Journal of Official Statistics, Vol. 20, No. 3, 2004, pp. 451–465

# List-based Web Surveys: Quality, Timeliness, and Nonresponse in the Steps of the Participation Flow

*Monica Pratesi*[1], *Katja Lozar Manfreda*[2], *Silvia Biffignandi*[3], *and Vasja Vehovar*[4]

The debate on quality issues in web surveys is open and lively (see http://www.websm.org, i.e., the Web Survey Methodology site). Data quality is required to satisfy the user's needs. Improving the survey process quality is a precondition for obtaining product quality at acceptable cost. This article contributes to the debate focusing on the timeliness of web surveys. From our perspective the timeliness of data collection is mainly due to the timeliness of response from the members of the eligible population. We identify several steps in the web survey process and divide the final response rate into different components, one for each step. We model the survival of eligible respondents, finding out which participants come farthest in the process of a web survey. The analysis contributes to the efforts to explore the nonresponse process and to shorten the individual survey period length.

*Key words:* Web survey process; drop-outs; survival analysis; nonresponse.

## 1. Introduction

The spread of Internet technology in the target population (individuals or establishments) affects several dimensions of quality in the design of web surveys. Among these, the timeliness of data collection is considered one of the most competitive advantages of the Internet with respect to other survey modes. From our perspective, the timeliness of data collection is linked to the timeliness of response by the eligible population and to the process of participation in a list-based web survey. By a list-based web survey we mean one where a list of individual units (either individuals or establishments) that were invited to participate in the survey exists. Such surveys are different from other types of web surveys, e.g., from self-selected web surveys where general invitations to participate in the survey are usually placed on web sites, and anyone can access the web questionnaire. For a discussion of types of web surveys, see Couper (2000).

In list-based web surveys timeliness has several aspects that need to be investigated. We refer to the survey period length, the speed of data collection, and processing tools and their influence on contact and response rates. The Internet could theoretically shorten the data collection period, but in practice the average time of response often amounts to

[1] Department of Mathematics, Statistics, Computational Science and Applications of the University of Bergamo, Via dei Caniana 2, I-24129 Bergamo, Italy. Email: monica.pratesi@unibg.it; m.pratesi@ec.unipi.it
[2] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia. Email: katja.lozar@uni-lj.si
[3] Department of Mathematics, Statistics, Computational Science and Applications of the University of Bergamo, Via dei Caniana 2, I-24129 Bergamo, Italy. Email: silvia.biffignandi@unibg.it
[4] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia. Email: vasja.vehovar@uni-lj.si
**Acknowledgment:** This article was drafted and revised within the MIUR 40% project 1999 and 2001 for the Italian research collaboration (Biffignandi and Pratesi).

several days, and also the reaction to the e-mail invitation and follow-ups results in late first access to the questionnaire (Biffignandi and Pratesi 2002; Bosnjak et al. 2001).

The process of participation in a list-based web survey starts with sending out e-mail or mail invitations to individuals on the list and ends with the response or nonresponse of members of the eligible population. Each member of the population has his or her own behavior after receiving the invitation. The survival time in the survey process has to be modeled in order to find out which participants come farthest in the survey, and which steps of the survey participation process contribute most to the timeliness of response and of data collection.

In this article we identify several steps of the web survey process (Vehovar et al. 2002) and divide the final response rate into different components, one for each step. For example, for telephone pre-recruited web surveys of Internet users, the following steps can be identified after the telephone pre-recruitment (telephone interview, e-mail address collection): e-mail recruitment (sending out e-mail invitations, absorption (acceptance) of e-mails by the network), access to the questionnaire, and questionnaire completion step. The steps of participation in the web survey are illustrated in Section 2. Survival in the survey process is modeled in Section 3. The proposed models are applied to a web survey targeting the general Internet population in Slovenia. The case study is discussed in Section 4.

## 2. The Participation Process

An overview of the stages in the participation process for web surveys in general, both for individuals and establishments, is given by Biffignandi and Pratesi (2000a and 2000b), Lozar Manfreda (2001), and Vehovar et al. (2002). The complete set of stages in a telephone pre-recruited web survey is discussed in Lozar Manfreda et al. (2002). Here we are interested only in the stages from the absorption of the e-mail invitation to the completion of the web questionnaire.

Given an e-mail list of size $N$, noncontact ($NC$) and contact ($C$) are the first results (possible events, outcomes) of the survey process. Noncontact leads to nonparticipation in the survey, whereas contact presents a further dichotomous possibility of participation or nonparticipation.

In this case the step of completing the questionnaire is described in relation to the empirical data available, as described in Section 4 (a more detailed elaboration of the completion process of the web questionnaire is described in Bosnjak et al. 2001). We distinguish four steps in the participation flow presented in Sections 2.1 through 2.4.

### *2.1. E-mail invitation*

E-mail invitations to answer the web questionnaire are sent. Not all the intended recipients notice and read the invitation. Missed contacts can be caused by an error in the e-mail address ($WE$) (e.g., the e-mail is returned because of nonexistent or unknown receiver or domain) or because of temporary network problems in the domain of the e-mail address ($NE$). When the list of e-mail addresses contains wrong e-mail addresses ($WE$), or there is a network problem ($NE$), this causes a missed contact and produces a *coverage error*. Checking the list before the survey starts, in order to discover and correct any wrong

e-mail addresses, can reduce this error. However, recipients (individuals or establishments) may also change their e-mail addresses after the survey has already started, making it difficult to correct them during the survey period. Only those whose e-mail address is absorbed by the network are included in the next stages of the web survey process.

### *2.2. Access to the introductory web questionnaire page*

The recipient of the e-mail invitation may decide to access the introductory page of the questionnaire. Those who access this page may decide to proceed to the next stages. Not all those who access it proceed immediately to completion of the questionnaire. Some need several accesses on different days to complete the questionnaire; others limit their reaction to the first access, even after having been followed up by additional e-mail invitations. Every recipient who accesses the introductory page is considered to have been contacted by the survey organization even if he or she only accessed but did not answer any questions. Other recipients limit their reaction to the sending of a return receipt of the e-mail invitation, without accessing the introductory page. These too are considered as having been contacted by the survey organization.

### *2.3. Clicking "Start the questionnaire"*

Those who access the introductory page may decide to begin to complete it. Evidence for this is clicking the Start button. Some of the recipients who access the introductory page may decide to click the Start button but then do not proceed to answering the survey questions; they simply start the questionnaire and quit it. Others proceed towards completion of the questionnaire. Remember that this stage is not necessarily present in all web surveys. It is possible that the survey questions already appear on the introductory page. However, since in most cases some introductory page is present, we decided to consider this stage separately.

### *2.4. Completion of the questionnaire*

Those who begin to answer the questionnaire may decide: a) to complete it (complete response *CR*, i.e., those who, through answering, arrive at the final part of the questionnaire, although they do not necessarily answer all the questions (some item nonresponse is possible)) or b) to quit it after having completed it only partially (partial response, due to either voluntary or involuntary interruptions, *PR*). Note that interruption is not necessarily voluntary. For example, there may be a technical problem on the part of the survey organization resulting in respondents quitting the questionnaire. Evidence of both behaviors is documented in log files after submit buttons on each questionnaire page are clicked. In some cases the recipient explicitly refuses to participate, sending an e-mail message to the survey organization asking to be excluded from the survey.

Each member of the eligible population whose e-mail address is absorbed by the network and who notices and reads the invitation is a candidate for participating in the survey. At the end of the data collection period each sampled person who reacts to the initial e-mail invitation or to the follow-up messages passes through one of the described

stages. At some time (immediately or many days after the invitation) he or she withdraws from the survey process, either after completing the questionnaire (full or partial completion), after only clicking the start button without answering any question, or after only one or more accesses to the introductory page without proceeding further.

There are some members of the eligible population whose e-mail addresses are absorbed by the net, but who do not react to the invitation. This group of people is comprised of individuals with whom contact has not been made (they do not see and read the invitation) and individuals who are contacted (they notice and read the invitation) but decide not to react. In other words, the latter implicitly refuse to participate in the survey. It is impossible to distinguish one type of situation from the other. The e-mail invitations that are not followed by a reaction are called nonreactions. In our context, the absence of reaction of the intended recipient is considered completely due to a missed contact. Under this assumption, the nonreactions plus the missed contacts due to incorrect e-mail addresses or to network problems in the domain of the e-mail address make up the noncontact set ($NC$).

At the end of the planned survey period, the rates presented in Table 1 give a picture of the level of participation obtained in the survey.

*Table 1. Contact and response rates associated with the participation flow*

| | |
|---|---|
| Absorption rate | $\frac{N - WE + NE}{N}$ |
| Contact rate | $\frac{C}{N}$ |
| Response rate given contact | $\frac{PR + CR}{C}$ |
| Response rate of the absorbed e-mails | $\frac{PR + CR}{N - WE - NE}$ |
| Global response rate | $\frac{PR + CR}{N}$ |

## 3. Modeling the Survival in the Survey Process

Timeliness of data collection depends on the distribution of the individual reaction times and on the number of drop-outs from each of the previous stages. Each sampled person may decide to quit the survey process at one of the stages after the absorbed invitation or to proceed toward the next step until the full completion of the web questionnaire. The individual reaction times vary among sample persons. In theory, all sampled persons could cooperate immediately, avoiding the nonresponse problem, and their survival in the survey process could be as short as allowed by the communication times on the Internet. In practice, however, some individuals cooperate immediately, others respond only after many invitation messages, still others do not react at all or limit their reaction to one or more visits to the web questionnaire introductory page without answering any questions. Only a part of the eligible population reacts immediately to the invitation. Quick reaction means short individual survey period length and, thus, short survival time in the survey process and timeliness of data collection.

The reaction time is the result of both the demographic and socioeconomic characteristics of participants and the survey conditions. By the latter we mean the absorption rate of the e-mail invitations and the time schedule of the follow-up messages. It is clear that timeliness of data collection can be achieved by promoting a fast response from all members of the eligible population. In order to understand which conditions favor a short survival time in the survey process and to encourage participation in the survey, we

model the survival time in the survey process by specifying the dependence on individual covariates of the exit time from this process.

In our analysis the survival time in the survey process is the time between the e-mail invitation and the drop-out. Drop-out can be due to full completion of the web questionnaire, partial completion of the web questionnaire, and access or login with no items completed. This last category includes those who only access the introductory page of the questionnaire and also those who click the Start button but never answer any questions.

Let $T_i$ be a random variable denoting the time of drop-out for sample person $i$ and let $J_i$ be a random variable denoting the type of drop-out (Type 1: full completion; Type 2: partial completion; Type 3: access only). Thus $J_i = 2$ means that the sample person $i$ exits from the survey process after having partially completed the questionnaire. The type-specific hazard of drop-out at time $t$ for sample person $i$ due to reason $j$, $h_{ij}(t)$, is defined by

$$h_{ij}(t) = \lim_{\Delta t \to 0} \frac{P(t < T_i < t + \Delta t, J_i = j | T_i = t)}{\Delta t} \quad \text{for} \quad j = 1, 2, 3 \tag{1}$$

The dependence on covariates of the type-specific hazard can be modeled through a general proportional hazard model for all three drop-out types

$$\log(h_{ij}(t)) = \alpha_j(t) + \beta_j x_i(t) \quad \text{for} \quad j = 1, 2, 3 \tag{2}$$

where $x_i(t)$ is a vector of covariates, some of which may vary with time. The coefficient vector $\beta$ is subscripted to indicate that the effect of the covariates may be different for different drop-out types. Also the dependence of the hazard on time $\alpha(t)$ may vary across drop-out type.

We are treating events other than those of immediate interest as a form of censoring. The censoring mechanism should be noninformative if the estimates are to be unbiased. In other words, we must assume that, conditional on the covariates, those sampled persons who are at a particularly high (or low) chance of one event type (e.g., completing the questionnaire) are not more (or less) likely to experience other kinds of events.

## 4. The Case Study

Two telephone surveys based on a probability sample of Slovene households over a period of three weeks in June and July 2001 were used to find the eligible population. The first survey (June 14-July 1, 2001) was on the use of information-communication technology among the general population between 10 and 75 years of age. Respondents who claimed to be Internet users were asked to give their e-mail addresses for the purpose of a web survey. From responding Internet users in this survey, 140 e-mail addresses were collected. The second survey (July 4-July 9, 2001) based on a probability sample of Slovenian households, collected 208 e-mail addresses by screening for the person within the household who most often uses the Internet.

The response rate for the two telephone surveys was 46% (903 respondents among 1,963 sampled household telephone numbers; establishment telephone numbers are excluded). This figure is typical of the response rate for Slovenian telephone surveys conducted by independent research agencies. In our case, the telephone pre-recruitment was done by a Slovene research agency, CATI Center (www.cati.si).

Altogether, 348 e-mail addresses were collected, and on July 12, 2001, e-mail invitations were sent to all 348 addresses. In the e-mail message they were given the URL address of the web questionnaire, which included their identification number. The first e-mail follow-up was sent to nonrespondents on July 30, 2001, and the second on August 28, 2001. Because it was summer (holiday time), time between follow-ups were somewhat longer than what is common in Internet surveys.

### *4.1. Participation flow*

The survey period (time between the first e-mail invitation and the last reaction) was 67 days. The number of absorbed e-mails was 294. The behavior of the 294 recipients of the e-mail invitation is described in Table 2. For each step of the participation flow, we indicate how many sample persons withdraw and how many remain in the survey process. Table 3 contains the rates that are associated with the participation flow and are calculated at the end of the survey period. We call complete respondents those who reached the final part of the questionnaire although they did not necessarily answer all the questions.

*Table 2. The steps of the participation flow*

| Survey period starts | Step 1: E-mail invitation | Step 2: Access to introductory page | Step 3: Clicking Start | Step 4: Questionnaire completion |
|---|---|---|---|---|
| Invitations sent = 348<br>Wrong e-mail = 54<br>No reactions = 81<br>Absorbed e-mails = 294 | *Drop-outs*: Access intro only = 11<br>*Stay in*: 283 | *Drop-outs*: Start quest only = 29<br>*Stay in*: 254 | *Drop-outs*: Partial comp. = 50<br>*Stay in*: 204 | *Drop-outs*: Full comp. = 123 |

The percentage of undeliverable e-mails was about 15%: 84.5% of the invitations sent were absorbed by the network (*absorption rate* defined as absorbed e-mails on invitations sent). The percentage of responses obtained from the absorbed e-mails was 59%. Evidence of contact with eligible respondents was obtained in 213 cases: those who participated after the contact were 81% (*response rate given contact*). The percent of respondents among all the invitations sent was 50% (*global response rate*).

*Table 3. Rates associated with the participation flow*

| | | |
|---|---|---|
| Rate of absorbed e-mails | 294/348 | = 0.85 |
| Contact rate | (11 + 29 + 50 + 123)/348 = 213/348 | = 0.61 |
| Global response rate | (123 + 50)/348 = 173/348 | = 0.50 |
| Response rate given contact | (123 + 59)/(11 + 29 + 50 + 123) = 173/213 | = 0.81 |
| Response rate of the absorbed e-mails | (123 + 50)/294 = 173/294 | = 0.59 |

### *4.2. Survival in the survey process*

The behavior of the contacted individuals after the e-mail invitation determines their survival in the survey process and how long they wait before the next stage of the process. We identify three groups of contacted individuals in relation to the three ways of dropping out of the process:

- 123 respondents complete the web questionnaire (Group1 – Full completion),
- 50 limit their participation to partial completion of the web questionnaire (Group 2 – Partial completion),
- 40 react to the invitation only by accessing the questionnaire. In this last category those who only access the introductory page of the questionnaire (11) and those who only click on the Start button (29) are combined (Group 3 – Access only).

The survival of each group is studied with their survivor functions estimated by the life table method. More precisely, the data are processed using the survival analysis procedures implemented by the SAS system (Allison 1998). Each survivor function gives the probability of surviving, in other words the probability of remaining in the survey process (not withdrawing) beyond time *t*. The origin of the time (0 point) is the date of the e-mail invitation. The survival for the three groups is illustrated in Figure 1.

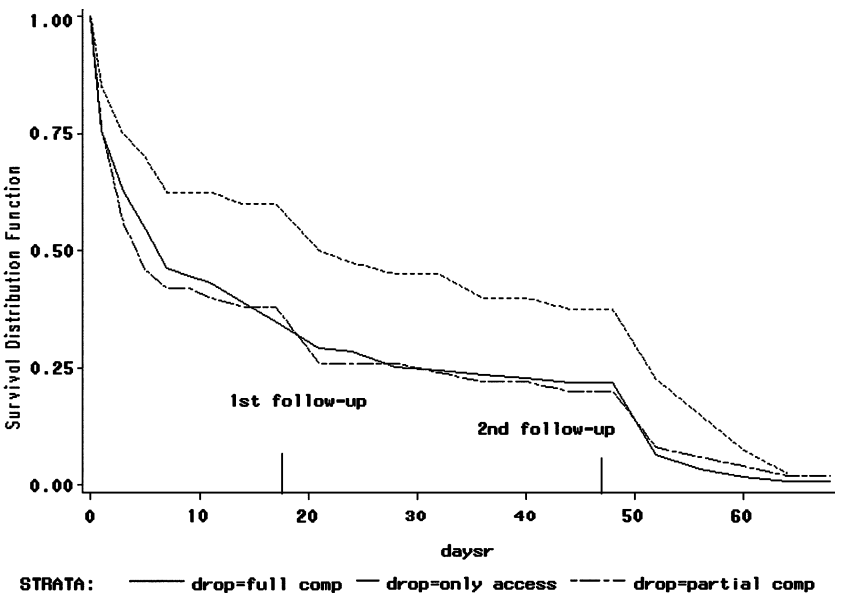


*Fig. 1. Drop-outs from the survey process*

The probability of survival decreases dramatically after sending out the e-mail invitations. Most reactions happen during the first 10 days of the survey period; 110 (31 partial and 79 complete) of the 173 responses were obtained before the first follow-up (during the first 18 days). The others were obtained after two follow-ups: the first sent 18 days after the initial invitation, the second after 48 days. There were an additional 24 responses (8 partial and 16 complete) after the first and before the second follow-up, and 39 responses (11 partial and 28 complete) after the second follow-up. People have a prompt reaction to follow-ups: the probability of remaining in the survey decreases two days after the first and the second follow-up.

The tests of equality of the survivor functions over the three groups are significant (Log Rank: Chi square 5.80, at $p = 0.0505$ level, Wilcoxon: Chi square 6.38 $p = 0.04$). The hypothesis that the three survivor functions are equal is therefore rejected. In other words, the three groups have different survival behavior in the survey. While respondents (partial and complete) might have similar determinants, drop-outs after only accessing the questionnaire can be considered a distinct phenomenon.

People who only access the questionnaire have a higher probability of survival in the survey process than the partial or complete respondents. That means that they react to the e-mail invitation later than partial and complete respondents. The gap between their survival curve and the respondents' survival curve is evident. The average survival time in the survey process of those who accessed the questionnaire but did not complete any item is 27 days (median 21 days) from the invitation, compared to 16 days for partial respondents and 17 days for complete respondents (see Table 4). The first group's longer survival in the survey process is due also to their larger number of accesses to the questionnaire in comparison with respondents: 2.25 times (std dev 1.07) versus 1.64 times for partial and 1.48 times for complete respondents (see Table 5).

*Table 4. Average survival time (in days) in the survey period*

| Type of drop-out | Average time | | | |
|---|---|---|---|---|
| | Mean | Median | Std dev | *N* |
| Access only | 26.95 | 20.5 | 23.58 | 40 |
| Partial completion | 16.18 | 4 | 21.06 | 50 |
| Full completion | 16.78 | 5 | 20.34 | 123 |
| Total | 18.54 | 4 | 2.06 | 213 |

*Table 5. Average number of accesses by type of drop-out*

| Type of drop-out | Average number of accesses | | | |
|---|---|---|---|---|
| | Mean | Median | Std dev | *N* |
| Access only | 2.25 | 1 | 1.07 | 40 |
| Partial completion | 1.64 | 1 | 1.24 | 50 |
| Full completion | 1.48 | 1 | 1.07 | 123 |
| Total | 1.66 | 1 | 2.06 | 213 |

The partial or complete respondents have a common pattern of survival in the survey process: their survival curves are virtually indistinguishable. People who participate in the survey by answering the questionnaire items (partial completion and full completion) stay in the survey process for a shorter period of time (16 and 17 days, respectively) than the access-only group (27 days), but the partial respondents, statistically, do not respond more quickly than the complete respondents. The distribution of responses is asymmetric, with median survival time of respondents equal to 4 days, since the majority of responses are obtained before the first follow-up.

Most completions are done on the day of the first access. People who accessed the questionnaire several times in the same day are grouped together and considered as if they reacted only on first access. Accesses made on different days are called separate accesses. 32.65% of the people contacted complete the questionnaire on first access partial completion is also done mostly on first access (see Table 6).

*Table 6. Drop-outs from the survey process by number of accesses to the questionnaire*

| Type of drop-out | Frequency | | Percentage of absorbed e-mails | |
|---|---|---|---|---|
| Access only | | 40 | | 13.61 |
| Single access | 25 | | 8.51 | |
| Multiple accesses | 15 | | 5.10 | |
| Partial completion | | 50 | | 17.01 |
| Single access | 39 | | 13.27 | |
| Multiple accesses | 11 | | 3.74 | |
| Full completion | | 123 | | 41.83 |
| Single access | 96 | | 32.65 | |
| Multiple accesses | 27 | | 9.18 | |
| No reaction | | 81 | | 27.55 |
| Total | | 294 | | 100 |

Figure 2 shows the evolution of the first accesses to the questionnaire during the 68 days of the survey period. The first accesses may result in complete answers or some other drop-out type. The *Y*-axis is the percent of first accesses and of completion at first access (both percent of all invitations), and the *X*-axis is the day from the e-mail invitation. The accesses were gathered after the dates of the first e-mail invitation and of the follow-ups.

The average time before the first access is generally lower than the global survival time (the general mean overall 13 days, the median 6 days; the mean for complete respondents only 12 days, and the median 4 days; see Table 7). The number of second accesses is not high (68 out of 213 contacted), and among these there are some late full completions (mean time of the second access is 25 days, with median 24 days; see Table 8).

### *4.3. Modeling the survival*

Since most of the reactions were obtained at the first access, we decided to model the survival of participants from the e-mail invitation to the first access, distinguishing drop-outs at the first access from drop-outs later in the survey period.

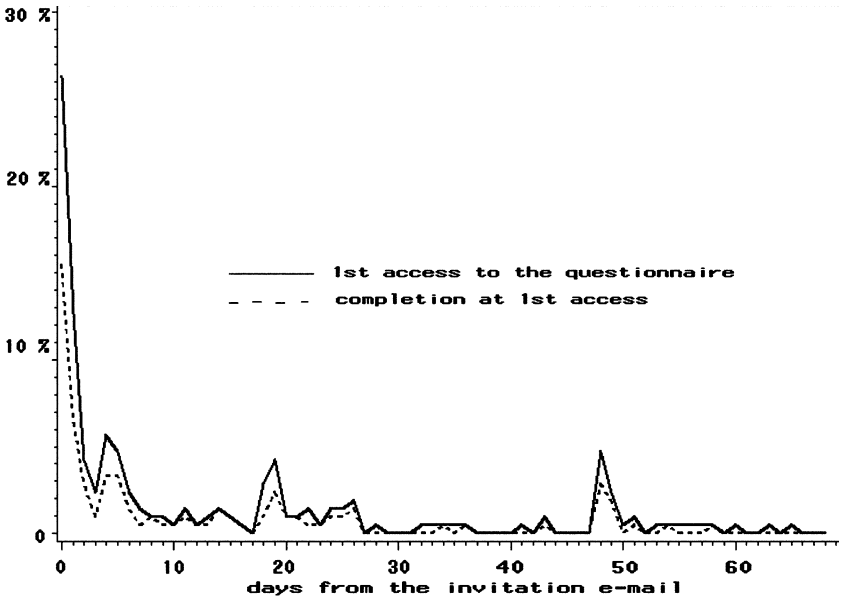


*Fig. 2. Distribution of first accesses and completion at first access (percent of all invitations) by number of days after the first e-mail invitation was sent*

*Table 7. Average time (in days) before the first access to the questionnaire*

| Type of drop-out | Average time | | | |
|---|---|---|---|---|
| | Mean | Median | Std dev | *N* |
| Access only | 17.07 | 5.5 | 20.55 | 40 |
| Partial completion | 13.38 | 3 | 18.96 | 50 |
| Full completion | 12.06 | 4 | 12.06 | 123 |
| Total | 13.31 | 6 | 17.61 | 213 |

*Table 8. Average time (in days) before the second access to the questionnaire*

| Type of drop-out | Average time | | | |
|---|---|---|---|---|
| | Mean | Median | Std dev | *N* |
| Access only | 19.63 | 19 | 15.67 | 19 |
| Partial completion | 22.69 | 12.5 | 23.33 | 16 |
| Full completion | 24.64 | 24 | 21.95 | 33 |
| Total | 22.78 | 21.5 | 20.55 | 68 |

The effect of covariates on each type of specific hazard can be tested by fitting a Cox model (Equation (2)) to each type of drop-out. The goal is to test whether the effect of the covariates is the same or different across types of events (drop-outs) and to find out which participants come farthest in the process of the web survey.

The models included several variables collected at the pre-recruitment stage of the survey (during the pre-recruitment telephone survey). A set of standard social demographic variables, variables on the technical equipment of participants, characteristics of Internet

usage, problems with the Internet and attitudes towards surveys were collected. A complete description of the variables can be found in Lozar Manfreda et al. (2002). Here we report only the meaning and the categorization of the significant independent variables. The models included the following variables:

- Time between sending (receiving) the e-mail invitation to participate in the survey and drop-out at the first access or later, measured in hours, minutes and seconds.
- Place from where the Internet is accessed: frequent access from home ($n = 118$); frequent access otherwise ($n = 95$);
- Purpose of Internet usage: use the Internet often or regularly for business purposes without considering the place of access ($n = 84$); use the Internet otherwise ($n = 129$); access from work and reason is work ($n = 57$); access otherwise ($n = 156$);
- Age of the contacted person: age between 21 and 40 years ($n = 137$); other ages ($n = 76$);
- Educational level of the contacted person: university level ($n = 77$); other educational level ($n = 136$).

The models adopted are the following:

- Model 1 – *Late reactions*: The model applies to sample persons who need several separate sessions of work to start the questionnaire, or to access the introductory page, or to complete the questionnaire fully or partially. They are considered sample persons who react late to the invitation. The model concerns the competing risk of having a late reaction (see Table 9).

*Table 9. Model 1: Late reactions versus others*

| Variables in the equation | $\beta$ | S. E. | Chi-Square | Sig. | H. R. Exp($\beta$) |
|---|---|---|---|---|---|
| Use of Internet for business[a] | 0.80 | 0.28 | 8.21 | 0.004 | 2.22 |

[a] Variables entered in Step 1: bus
Testing Global Null Hypothesis: Likelihood Ratio test: Chi-Square = 8.14, $p > 0.0043$

- Model 2 – *Access only to the questionnaire*: The model refers to sample persons who only access the questionnaire without completing it (fully or partially). The model concerns the competing risk of only accessing the questionnaire (see Table 10).
- Model 3 – *Slow respondents*: The model refers to the sample persons who do not complete (fully or partially) the questionnaire in only a single session of work. They are considered slow respondents. The model concerns the competing risk of not being a fast respondent (see Table 11).

*Table 10. Model 2: Only accesses to the questionnaire versus others*

| Variables in the equation | $\beta$ | S. E. | Chi-Square | Sig. | H. R. Exp($\beta$) |
|---|---|---|---|---|---|
| [a]Age between 21 and 40 years | −1.01 | 0.34 | 8.82 | 0.0030 | 0.36 |
| Frequent access from home | −0.71 | 0.34 | 4.30 | 0.0380 | 0.49 |

[a] Variables entered in Step 1: age young, home
Testing Global Null Hypothesis: Likelihood Ratio test: Chi-Square = 10.33, $p > 0.0057$

*Table 11. Model 3: Slow respondents versus others*

| Variables in the equation | $\beta$ | S. E. | Chi-Square | Sig. | H. R. Exp($\beta$) |
|---|---|---|---|---|---|
| [a]University degree | −0.62 | 0.29 | 4.59 | 0.0322 | 0.54 |
| Access from work and for work | 0.94 | 0.30 | 9.68 | 0.0019 | 2.55 |

[a] Variables entered on Step 1: edu_u, wbus
Testing Global Null Hypothesis: Likelihood Ratio test: Chi-Square = 9.84, $p > 0.0073$

In deciding on the best-fitted model, we compare the models with standard tests (likelihood ratio test for nested models and Akaike information criterion for models that are not nested). We rely on the model chi-square testing the null hypothesis that all the coefficients in the model are zero. The effect of the covariates has been tested with a forward selection procedure. In order to test the significance of individual independent variables, the chi-square statistic is used. Hazard ratios for the outcome category of interest are obtained from the exponential value of the $\beta$ coefficient (Exp($\beta$)). The hazard ratio describes the effect of a significant independent variable on the risk of being in the outcome category: when values of all the other independent variables included in the model are kept constant, a hazard ratio greater than 1 indicates that the risk of being in the outcome category is increased by the independent variable, while a hazard ratio lower than 1 indicates a decrease in the risk of being in the outcome category.

The best-fitted Cox model for *Late reactions* is presented in Table 9. The model is statistically significant, rejecting the global null hypothesis that all the coefficients in the model are zero ($p > 0.0043$). The strongest effect in Model 1 is the reason for Internet usage. People who use the Internet often and regularly for business purposes seem to postpone participation in the survey. Internet users motivated by business are two times as likely as other users to exit later from the survey process (Hazard Ratio = 2.21). They therefore react later to the survey invitation, either by completing the questionnaire or by only accessing it.

Exits after having only accessed the questionnaire show a different pattern (see Model 2). The best-fitted model for A*ccess only to the questionnaire* is presented in Table 10. The model is statistically significant ($p > 0.0057$). The purpose of Internet use does not have a significant influence on the risk of only accessing the questionnaire. We do see an effect related to the place of access to the Internet and to the age of the Internet users: the age of the person contacted yields a .70 decrease in the risk of exit after having only accessed the questionnaire (Hazard Ratio = 0.36). Moreover, connecting from home yields a .50 decrease in the risk (Hazard Ratio = 0.49). The evidence shows that people between 21 and 40 years of age seem to be more willing to participate in the survey than others, and that, at the same time, people connecting to the Internet from home are more likely to go through the steps of the survey process than people connecting from school, university, work or other places (library, friends, relatives, clubs, Cyber cafés, and other).

The best-fitted model for S*low respondents* is presented in Table 11. The model is statistically significant ($p > 0.0073$). The only significant influences on the risk of being a slow respondent are educational level and purpose of Internet usage: those contacted who have a university degree have a 46 percent lower risk of not leaving the survey immediately (Hazard Ratio = 0.54), while the risk is doubled for Internet users for business purposes

who connect to the Internet from their workplace (Hazard Ratio = 2.55). It seems that educated people need less time to go through the steps of the participation process. Instead, people who are busy (at least busy when they receive the invitation) and used to the Internet because they connect from the workplace, seem to spend less time responding to the survey questions and seem to have a higher risk of being slow respondents.

## 5. Concluding Remarks

The method proposed for investigating participation in web surveys and the timeliness of the participant's survey response permits measurement of the times of reaction to e-mail invitations and modeling of survival in the survey process.

We have identified different levels of cooperation in the survey by distinguishing the steps in the survey process. After telephone pre-recruitment (telephone interview, e-mail address collection), the following steps have been identified: e-mail recruitment (sending e-mail invitations to take part in the web survey, absorption of e-mails by the network), access to the questionnaire, and questionnaire completion. In order to identify the characteristics of those who come farthest in the participation flow, people who fully or partially completed the questionnaire using only one session of work have been treated separately from people who browsed the pages during several sessions of work.

In our telephone pre-recruited web survey, the final response rate was separated into different components, one for each step in the survey process. More than half the eligible members of the target population proceeded towards the final step of the survey process: the contact rate was 72.4%, at net of the e-mail invitations that were not absorbed by the network. The final response rate (response rate of absorbed e-mails) was 58.8%. The response rate is higher if it is measured on contacts: we obtained 81.2% of responses among people contacted.

The length of the whole survey period was 68 days. The average time of response was approximately 16 days, and most of the complete responses were obtained on the first access to the web questionnaire. In fact, the median time of the first access to the questionnaire was 4 days (modal time 0).

Timeliness of data collection was highly influenced by the behavior at the first access to the questionnaire. As we have seen, most of the responses were obtained at the first contact, and this was highly influenced by the effect of the first e-mail invitation and of the two follow-ups. Most responses were also obtained before the first follow-up. This effect might be partially due to telephone recruitment. The 348 Internet users who gave their e-mail addresses were told that they would be invited to take part in a web survey.

Survey process survival of the eligible respondents has been modeled by comparing risk models in order to find out which participants come farthest in the process of a web survey, contributing to its timeliness. Prompt reactions at the first access seem related to the purpose of Internet usage: people who connect for business purposes do not seem to be willing to participate in the survey. However, well-educated users tend to respond faster than others. Also, people who connect from work for business purposes tend to respond later than others. Young users who connect from home are more likely to proceed from the contact stage to the response stage.

The results obtained are obviously limited to our target population; however, they suggest that:

- after having captured the interest of the eligible respondent, the first access to the questionnaire is the best occasion for response to be obtained;
- timeliness of survey response is mainly achieved by promoting the first access to the questionnaire at the beginning of the survey period through a suitable timetable of e-mail invitations.

Survival in the survey process depends mainly on the features of participants' Internet usage: purpose of Internet usage and place of connection seem to be the key factors regarding the speed of reaction of eligible respondents.

Future research will focus on the collection of additional data on Internet users. Our findings suggest that nonresponse rates and quick reactions in web surveys are sensitive to the individual's characteristics: demographics were not significant in this application, but they could reasonably be included in other case studies. In the near future web-based data collection will require additional theoretical work. Empirical studies are also important to gain further experience with the approach we propose. Also, since many applications of this approach would use information extracted from the web surveys' error profile, it is important to develop software and to promote knowledge about web surveys that would facilitate such an information extraction. The thematic network project "Web-based data collection. Online knowledge base and co-operation platform on the methodology of web surveys and other types of web-based data collection" (acronym WebSM site) on which we are currently working is designed to establish a web portal on web surveys and is a step in this direction. This project is conducted by the following partners: University of Ljubljana Faculty of Social Sciences, Centre of Methodology and Informatics (Slovenia), Linkoeping University, Department of Mathematics (Sweden), University of Bergamo, Department of Mathematics, Statistics, Informatics and Applications (Italy), Zentrum für Umfragen, Methoden und Analysen, and AG OnlineResearch (Germany).

## 6. References

Allison, P.D. (1998). Survival Analysis Using the SAS System. A Practical Guide. SAS Institute.

Biffignandi, S. and Pratesi, M. (2000a). Modeling Firm Response and Contact Probabilities in Web Surveys. Proceedings of the Second International Conference on Establishment Surveys, Buffalo, New York, 17–20 June, 1528–1533.

Biffignandi, S. and Pratesi, M. (2000b). Le indagini via internet sulle imprese: aspetti metodologici e un'analisi dei rispondenti. In Tecnologie informatiche e fonti amministrative nella produzione di dati, C. Filippucci (ed.), Franco Angeli, Milan. [In Italian]

Biffignandi, S. and Pratesi, M. (2002). Internet Surveys: The Role of Time in Italian Firms' Response Behaviour. Research in Official Statistics, 2, 19–33.

Bosnjak, M., Tuten, T.L., and Bandilla, W. (2001). Participation in Web Surveys – A Typology. ZUMA Nachrichten, 48, 7–17.

Couper, M.P. (2000). Web Surveys: A Review of Issues and Approaches. Public Opinion Quarterly, 64, 464–494.

Lozar Manfreda, K. (2001). Web Survey Errors. Doctoral dissertation. Faculty of Social Sciences, University of Ljubljana, Ljubljana.

Lozar Manfreda, K., Biffignandi, S., Pratesi, M., and Vehovar, V. (2002). Participation in Telephone Pre-recruited Web Surveys: Who Comes Farthest? Paper presented at the 57th Annual AAPOR Conference, St. Petersburg Beach, Florida.

Vehovar, V., Batagelj, Z., Lozar Manfreda, K., and Zaletel, M. (2002). Nonresponse in Web Surveys. In Survey Nonresponse, R.M. Groves, D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds). New York: John Wiley and Sons.

Received November 2002
Revised March 2004