A standard with quality indicators for web panel surveys: a Swedish example

Mats Nyfjäll
Statisticon AB, Uppsala, Sweden mats.nyfjall@statisticon.se

Abstract

During the last 10 to 15 years, web panels have been established as a cost effective and quick method in market surveys. The generalizability of the results, i.e. the inference, to a certain population depends primarily on two important aspects: the recruitment method and the total non-response in both the recruitment phase and the specific survey. Besides those major aspects there are several other quality aspects such as the design of the specific survey, management of the panel, rules of quarantine, attrition, conditioning, etc. Traditional measures of quality in surveys, such as precision (e.g. confidence intervals), are not applicable for web panel surveys. This means that the method constitutes almost a paradigm shift since quality aspects have to be discussed in new terms. The Swedish Survey Society has assigned a committee to propose (metric) measures that can be used to assess quality aspects of results from web panel surveys. The purpose is to establish a standard in Sweden for quality indicators that can be used and how they can be calculated. This paper will present the proposed standard as well as reactions from the market research community in Sweden.

Key Words: generalizability, recruitment, paradigm shift, metric standard

1. Introduction

Since the beginning of year 2000, approximately, web panel surveys have become more and more usual in market-, opinion- and social surveys. Within official statistics, often produced by National Statistical Institutes (NSI:s), and academic research web panel surveys seem not to be frequently used.

An advantage with web panel surveys, often claimed by the web panel provider, is that they are quick and cost effective. This is due to the fact that the cost for panel recruitment is shared by several surveys and that the panelists are willing to participate. Through access to detailed background variables regarding each panelist samples can be effectively stratified and rare populations identified.

However, for a statistician there are several problems with web panel surveys. Due to the usually large non-response rate it is difficult to assess the quality of results for web panel surveys based on traditional statistical theory. Another complicating circumstance is that panel recruitment sometimes is based on self-selection instead of probability based methods.

There are some statistical literature and standard documents trying to address web panel surveys from a statistical point of view. The ISO standard 26362 *Access panels in market, opinion and social research – Vocabulary and service requirements* is an important document in this direction. *ESOMAR 28 questions to help buyers of online samples* (2012) and the AAPOR *Report on Online Panels* (2010) are also two important contributions. A common trait in these documents is that they describe and discuss procedures, processes and requirements that need to be fulfilled. However, they do not explicitly specify numerical measures – metrics – for web panel survey. Two important articles presenting metrics for web panels are Callegaro and DiSogra (2008) and DiSogra and Callegaro (2009). Despite these articles "there are no widely-accepted definitions of outcomes and methods for calculation of rates similar to AAPOR's *Standard Definitions* (2009) that allow us to judge the quality of results

from surveys using online panels", quote from AAPOR (2010).

In Sweden, an attempt to form a standard of metrics for web panel surveys has been undertaken. In 2009 the Swedish Survey Society formed a committee whose purpose was to propose metric measures that can be used to assess quality aspects of results from web panel surveys. Before the standard is described some short historical notes regarding the situation in Sweden may be warranted.

In 2005 the Swedish Survey Society (2005) published *Standard for Response Rate Calculation* (manuscript exists in Swedish only). The standard was adapted for Swedish conditions and used the AAPOR (2004) *Standard Definitions: Final Dispositions of Case Codes and Outcomes for Surveys* as an important source of inspiration. Even though AAPOR (2004, 2011) are excellent documents they cannot constitute a standard in Sweden due to several reasons. One obstacle is the language barrier; these kinds of documents will simply not be read by the broad survey industry in Sweden. That was one of the major reasons for writing a Swedish standard. The standard for response rate calculation has since then been broadly accepted as a standard in Sweden. Both by the market research industry as well as Statistics Sweden (the Swedish NSI). The standard is also used as literature in courses in survey methodology within the universities.

It was with this background that the Swedish Survey Society formed the web panel committee in 2009. One humble expectation was that a standard with metric measures for web panel surveys could gain some, or perhaps, the same acceptance. The committee consisted of seven statisticians from different fields; the academy, consulting firms, web panel providers and the Swedish ISO committee. Writing this kind of document requires a balancing act. The document has to present the material in a strict statistical matter, but if the presentation is too technical it will be difficult to penetrate and gain broad acceptance. The Swedish Survey Society has no formal jurisdiction to implement the proposed metric measures as a standard. One has to remember that one important aspect whether a document of this kind will be accepted as a standard or not depends on the easiness of the material as well as to whether the metric measures make any sense and are reasonably easy to calculate.

At the end of 2012 a remittance version of the standard was presented. About 15 companies and individuals were asked to comment on the proposed standard. Based on their comments a final version of the standard will be presented sometime in the summer of 2013. Below we present the metric measures proposed in the standard. The measures can be categorized into three groups; measures regarding participation rates, measures regarding the panel and measures regarding a specific survey.

2. The Standard

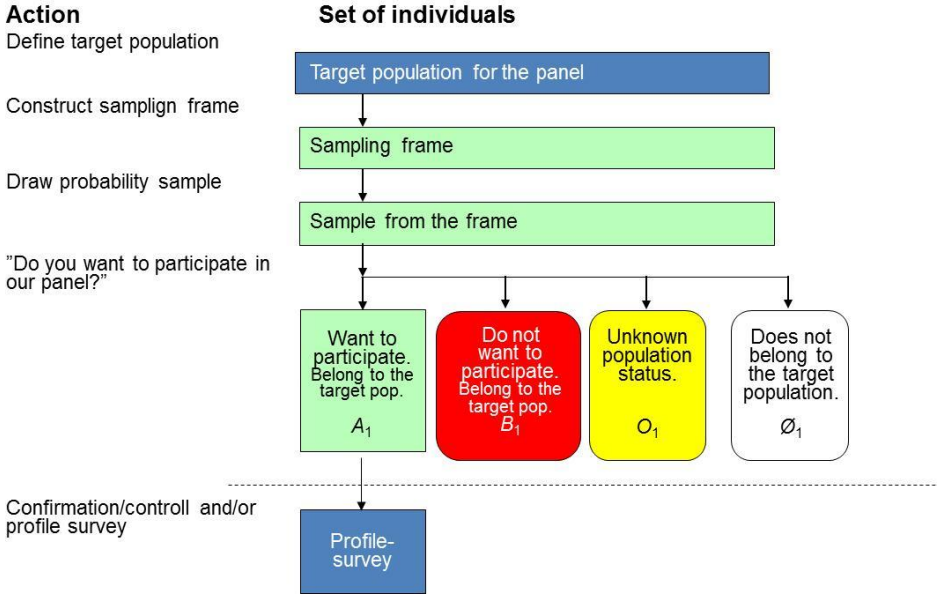
2.1 Measures regarding participation rates

There are four metric measures regarding participation rates; recruitment rate (RA[1]), profile rate (PA), participation rate (DA) in a specific survey and the cumulative participation rate (KA). These measures, as well as some other proposed measures, are largely inspired by Callegaro and DiSogra (2008) and slightly adapted to Swedish conditions. In figure 1 the recruitment phase for probability sampling from the target population is outlined. From the frame a probability sample is drawn according to a specific design and the sampled individuals are contacted, often by postal mail or by telephone, and asked whether they want to participate in a web panel. This type of (direct) recruitment is based on statistical theory and is the most expensive one. With

[1] We have kept the Swedish notation *RA* instead of changing into *RR* which is more intuitive in English for *R*ecruitment *R*ate.

this type of recruitment the only purpose of the "recruitment survey" is to recruit panelists. Sometimes the recruitment is integrated in another survey (indirect recruitment), and the recruitment question is one among many different questions. In both kinds of recruitments the sampled individuals can be classified into one of the four categories: (i) want to participate in the panel (the number is denoted A_1), (ii) do not want to participate (B_1), people not belonging to the population ($Ø_1$) or (iv) no contact is established which means that their eligibility status is unknown (O_1).

Figure 1. Recruitment phase. Direct recruitment based on probability sample.



The recruitment rate (*RA*) can then be defined as

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

where u_1 is an estimate of the proportion in category O_1 belonging to the population. It is unusual that the recruitment is solely done at one single occasion. More common is that recruitment is done at several separate occasions. In this situation RA can be calculated roughly as a mean over the separate recruitment phases.

The people in category A_1 who accept to participate in the panel then has to participate in a profile survey. This procedure is sometimes referred to as double opt-in. In single opt-in, no profile survey is conducted. A figure similar to figure 1 can be constructed for the profile survey, but due to space limitations it is omitted here. The profile rate can then analogously be defined as

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

where the terms are defined analogously with expression (1a). After the profile survey the panel is ready for use. Specific surveys can be conducted directly towards the whole panel or subgroups in the panel. Normally a probability sample is drawn from the panel and after the data collection the sampled individuals can be classified into the categories responders (A_3), non-responders (B_3 , known to belong to the population), over coverage ($Ø_3$) and unknown eligibility status (O_3). Based on these categories the participation rate in a specific survey can then analogously be defined as

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

where u_3 is an estimate of the proportion in category O_3 belonging to the population.

Based on *RA, PA* and *DA* we can define the cumulative participation rate as

$$KA = RA \times PA \times DA \quad (1d)$$

The cumulative response rate takes into account all steps in the process from the sample in the recruitment to the response set in a specific survey. This means that *KA* resembles the response rate in a traditional probability sample survey and can hence be used to comparisons between web panel surveys and traditional probability sample surveys. Note that the *RA*, and hence *KA*, can only be calculated if the recruitment is based on a probability sample, which means that panels with recruitment based on self-selection cannot calculate this measure.

2.2 Measures regarding the panel

A common trait for measures regarding the panel is that they cannot in a simple way be related to good or bad quality in survey results. The association between the measures and quality is complicated and can vary between surveys. Merely the ability to calculate the measures can be seen as an indicator of good survey culture at the panel provider.

Panel size

The panel size can be seen as basic information regarding the panel and not as a measure for the panel quality. The panel size should be stated for a given date since the panel size varies over time. The measure for panel size is defined as

$$PS = \text{Number of active members} \quad (2)$$

The definition of an active panel member is based on the ISO 26362 (2009) definition.

Response burden

The measure response burden highlights the number of surveys the panelists are confronted with during a certain period of time. The recommendation is to use calendar year. Let

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

$I_t =$ the total number of invitations during year *t* sent out to the active panelists in PS_t

$K_t =$ the total number of complete responses during year *t* from the active panelists in PS_t.

Based on this we propose the two following measures for response burden

$$UB_1 = \frac{I_t}{PS_t} \quad (3a)$$

$$UB_2 = \frac{K_t}{PS_t} \quad (3b)$$

The measures can be interpreted as

$UB_1 =$ the average number of invitations per panelist during year *t* for the panelists who were active December 31 year *t*.

$UB_2 =$ the average number of complete responses per panelist during year *t* for the panelists who were active December 31 year *t*.

Dominance from certain panelists

Sometimes there is a relatively small amount of panelists who stands for the most responses. This can lead to bias in the results if these panelists are not "representative" for the rest of the panel or the underlying population. For the measure, we consider the panelists who are active a certain date e.g. December 31 year *t*. The measure uses K_t

from response burden above and the following quantity: B_{20} = the total number of complete responses during year t from the 20 percent most active panelists. The measure to cast some light on the problem with dominance is then defined as

$$D_{20} = \frac{B_{20}}{K_t} \quad (4)$$

This measure can be interpreted as the percentage of all complete responses that the 20 percent most active panelists stands for which can be used for discussions whether dominance from certain panelists exist.

In addition, we also propose a measure for attrition rate which is inspired by Calegaro and DiSogra (2008) and a measure for the proportion of newly recruited.

2.3 Measures regarding a specific survey

In a specific survey the topic and the questions can vary a lot. Some of the measures require choosing one or two questions. It is then up to the panel provider in consultation with the buyer to carefully choose those questions.

Conditioning

According to ISO 26362 (2009) the panel provider shall try to identify and remove inattentive and fraudulent panelists. Another type of conditioning is when panelists change their attitudes or become more observant as a consequence of survey participation. This is a type of conditioning that can be called professionalism. We propose three measures regarding conditioning. The idea behind all three measures is that differences in response pattern should be detected between different groupings of the respondents. For each measure the responders should be divided into three groups according to the number of surveys they have participated in. One possible grouping is

- Group I: responders who have participated in 5 or fewer surveys
- Group II: responders who have participated in 6 to 20 surveys
- Group III: responders who have participated in 21 or more surveys

Then

1. Choose one or a couple of questions with relatively high partial non-response
2. Register the individual time to finalize the survey for all responders
3. Choose one of the most important subject matter questions and dichotomize it (for simplicity) where one category is for example the most positive scale alternative and the other category is the other alternatives.

Three measures that on an overall level can be used to indicate presence of conditioning is to calculate the following quantities for each of the three groups

$$AS_1 = \frac{\text{the number of partial non-responses in the group}}{\text{the total number of persons in the group}} \quad (5a)$$

$$AS_2 = \text{the median time to finalize the survey in the group} \quad (5b)$$

$$AS_3 = \frac{\text{the number "most positive-answers" on the question in the group}}{\text{the total number of responses in the group}} \quad (5c)$$

If there is no conditioning the three groups should have roughly the same values for all three measures. It is not clear cut what constitutes small and large differences so the results must be carefully interpreted by the panel provider together with the buyer.

In addition, we also propose two measure suggested by Calegaro and DiSogra (2008); Absorption rate (a measure that indicates the quality of the register of email addresses) and Break-off rate (the number of opened but not completed web questionnaires compared to the sum of completed and not completed questionnaires).

3. Discussion

With the development of web panel surveys statistical methodological aspects has not always been in focus. Instead, the progress has been driven by clever and feasible IT-solutions. Statisticians have not played the central role they should when new modes of data collections emerge, this is true at least in Sweden.

Web panel surveys pose a new challenge for statisticians since results cannot be evaluated with traditional statistical tools, e.g. precision such as confidence intervals. Even if the panel recruitment is done with probability samples, the non-response in all phases is so large that the final response set usually cannot be regarded as a probability sample from the population. Whether the inference to the underlying population is correct or not, cannot be evaluated based on a statistical scientific platform. Instead, new types of measures based on different considerations have to be proposed and used to make quality discussions on the whole feasible.

In the new proposed standard in Sweden some measures are presented. We are aware of the fact that the measures are not always easy to interpret and their connections to quality in survey results are not unambiguous. Since the measures have been presented in national conferences and workshops and the standard has been scrutinized in a remittance version, we have some empirical evidence that it has been well received. One aspect we focus on in the report is transparency, i.e. that the whole web panel survey industry have to be more transparent with their procedures and methodologies. One apparent empirical result from the standard is that quality issues nowadays are discussed in a coherent manner in Sweden, which has not been the case during the last 10 years. The proposed measures facilitate the communication between panel providers and users of the results, which is hopeful.

Up till now there are four companies that have declared willing to calculate the proposed measures for their panels and a chosen specific survey. Their results will be publicly published on the Swedish Survey Society web page and can serve as a bench-mark for other companies and users. Our hope is that more companies are willing to calculate the measures for their panels and that their results also can be published.

Besides the proposed measures the report also suggests that the web panel provider should in detail verbally describe: (i) the recruitment method, (ii) the sampling design used for the specific survey and (iii) the calculation of any sampling weights.

We hope that serious discussions in Sweden regarding quality can emerge from the proposed standard. However, there is much work to be done before web panel surveys can be recognized as a scientific reliable method.

References

AAPOR (2004). *Standard Definitions: Final Dispositions of Case Codes and Outcomes for Surveys.*

AAPOR (2011). *Standard Definitions: Final Dispositions of Case Codes and Outcomes for Surveys*. 7:th Edition.

AAPOR (2010). *AAPOR Report on Online Panels*.

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

DiSogra, C. och 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*.

SS-ISO 26362:2009 Access panels in market, opinion and social research – Vocabulary and service requirements (ISO 26362:2009, IDT)

Swedish Survey Society (2005). *Standard for response rate calculation*. Manuscript in Swedish only.