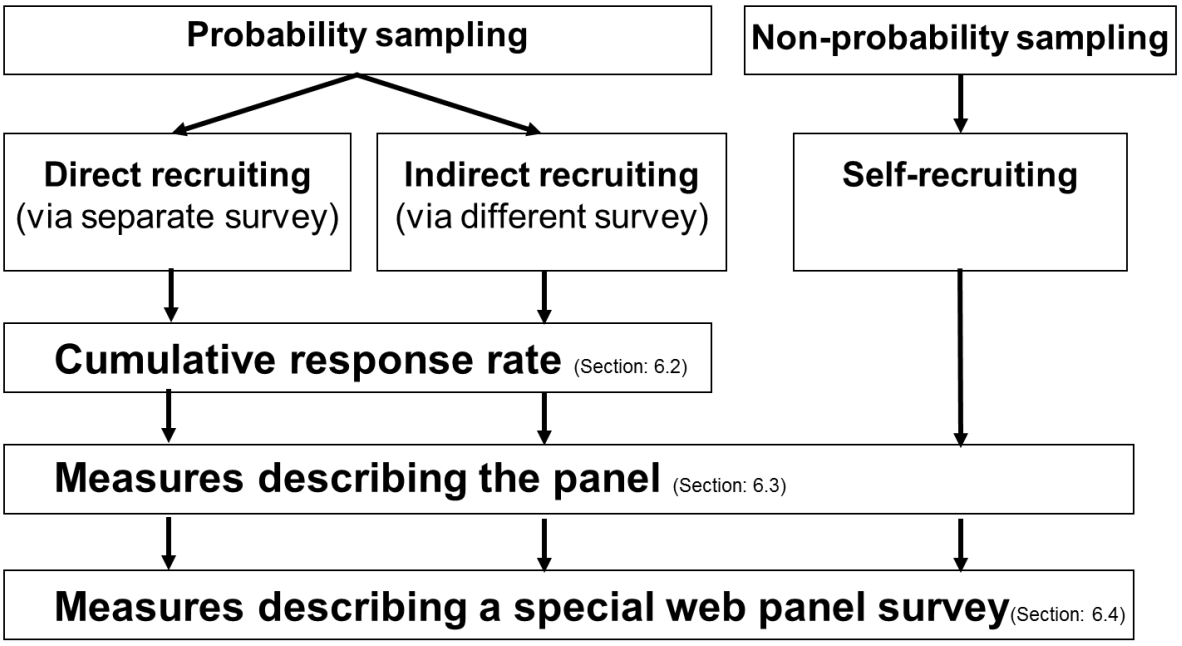
Yeager, D.S., Krosnick, J.A., Chang, L.C., Javitz, H.S., Levendusky, M.S., Simpser, A., and Wang, R. (2011). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-probability Samples. *Public Opinion Quarterly* 75:709-747.

Appendix 1: Summary of numerical measures

Category	Measure	Name	Explanation	Eq. No.
Measures describing participation	*RECR*	Recruitment rate (Section 6.2.1)	Proportion in recruitment survey willing to become panel members	(1a)
	PROR	Profile rate (Section 6.2.2)	Proportion in profile survey willing to become panel members	(1b)
	PARR	Participation rate (Section 6.2.3)	Participation rate in a specific survey	(1c)
	CUMPR	Cumulative participation rate (Section 6.2.4)	*CUMPR* = = *RECR*× *PROR*× *PARR*	(1d)
Measures describing the panel	*PS*	Panel size (Section 6.3.1)	# active panel members	(2)
	NRR	New recruitment rate (Section 6.3.2)	proportion active panel members with one year or less in the panel	(3)
	RB_1	Response burden (1) (Section 6.3.3)	# invitations per year for an average panel member	(4a)
	RB_2	Response burden (2) (Section 6.3.3)	# completed questionnaires per year for an average panel member	(4b)
	ATTR	Attrition rate (Section 6.3.4)	Proportion of panel members leaving the panel during a year	(5)
	DOM_{20}	Dominance (Section 6.3.5)	Proportion completed questionnaires per year from the 20% most active panel members	(6)

Measures describing the specific survey	$COND_1$	Conditioning (1) (Section 6.4.1)	Proportion non-respondents on a question in groups with extensive versus little survey experience	(7a)
	$COND_2$	Conditioning (2) (Section 6.4.1)	Median time to complete questionnaire in groups with extensive versus little survey experience	(7b)
	$COND_3$	Conditioning (3) (Section 6.4.1)	Proportion "extreme" answers on a question in groups with extensive versus little survey experience	(7c)
	ABSR	Absorption rate (Section 6.4.2)	Proportion of invitations to a survey that reached the addressee	(8)
	BOR	Break-off rate (Section 6.4.3)	Proportion opened but not completed questionnaires in a survey	(9)

Appendix 2: Numerical measures in relation to recruitment procedure



Appendix 3: English terms and notation with Swedish counterparts

English		Swedish	
Recruitment rate	*RECR*	Rekryteringsandel	*RA*
Profile rate	*PROR*	Profilandel	*PA*
Participation rate	*PARR*	Deltagarandel	*DA*
Cumulative participation rate	*CUMPR*	Kumulativ deltagarandel	*KA*
Panel size	*PS*	Panelstorlek	*PS*
# active panel members ≤ 1 year	*ACT*	# aktiva panelmedl. ≤ 1år	*V*
New recruitment rate	*NRR*	Andel nyrekryterade	*AN*
# invitations	*INV*	# inbjudningar	*I*
# completed questionnaires	*COMQ*	# komplett besv. enkäter	*K*
Response burden	*RB*	Uppgiftslämnarbörda	*UB*
Average panel size	*AVPS*	Genomsn. panelstorlek	$\overline{PS}$
Attrition rate	*ATTR*	Utträdesandel	*UA*
# compl. quest´s from 20% most active	CQ_{20}	# besv. enk. från 20% mest akt.	B_{20}
Dominance	DOM_{20}	Dominans	D_{20}
Conditioning	*COND*	Anpassat svarsbeteende	*AS*
# invitations in spec. survey	*INV*	# inbudn. i spec. unders.	*INBJ*
# non-deliverable invitations	*NDI*	# olevererbara inbjudn.	*OL*
Absorption rate	*ABSR*	Nåbarhetsandel	*NA*
# opened, not compl. questionnaires	*ONQ*	# avbrutna enkäter	*EjSV*
# opened and compl. questionnairess	*OCQ*	# besvarade enkäter	*SV*
Break-off rate	*BOR*	Avbrottsandel	*AA*