

## **List assisted Web Surveys: quality, timeliness and non-response in the steps of the participation flow**

Biffignandi S., Pratesi M.,  
University of Bergamo<sup>1</sup>, Italy

Lozar Manfreda K., Vehovar V.  
University of Ljubljana, Slovenia

**Abstract:** The debate on quality issues in Web surveys is open and lively. Data quality is required to satisfy user needs. Improving the survey process quality is a precondition to obtain product quality at an acceptable cost. This paper focuses on timeliness as a dimension of the quality of Web surveys. The authors propose to identify several steps of the Web survey process and to divide the final response rate into different components, one for each step. The aim of the research is to model the survival of the eligible respondents finding out which participants come farthest in the process of a Web survey. The analysis contributes to the efforts to reduce non-response and to shorten the individual survey period length.

**Keywords:** Web survey, Internet survey, online survey, survey process, withdrawals, survival, non-response, data quality, timeliness, participation.

### **1. Introduction**

The spread of technological culture and the penetration of Internet technology in the target population (individuals or establishments) affect several dimensions of the quality of the statistical design of Web surveys. Among these, timeliness of data collection is considered one of the most competitive advantages of the Internet with respect to the other survey modes.

In list assisted Web surveys<sup>2</sup> timeliness has several aspects to be investigated. We refer to the survey period length, the speed of the data capturing and processing tools and their influence on contact and response rate. The Internet could theoretically shorten the data collection period, but in practice the average time of response often amounts to several days and also the reaction to the e-mail invitation and follow ups results in a late first access to the questionnaire (Biffignandi and Pratesi, 2001, Bosnjak et al., 2001).

The process of participation in a list assisted Web survey starts with sending out e-mail or mail invitations to individuals from the list and ends with response or non-response of the members of the eligible population. Each member of the population has his/her own behavior after the invitation. The permanence in the survey process has to be modeled in order to find out which participants come farthest in the process of the Web survey and what steps of the survey participation process mostly contribute to the timeliness of data collection.

In this paper 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

---

<sup>1</sup> This paper is produced within the research activity of the University of Bergamo in the Murst 40% 1999 research grant on "Qualità delle statistiche economiche: le nuove strategie di formazione dei dati (Quality of economic statistical information: new data production strategies)", local unit scientific coordinator S. Biffignandi.

<sup>2</sup> By list assisted Web surveys we mean Web surveys where a list of individual units (either individuals or establishments) who were invited to participate in the survey exists. They are different from other types of Web surveys, i.e. from self-selected Web surveys where general invitations to the survey are placed usually on Web sites and anyone can access the Web questionnaire.

invitations, absorption of e-mails by the network), access to the questionnaire, and questionnaire completion step. The aim of this research is to model the permanence of eligible respondents in the survey process. In this way we can find out which participants come farthest in the process of Web survey thus contributing to its timeliness. The steps of the participation in the Web survey are illustrated in Section 2. The permanence in the survey process (survival) 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 of 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  $N$ , non-contact (NC) and contact (C) are the first results (possible events, outcomes) of the survey process. The non-contact leads to non participation in the survey, whereas the contact presents a further dichotomous possibility of participation or non participation. We distinguish four steps in the participation flow<sup>3</sup>:

1. *E-mail invitation*: E-mail invitations to the Web questionnaire are sent. Not all the intended recipients notice and read the invitation (see Table 1). Missed contacts can be caused by an error in the e-mail address (WE) (i.e., the e-mail is returned because of non-existent, unknown receiver) or due to 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*. This error can be minimized by checking the list before the survey starts. However, the recipient (individual or establishment) can change their e-mail address also after the survey had 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. *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 the introductory page may decide to proceed to the next stages. Not all those who access the introductory page proceed immediately to the completion. 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 contacted by the survey organization even if he/she only accessed but did not answer any questions (OE, see Table 1B). Other recipients limit their reaction to the sending of a return receipt of the e-mail invitation, without accessing the introductory page. Also these are considered as contacted by the survey organization (ORR, see Table 1B).
3. *Clicking "Start the questionnaire"*: Those who access the introductory page may decide to begin to complete it. The evidence 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

---

<sup>3</sup> 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 of the Web questionnaire is described in Bosnjak et al. (2001).

answering the survey questions: they simply start the questionnaire and quit it (OSB). Others proceed towards the completion of the questionnaire<sup>4</sup>.

4. *Completion of the questionnaire*: Those who begin to answer the questionnaire may decide to complete it (completed response event CR<sup>5</sup>) or to quit the questionnaire after having completed it only partially (partial response event, due either to voluntary or involuntary interruptions<sup>6</sup> PR). In both cases the evidence of both behaviours is written in log files after submit buttons at each questionnaire page are clicked (see Table 1A). In some cases the recipient explicitly refuse to participate in the survey sending an e-mail message to the survey organization asking to be excluded from the target population (ER see Table 1B).

Table 1. Event: contact (C)

TABLE 1A - EFFECT: PARTICIPATION IN THE SURVEY			
CAUSE	EVIDENCE	MEASURE	
Completed response (CR)	Pressing of submit buttons	Response rate	Contact rate
Interruption of the questionnaire, completion or partial completion, voluntary or involuntary interruptions (PR)	Pressing of submit buttons		

TABLE 1B - EFFECT: NON PARTICIPATION IN THE SURVEY			
CAUSE	EVIDENCE	MEASURE	
Only return receipt (ORR)	Return receipt	Non-response rate (which includes the refusal rate that refers only to explicit refusals)	Contact rate
Only entering the introductory page (OE)	Log file		
Only clicking the Start button (OSB)	Log file		
Explicit refusal statement (ER)	E-mail message		

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 to participate in the survey. At the end of the period devoted to the data collection (in many empirical situations a couple of months is needed) each candidate who reacts to the initial e-mail invitation or to the follow-ups passes through one of the described stages. At some time (immediately or after many days after the invitation) he/she withdraws from the survey process either after completing the questionnaire (full or partial completion), after only one or more accesses to the introductory page without proceeding further, or after only clicking the start button without answering any question.

There are members of the eligible population, whose e-mail address is absorbed by the net, but who do not react to the invitation. This group of people is composed of individuals with whom the 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 whole set of e-mail invitations that are not followed by a reaction is called NoR (see Table 2). In our context, the absence of reaction of the intended recipient is considered completely due to a missed contact. Under this assumption, the *non-reaction rate* plus the coverage rate make up the *non-contact rate* (see Table 2).

<sup>4</sup> We have to remember here 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.

<sup>5</sup> Here we intend those who through answering come the final part of the questionnaire, although they do not necessarily answer all the questions (some item non-response is possible). For detailed patterns of (non)response behavior see Bosnjak et al. (2001).

<sup>6</sup> Interruption is not necessarily voluntary. For example, there may be a technical problem at the survey organization side because of which respondents quit the questionnaire.

Table 2. Event: non-contact (NC)

EFFECT: NON PARTICIPATION IN THE SURVEY			
CAUSE		EVIDENCE	MEASURE
Wrong e-mail address (WE) or network problem (NE)		Message failed delivery	Coverage rate
Right e-mail address and no reaction (NoR)		Nothing	Non-reaction rate
			Non-contact rate

At the end of the planned survey period, the following rates give the picture of the level of participation obtained in the survey:

<b>Non-contact rate:</b>	$\frac{WE + NE + NoR}{NC + C} = \frac{WE + NE + NoR}{N}$
<b>Gross contact rate:</b>	$\frac{ORR + OE + OSB + ER + PR + CR}{NC + C} = \frac{C}{N}$
<b>Net contact rate:</b>	$\frac{ORR + OE + OSB + ER + PR + CR}{NC + C - WE - NE} = \frac{C}{N - WE - NE}$
<b>Response rate given the contact:</b>	$\frac{PR + CR}{C}$
<b>Response rate of the absorbed e-mails:</b>	$\frac{PR + CR}{NC + C - WE - NE}$
<b>Global response rate</b>	$\frac{PR + CR}{NC + C} = \frac{PR + CR}{N}$
<b>Non-response rate given the contact<sup>7</sup>:</b>	$\frac{ORR + OE + ER + OSB}{C}$

### 3. Modeling the permanence in the survey process

Timeliness of data collection depends on the distribution of the individual times of reaction and on the number of the withdrawals from each of the previous stages. Each candidate may decide to quit the survey process at one of the stages after the absorbed invitation or to proceed toward the next step till the full completion of the Web questionnaire. The individual reaction time varies among the candidates. In theory all the candidates could cooperate immediately avoiding the non-response problem and their permanence in the survey process (survival) could be as short as allowed by the communication times on the Internet.

In practice, on the other hand, some cooperates immediately, others respond only after many invitation messages; others do not react at all or limit their reaction to one or more visits to the Web questionnaire introductory page without answering any question.

Only a part of the eligible population reacts immediately to the invitation. Quick reaction means short individual survey period length and, in other words, short permanence (survival time) in the survey process and timeliness of data collection.

The reaction time is the result of both: the characteristics of participants and the survey conditions. By the latter we intend the absorption rate of the e-mail addresses and the time schedule of the follow up messages. It is clear that the timeliness of the data collection can be achieved by promoting a fast response of all members of the eligible population. In order to understand which conditions favor a short permanence in the survey process and encourage the participation in the survey, we model the duration of the permanence in the survey process specifying the dependence on individual covariates of the times of exiting from the survey process.

<sup>7</sup> This rate is defined under the assumption that NoR is due only to non contact with the eligible members of the population. Indeed in the NoR there is a quota of hidden implicit refusals that should be added to the numerator of the non response rate.

In our analysis the duration of permanence in the survey process is the time between the e-mail invitation and the withdrawal from the process. The withdrawal can be of three types: withdrawal due to full completion of the Web questionnaire, withdrawal due to partial completion of the Web questionnaire, withdrawal after only accessing the questionnaire without answering any questions. This last category includes those who only access to the introductory page of the questionnaire and those who also only click the Start button, but neither answer any question.

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

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

The dependence on covariates of the type specific hazard can be modelled through a general proportional hazard model for all three withdrawal types

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

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 withdrawal type. Also the dependence of the hazard on time  $\alpha(t)$  may vary across withdrawal type.

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

#### 4. The case study

Two telephone surveys from 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 address 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) collected 208 e-mail addresses by screening (the person within the household who most often uses the Internet was asked for) on a probability sample of Slovenian households. Altogether, 348 e-mail addresses were collected and on July 12, 2001 348 e-mail invitation were sent. 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 non-respondents on July 30, 2001 and the second on August 28, 2001. Because it was summer, time intervals between follow-ups were somewhat larger than usual in Internet surveys.

#### 4.1 The participation flow

The survey period (time interval between the first e-mail invitation and the last reaction) was equal to 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 3. For each step of the participation flow we indicate how many candidates withdraw and how many remain in the survey process. Table 4 contains the rates that are associated with the participation flow and are calculated at the end of the survey period. We call complete respondents the candidates who reached the final part of the questionnaire although they did not necessarily answer all the questions.

The percentage of undeliverable e-mails was about 15%: 84.48% of the invitation sent were absorbed by the network (*absorption rate* defined as absorbed e-mails on invitations sent). 72% of the absorbed e-mails were followed by a reaction. The recipients read and accessed the questionnaire, completed it partially or reached its final part (*net contact rate* defined as contacts on absorbed e-mails). The percentage of responses obtained from the absorbed e-mails was 59%: this rate includes also those who did not react to the absorbed message. The evidence of contact with eligible respondents was obtained in 213 cases: those who participated after the contact were 81% (*response rate given the contact*). The quota of respondents among all the invitations sent was lower and equal to 49% (*global response rate*).

Table 3: 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 = <b>348</b>				
Wrong e-mail=54				
Absorbed e-mails= <b>294</b>	<i>Withdrawals:</i> Access intro only =11			
	<i>Stay in:</i> 283	<i>Withdrawals:</i> Start quest only =29		
		<i>Stay in:</i> 254	<i>Withdrawals:</i> Partial comp.=50	
			<i>Stay in</i> 204	<i>Withdrawals:</i> Full comp.=123
				Non reactions= <b>81</b>

Table 4: Rates associated with the participation flow

Rate of absorbed e-mails	294/348	=0.8448
Net contact rate	(11+29+50+123)/294 213/294	=0.7245
Gross contact rate	(11+29+50+123)/348 213/348	=0.6121
Global response rate	(123+50)/348 173/348	=0.4971
	123/348	=0.3534
Response rate given the contact	173/213	=0.812
Response rate of the absorbed e-mails	173/294	=0.5884

#### 4.2 The permanence in the survey process

The behavior of the contacted individuals after the e-mail invitation determines the duration of their permanence in the survey process and how long they wait before the next stage of the survey

process. We identify three groups of contacted individuals in relation to the three ways of withdrawing from the survey process:

- 113 respondents withdraw after full completion of the Web questionnaire (Group 1),
- 50 limit their participation to the partial completion of the Web questionnaire (Group 2),
- 40 react to the invitation only accessing the questionnaire. In this last category those who only access to the introductory page of the questionnaire (11) and those who only click on the Start button (29) are gathered (Group 3).

The permanence of each group is studied with their survivor functions estimated by the life table method<sup>8</sup>. Each survivor function gives the probability of surviving, in other words the probability of remaining in the survey process (not withdrawing) beyond the time  $t$ . The origin of the time (0 point) is the date of the e-mail invitation. The permanence for the three groups is illustrated in Figure 1.

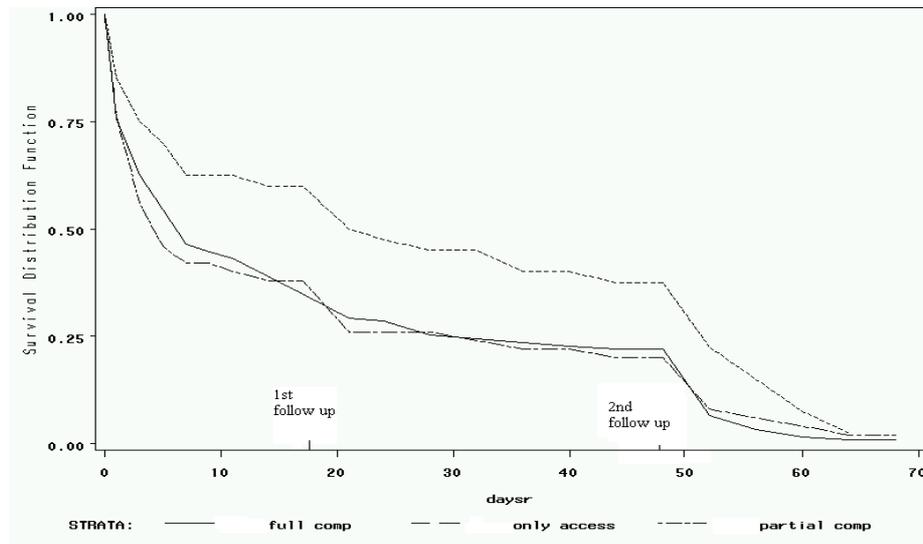


Figure 1: Withdrawals from the survey process

The probability of permanence decreases dramatically after sending out the e-mail invitations. Most reactions happen in 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 (in first 18 days). The others are obtained after the two follow-ups: the first sent after 18 days from the initial invitation, the second after 48 days. There are 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 after 2 days from the first and the second follow-up.

The hypothesis that the three survivor functions are equal is rejected<sup>9</sup>. In other words the three groups have different permanence behaviour in the survey. While respondents (partial and complete) might have similar determinants, withdrawals 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 compared with 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 permanence in the survey process (survival time) of those who limit their reaction to the access to the questionnaire (access to the introduction page only or also click the start button) is 26.95 days from the invitation (median 20.5, std dev. 23.58),

<sup>8</sup> The data are processed using the survival analysis procedures implemented by the SAS system (Allison P.D. 1998 ).

<sup>9</sup> The tests of equality of the survivor functions over the three groups are significant (Log Rank: Chi square 5.80 df 2, at  $p=0.0505$  level, Wilcoxon: Chi square 6.38 df 2  $p=0.04$ ).

while only 16.18 and 16.78 for partial and complete respondents (see Table 5). The longer permanence in the survey process is due also to their greater 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 6)

Table 5: Average time of permanence (survival) in the survey period

Type of Withdrawal	Average time			
	Mean	Median	Std dev	N
Only access	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 6: Average number of accesses by type of withdrawal

Type of Withdrawal	Average number of accesses			
	Mean	Median	Std dev	N
Only access	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 permanence in the survey process: their survival curves are virtually indistinguishable. People who participate in the survey answering the questionnaire items stay in the survey process less than the others, but the partial respondents are not faster than the complete respondents. On average, the respondents who withdraw from the survey process within two weeks either partially complete the questionnaire or come to the final part of the questionnaire. The average survival time for complete respondents is 16.78 days (std dev 20.34) versus the 16.18 days (std dev 21.06) for partial respondents. The distribution of responses is asymmetric with median permanence of respondents equal to 4 days as the larger part of responses is obtained before the first follow up.

Most full 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 at the first access<sup>10</sup>. The accesses done on different days are called separate accesses. 32,65% of the contacted people complete the questionnaire at the first access to the questionnaire and also the partial completion are done mostly at the first access (see Table 7).

<sup>10</sup> The respondent can fully or partially complete the questionnaire using only one session of work or he/she can browse the pages in several different sessions of work. However, in our case if several different sessions were used, the respondent should have started each time from the beginning, since the software used did not allow continuation where previously left. Each session has a starting time and date and an ending time and date. Some respondents do several sessions of work in the same day, others connect to the questionnaire in different days. Each session starts with the access to the questionnaire, in other words with the browsing of the questionnaire pages. The access could be limited to the introductory page or it could be a browsing of all the questionnaire pages. For simplicity, in this paper we call access the whole session of work. The first access is the first session of work, the second access is the second session of work, and so on. There were 68 second accesses and 57.35% of them were done in the same day of the first access. Third accesses were 32 and 46.98% of them were done in the same day of the first access.

Table 7: Withdrawals from the survey process by number of accesses to the questionnaire

Type of Withdrawal	Frequency		Percent	
Only access	Only first single access	25	8.51	13.61
	Separate accesses	15	5.10	
		40		
Partial completion	At first single access	39	13.27	17.01
	At separate accesses	11	3.74	
		50		
Full completion	At first single access	96	32.65	41.83
	At separate accesses	27	9.18	
		123		
No reaction				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. They may result in complete answer or some other withdrawal type. The first accesses were gathered after the dates of the first e-mail invitation and of the follow-ups.

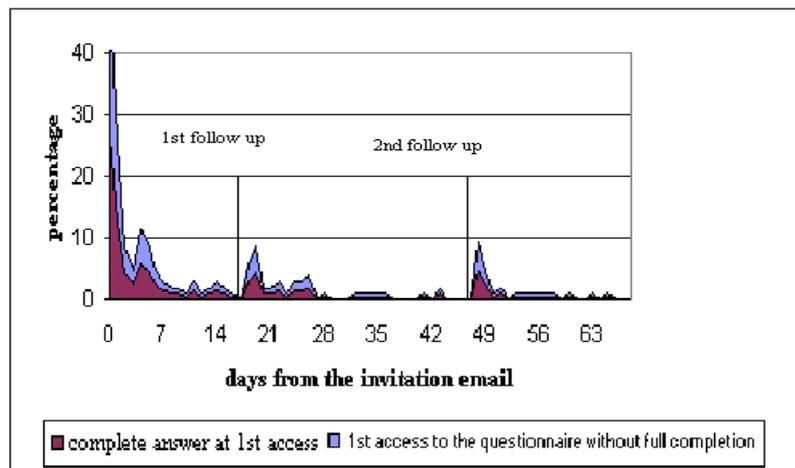


Figure 2: Distribution of first accesses in days from the first e-mail invitation

The average time before the first access is generally lower than the global permanence time (13.31 days the general mean, 6 days the median, 12.06 days the mean for the complete respondents, 4 days the median, see Table 8). The number of second accesses is not high (68 over 213 contacted) and among them there are some late full completions (mean time of the second access is 24.64 days, with median 24 days, see Table 9).

Table 8: Average time before the first access to the questionnaire

Type of Withdrawal	Average time			
	Mean	Median	Std dev	N
Only access	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 9: Average time before the second access to the questionnaire

Type of Withdrawal	Average time			
	Mean	Median	Std dev	N
Only access	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

As most of the reactions were obtained at the first access we decided to model the permanence of participants from the e-mail invitation to the first access, distinguishing the withdrawals at the first access from the withdrawals happening later in the survey period.

The effect of covariates on each type of specific hazard can be tested by fitting a Cox model (equation (2)) to each type of withdrawal. The goal is to test whether the effect of the covariates is the same or different across types of events (withdrawals) and to find out which participants come farthest in the process of the Web survey<sup>11</sup>.

The models included several variables collected at the pre-recruitment stage of the Web 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 categorisation 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 withdrawal at the first access or later, measured in hours, minutes and seconds.
- Place from where the Internet is accessed: home = 1, frequent access from home ( $n=118$ ); home=0, otherwise ( $n=95$ );
- Purpose of Internet usage: bus=1, use the Internet often or regularly for business purposes without considering the place of access ( $n=84$ ); bus=0, otherwise ( $n=129$ ); wbus=1, access from work and reason work ( $n=57$ ); wbus=0 otherwise ( $n=156$ );
- Age of the contacted person: ageyoung = 1, age between 21 and 40 years ( $n=137$ ); ageyoung=0 otherwise ( $n=76$ );
- Educational level of the contacted person: edu\_u=1 university level ( $n=77$ ); edu\_u=0 otherwise ( $n=136$ );

The adopted models are the following:

- Model 1 – *Late reactions*: The model regards the candidates who need several separate sessions of work for starting the questionnaire, or accessing the introductory page, or completing full or partially the questionnaire. They are considered as candidates who react late to the invitation. The model concerns the competing risk of having a late reaction.
- Model 2 – *Only accesses to the questionnaire*: The model refers to the candidates who only access to the questionnaire without completing (full or partially) it. The model concerns the competing risk of only accessing the questionnaire.
- Model 3 – *Fast respondents*: The model refers to the candidates who complete (full or partially) the questionnaire in only a single session of work. They are considered as fast respondents. The model concerns the competing risk of not being a fast respondent.

The strongest effect in Model 1 is the reason of Internet usage. People who use the Internet often and regularly for business purposes seem to postpone the participation in the survey. Internet users motivated by business are 2.215 times as likely as other users to exit later from the survey process.

<sup>11</sup> The fitted models have been evaluated doing a comparison of the models with standard tests (likelihood ratio test for nested models and Akaike information criterium for not nested models). The effect of the covariates has been tested with a forward selection procedure. All the fitting has been done by SAS procedures.

They therefore react later to the survey invitation, either by completing the questionnaire or only by accessing it.

Exits after having only accessed the questionnaire show a different pattern (see Model 2). Purpose of Internet use is not important. We see an effect of the place of access to the Internet and of the age of Internet users: the age of the contacted person yields a .70 decrease in the risk of exit after having only accessed the questionnaire. Also connecting from home yields a .50 decrease in the risk. The evidence is that people between 21 and 40 years old seem to be more willing to participate in the survey than the others and that, at the same time, people connecting to the net 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, Cybercaffes, other).

For not fast respondents, educational level and purpose of Internet usage are the only significant variables: contacted people with a university degree have a 43 percent lower risk of not leaving the survey immediately, while the risk is higher for Internet users for business purposes who connect to the Internet from their workplace (see Model 3). 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 in responding to the survey questions and seem to have a higher risk of being not fast respondents.

### Model 1: Later reactions versus others

Testing Global Null Hypothesis: BETA=0						
Test	Chi-Square	DF	Pr >	ChiSq		
Likelihood Ratio	8.1367	1	0.0043			
Score	8.6380	1	0.0033			
Wald	8.2114	1	0.0042			
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr >	Hazard Ratio
bus	1	0.79535	0.27755	8.2114	0.0042	2.215

### Model 2: Only accesses to the questionnaire versus others

Testing Global Null Hypothesis: BETA=0						
Test	Chi-Square	DF	Pr >	ChiSq		
Likelihood Ratio	10.3258	2	0.0057			
Score	10.5118	2	0.0052			
Wald	10.2670	2	0.0059			
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr >	Hazard Ratio
ageyoung	1	-1.01304	0.34107	8.8222	0.0030	0.363
home	1	-0.71037	0.34243	4.3035	0.0380	0.491

### Model 3: Not fast respondents versus others

Testing Global Null Hypothesis: BETA=0						
	Test	Chi-Square	DF	Pr > ChiSq		
	Likelihood Ratio	9.8435	2	0.0073		
	Score	10.0459	2	0.0066		
	Wald	9.9992	2	0.0067		
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
edu_u	1	-0.62016	0.28961	4.5853	0.0322	0.538
wbus	1	0.93554	0.30073	9.6774	0.0019	2.549

## 5 Concluding remarks

The method proposed to investigate the participation in Web surveys and timeliness of the data collection allows measurement of the times of reaction to the e-mail invitations and modeling the permanence in the survey process.

We have identified different levels of cooperation in the survey by distinguishing the steps of the survey process. After the telephone pre-recruitment (telephone interview, e-mail address collection) the following steps have been identified: e-mail recruitment (sending e-mail invitations to the Web survey, absorption of e-mails by the network), access to the questionnaire, and questionnaire completion step. 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 in several different sessions of work.

In our telephone pre-recruited Web survey, the final response rate has been divided into different components, one for each step of 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 72.45%, at net of the e-mail not absorbed by the network. The final response rate (response rate of the absorbed e-mails) was 58.84%. The response rate is higher if it is measured on contacts: we obtained the 81.2% of responses among the contacted people.

The length of the whole survey period was 68 days: from the day of sending the first invitation to the last recorded access to the questionnaire. The average time of response was about 16 days and most of the complete responses was obtained at 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 already said, most of the responses was 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. The largest part of the responses was also obtained before the first follow-up. This effect might be partially due also to telephone recruitment. The 348 Internet users who gave their e-mail address could have been informed by the telephone calls for the recruitment.

The permanence in the survey process of the eligible respondents has been modeled by comparing risk models in order to find out which participants come farthest in the process of Web survey, contributing to its timeliness. The 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 the others. Further, people who connect from work for business purposes tend to later than the others. It is more likely to proceed from the contact stage to the response stage for young users who connect from home.

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 the response to be obtained;
- timeliness is mainly achieved by promoting the first access to the questionnaire at the beginning of the survey period through a suitable timetable of email invitations .

The permanence 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 of the speed of reaction of eligible respondents.

## **References**

Biffignandi S, Pratesi M. (2001), "Internet surveys: the role of time in firms' response behavior", paper presented at the EKS-NTTS2001 Conference, Crete, 18-22 June 2001.

Biffignandi S., Pratesi M. (2000a) "Modeling firm response and contact probabilities in Web surveys", in: Proceedings of the Second International Conference on Establishment Surveys, Buffalo, New York, 17-20 June 2000.

Biffignandi S., 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.

Bosnjak M., Tuten T.L., Bandilla W. (2001) "Participation in Web surveys – A typology", in: ZUMA Nachrichten 48 pages 7-17.

Couper M. P. (2000) "Web surveys: a review of issues and approaches", Public Opinion Quarterly, 64,4, pages 464-494.

Groves R. M., Couper M. P. (1998), "Non-response in Household Interview surveys". John Wiley and sons, New York.

Lozar Manfreda, K. (2001), "Web survey errors". Doctoral dissertation. Faculty of Social Sciences, University of Ljubljana, Ljubljana.

Lozar Manfreda K., Biffignandi S., Pratesi M., Vehovar V. (2002), "Participation in telephone pre-recruited Web Surveys: Who comes farthest?", paper presented at the 57<sup>th</sup> Annual AAPOR Conference, St. Pete Beach Florida

Vehovar V., Batagelj Z., Lozar Manfreda K., Zaletel M. (2002), "Non-response in Web surveys", in: Groves R.M., Dillman D. A, Eltinge J. L., Little R.J.A. (eds), Survey Non-response. John Wiley and sons, New York.