

Internet surveys: the role of time in Italian firms' response behaviour

Silvia Biffignandi and Monica Pratesi

*Dipartimento di Matematica, Statistica, Informatica e Applicazioni
Piazza Rosate, 2 I-24129 Bergamo
silvia@unibg.it*

Keywords: Internet surveys, data quality, solicitation plan, response rates, survival analysis.

Abstract

The methodological framework and data quality in Internet surveys have not yet been completely investigated. This paper focuses on the role of the solicitation plan in the participation into the survey. The survey period length influences data timeliness; the period length is connected to the solicitation plan. Starting from this point of view, in our paper, the reaction to the soliciting message is analysed modelling the individual survey period length. The event of interest is the self-interview after the solicitation and the aim is to determine how the occurrence and timing of response depend on several covariates. It is proposed to use the competing risk model for the effect of the solicitation plan. This model is applied to the response process of a list-based Internet survey on Italian firms in five provinces of the Lombardy region.

1. Introduction

The Internet is a tool which entails great potential into the future of statistical surveys. The current literature discusses many of the multifaceted aspects of the design of e-mail and web surveys ⁽¹⁾, but both the methodological framework and the quality assessment of the collected data have not yet been completely defined (Batagelj et al. 1998b; Couper et al., 1999; Clayton and Werking, 2000).

Without entering the debate on the dimensions of quality and on the different points of view which guide the specific definitions ⁽²⁾, we stress that quality of statistical information merges both quality characteristics of the final product, the data, and of the production process and the statistical design of the survey. The diffusion of technological culture and the penetration of Internet technology in the population of the eligible respondents (households or firms) affect several dimensions of the quality of the statistical design of the survey. Among these, timeliness of data collection is considered one of the most competitive advantages of the Internet with respect to the other survey tools.

⁽¹⁾ Classifications of the different types of Internet surveys are illustrated in Batagelj et al. 1998a, Couper 2000, Biffignandi and Pratesi 2000a, Bosnjak and Tuten, 2001.

⁽²⁾ Many publications, conferences and working groups have been organised on the topic. As examples, we quote the European LEG (Leadership Expert Group) on Quality and the International Conference on Quality in Official Statistics, Stockholm, 14–15 May 2001.

Timeliness is a characteristic both of the final product and of the production process ⁽³⁾. In the first case, it concerns the frequency or periodicity of the data ⁽⁴⁾; in the second case, it is connected to the survey procedures, particularly to the definition of the survey period length, the speed of the data capturing and processing tools. It should be observed that accuracy of data often comes into conflict with timeliness. A too short length of the survey period may lead to an increase in non-response rate, an excessive speed in data processing and capturing can increase the risk of errors and mistakes.

Timeliness of communication allowed by the Internet could theoretically permit compression of the data collection period. Nevertheless, in empirical investigations the response times are different from zero, they are heterogeneous among the contacted subjects. In order to fit the traditional survey research guidelines to the new context connected to this survey tool, further investigations on the role of the time in the response behaviour are required. This paper focuses on the effects of the solicitation plan in surveys carried out on firms. The solicitation plan is a crucial procedure for improving the response rates in personal, telephone and mail surveys; it is reasonable to assume that such is the case for e-mail surveys as well ⁽⁵⁾. The structure of the solicitation plan affects the survey period length; therefore, it is a crucial aspect for the timeliness dimension of the survey data quality. Given the new environment within which the self-interview process is developed, we hypothesise that the behaviour of the respondents during the survey period can be affected by some covariates. Modelling the respondents behaviour could help in the definition of the best practices for the procedures of the e-mail or web surveys planning.

The main reasons that motivate the survival analysis of the response process in the Internet survey are illustrated in Section 2. Section 3 is devoted to the specification of a competing risk model to study the effect of the solicitation plan. The main steps of the data collection process of the list-based Internet survey on 'technological communication and links among enterprises' are illustrated in Section 4. In this section the results of the application of the proposed models are also described.

2. The time in the response process in web surveys

Various procedures are applied in surveys that are carried out via the Internet. With reference to the time dimension of the response behaviour, the main distinction is connected to the contact strategy, which is adopted. The main classification is the distinction between contact achieved through the personalised addressing of the message

⁽³⁾ This classification is quoted in the Encyclopedia of Statistical Sciences, Quality Concept for Official Statistics (Elvers, Rosén, 1997).

⁽⁴⁾ Frequency refers to statistical data, which are produced according to a regular scheme; it includes reference time frequency, i.e. the periodicity of reference times for published statistics, dissemination frequency, i.e. the periodicity with which statistics are made public.

⁽⁵⁾ Note that in some e-mail and web surveys where a questionnaire has been sent only once, a poor response rate has been obtained and interpreted as a failure of the survey tool. Experiments with solicitation plans have proved that the same response rate as the one obtained in the corresponding mail survey can be achieved; as regards timeliness it has been shown that the bulk of response were collected more quickly than the mail survey (Schaefer and Dillman, 1998, Dillman, 1999). Metha and Sividas (1995) and Smith (1997) compared a single contact e-mail survey with multiple contacts; both papers showed gains in response rates.

and contact achieved indirectly via the 'navigational path', i.e. the contact is achieved when the interview is submitted to a person before entering a site or before registration with a group. In this paper we focus on the personalised contact strategy with a solicitation procedure. We refer to personalised contact via e-mail. Within a personalised contact procedure, different interviewing procedures are possible: a questionnaire as an attachment to the invitation e-mail or a web questionnaire; our empirical analysis refers to a web questionnaire.

The behaviour of the respondent after the e-mail of invitation is therefore, in our opinion, the new perspective in the definition of the survey period. The whole survey period is the period between the first e-mail message and the last self-interview. The individual length of the survey period depends on many factors and its length is also the result of the firms' reactivity to the solicitation plan.

Given the response of the firm, the attention is on how long the period is T_i between the e-mail of invitation and the self-interview. The firms who participate in the survey can complete the form in a few days from the first e-mail of invitation or can complete the form after some solicitation messages, at the end of the survey period. In this context, it is important to explore the reactivity of the respondents to the messages sent by the survey organisation: communication with the firm is on-line and it is possible to send many soliciting messages (stimuli) to encourage participation in the survey.

The effect of the solicitation messages can be explored through survival analysis of responses (Cox and Oakes, 1984). The event of interest is the self-interview after the solicitation and the aim is to determine how the occurrence and timing of response depend on several covariates. This process in e-mail and web surveys has three features that are difficult to handle with conventional statistical methods: importance of the timing of response, censoring, and time-dependent variables.

The analysis of the behaviour of the respondents could be done performing a logistic regression with a dichotomous dependent variable: response or non-response after a stimulus; however this analysis ignores the timing of response. It is natural to suppose that people who respond soon have, on average, a higher propensity to complete the self-interview than those who do not participate until the third or fourth week of the survey period after having been solicited many times.

One solution to the problem is to consider the time before the self-interview as a dependent variable in a conventional linear regression model. However the problem of censored cases still remains open. The problem is complex. A first simplifying assumption is that censored cases are members of the target population, which do not react before the end of the survey period. In order to estimate the conventional linear regression model, we could discard these cases, but in web-based surveys the non-response rate is often relevant ⁽⁶⁾.

⁽⁶⁾ A classification of the possible outcome of the invitation to participate in a web survey is in Biffignandi, Pratesi, (2000a).

In conventional methods it is not clear how time dependent variables can be appropriately incorporated into either the logit model for the occurrence of response or the linear model for the timing of response. These covariates are conventionally the individual characteristics of the respondents (in the case of survey on population of firms: legal form, size of the firm, age of the firm, sector of economic activity) and also variables reflecting the contact plan designed by the survey organisation (in the case of list-based web surveys: number and timing of the solicitation messages). Some of these covariates remain constant over the survey period; others could change at any time during the follow-up.

2. A competing risk model for the effect of the solicitation plan

Generally, all the self-interviews were treated as though they were identical. To evaluate the effect of the solicitation plan, it is essential to treat the self-interview due to the first solicitation message differently from self-interviews due to the second or third solicitation message. The competing risk to react after the first or the other solicitation messages can be handled under the hypotheses that the occurrence of one type of event removes the individual from risk of all the other event types. Firms who complete the self-interview after the first solicitation message are no longer at risk of completing after the second solicitation message. Firms who complete the web questionnaire after the second solicitation message are no longer at risk of completing after the third (and last) solicitation message. In other words a firm who completes after the first solicitation at time t_1 would have later compiled at time t_2 after the second solicitation if the compilation has not occurred in time t_1 .

Let T_i be a random variable denoting the time of the self-interview for firm i and let J_i a random variable denoting the type of self-interview (type 1: self-interview after 1 solicitation, type 2: self-interview after 2 solicitations, type 3: self-interview after 3 solicitations). Thus $J_i = 2$ means that the firm completes after the second soliciting message. The type specific hazard of self-interview at time t for firm i , is defined by

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

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

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

where $x_i(t)$ is a vector of covariates, some of which may vary with time, and the coefficient vector b is subscripted to indicate that the effect of the covariates may be

different for different self-interview type. Also the dependence of the hazard on time $a(t)$ may vary across self-interview 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 firms which are at a particularly high (or low) risk of one event type (e.g. completing the questionnaire after the first solicitation) are not more or (less) likely to experience other kind of events.

3. The case study

3.1. Data collection process and research procedures

The proposed model is applied to the results of a list-based web survey on 'technological communication and links among enterprises' carried out in five provinces of the Lombardy region in Italy. The questionnaire has been kept simple and is based mainly on qualitative answers and some percentage data. The interviewed firms were asked about their use of e-commerce, their collaboration with other enterprises and/or belonging to groups, markets, and employment. The questionnaire consisted of eight pages (six pages of 40 substantial items, one welcome page and one final page).

Our list-based web survey was extended to the whole population of firms who gave their e-mail address to the Chamber of Commerce of the five provinces (Brescia, Lecco, Varese, Cremona and Mantova) of the Lombardy region. This target population was composed of 2 047 firms and could be roughly considered as the population of Internet users firms. We sent 2 047 e-mail invitations. Theoretically, the e-mail addresses of the target population members should have been updated and well maintained, in practice the network-absorbed e-mails made up only 1 790 (12 % of explicitly wrong e-mail addresses).

The data collection was carried out according to the following steps.

- The first invitation to participate in the survey was an e-mail sent (survey presentation letter) to each firm in the e-mail list at the beginning of the survey period. In order to encourage participation, incentives were offered (survey report and other connected advantages). The discussion on the opportunity of promotion and incentives for cooperation in the first e-mail to reduce the non-response rate has been the topic of many contributions. Batagelj et al. (1998) have studied the effects of incentives on responses.
- The access procedure to the questionnaire compilation was simple. Each firm has received its own address for compilation; no identification code (id) and no password were required. The duplication of responses was not possible in our web survey because in the invitation e-mail sent to the firms we indicated the personalised URL address where the respondent could find the questionnaire. Clicking on the address the respondent could access the questionnaire. It has been planned to record all the subsequent accesses of each firm and the day, hour and duration of each access.

- During the survey period three soliciting e-mail messages were sent. Soliciting e-mail messages have been sent weekly. The first soliciting message was sent 14 days after the survey follow-up.

We made contact with 1 120 firms: 559 responded or refused, 561 only accessed the questionnaire. The remaining 927 firms were not contacted ⁽⁷⁾. The results of the contact and response process are summarised in the following table:

1.1.1.1
1.1.1.2 *Contacted and not contacted firms*

	Complete responses	442
	Explicit refusals	117
	Visits to the web questionnaire	561
Contacts		1120
	Wrong e-mails	257
	No reactions ⁽⁸⁾	670
No contacts		927
Total target population		2047

We obtained 442 completed interviews and 117 explicit refusals. By calculating the contact and response rates according to the criteria proposed in Biffignandi and Pratesi (2000a, 2000b) the results are: gross contact rate 54.7 % (1 120/2 047); net contact rate 56.8 % (1 120/1 790); response rate given the contact 39.5 % (442/1 120), overall response rate 21.9 % (442/2 047).

We decided the periodicity and the number of the solicitation messages on the basis both of the empirical evidence obtained from the pilot survey and of practical sense. We found out in the pilot survey (same list-based web survey design on the about 500 Internet user firms of the province of Bergamo, see Biffignandi and Pratesi 2000b) that the reaction to the invitation e-mail was prompt (two working days from the sending). We decided on the basis of good sense that two weeks were a reasonable period before the first solicitation. In order to shorten the survey period length the other two solicitations were sent once a week. A week should have been enough given the prompt reaction of the firms to the e-mail messages. This solicitation plan worked well in the pilot survey and we decided to repeat it in the other five provinces ⁽⁹⁾.

The final aim of the research is to find out which individual features are related to the hazard of responding after 0, 1, 2, 3 solicitations. The findings are useful because they indicate which respondent categories are worth being solicited and how many times it is useful to capture their attention with a solicitation message.

⁽⁷⁾ For instance, some firms could not be reached as e-mail is not currently used; some firms could have ceased their activity.

⁽⁸⁾ No visiting of the questionnaire has been considered as a rough indication of no reaction: 670 cases.

⁽⁹⁾ Actually, a fourth soliciting e-mail was sent (excluding the pilot province) after the end of the survey period (three months after the first e-mail message). This last soliciting message aims to remind the firm to fill in the questionnaire or to react in order to interrupt the surveying process. The primary objective of the last soliciting e-mail has been the investigation of the characteristics of the firms, which had no reaction to the first e-mail (i.e. no questionnaire compilation, no explicit refusal, and no effectively detected wrong e-mail address). The result of this last soliciting message was not encouraging: only 10 firms reacted declaring that they wanted to be excluded by our mailing list for surveys. They were included in the explicit refusals category. This was the reason why we chose only three exit modes in our analysis.

We limit our analysis to the individual response time of the 442 firms who completed the self-interview. The aim is not to make inference on the behavior of the population of Internet users. We don't have a random sample of firms but a self-selected sample of respondents. We did not adjust the estimates to take into account individual non-response: we considered the 442 firms simply as a set of web respondents with the aim of exploring their reaction to the solicitation plan.

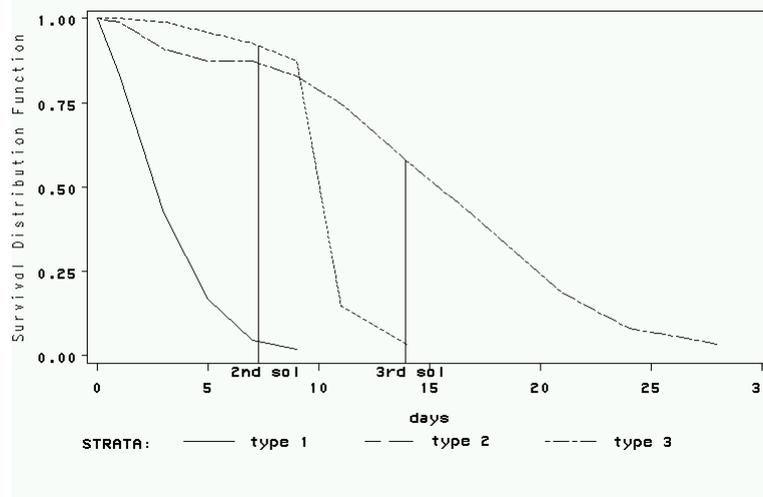
3.2. The survival function

In this study the mode of exit for self-interview of a firm was classified into three categories:

1. self-interview after the first solicitation sent after 14 days from the e-mail of invitation (type 1)
2. self-interview after the second solicitation sent after 21 days from the e-mail of invitation (type 2)
3. self-interview after third solicitation sent after 28 days from the e-mail of invitation (type 3)

The survivor functions of each group are estimated by the life table method⁽¹⁰⁾. Each survivor function gives the probability of surviving, in other words the probability of not completing the self-interview, beyond the time t. The origin of the time (0 point) is the date of the first solicitation message. The solicitation dates are used as a proxy of the contact time. The process of self-interview for the three event types is illustrated in Figure 1⁽¹¹⁾.

Figure 1: Survival distribution function of three types of self-interview



⁽¹⁰⁾ All the processing was carried out using the survival analysis procedures implemented by the SAS system (Allison 1998).

⁽¹¹⁾ If all the solicitation messages had followed the calendar rule the survival distribution functions for the second and third type of self-interview should have started from 1. This is not the case because there are seven second solicitation messages and 37 third solicitation messages sent before the planned date. This was due to a misunderstanding during the survey planning.

Firms have a prompt reaction to the first and second solicitation messages: the probability of surviving decreases dramatically after two days from the first solicitation and the two solicitations. The firms solicited three times live longer than the others and seem to be less reactive to the solicitation stimuli.

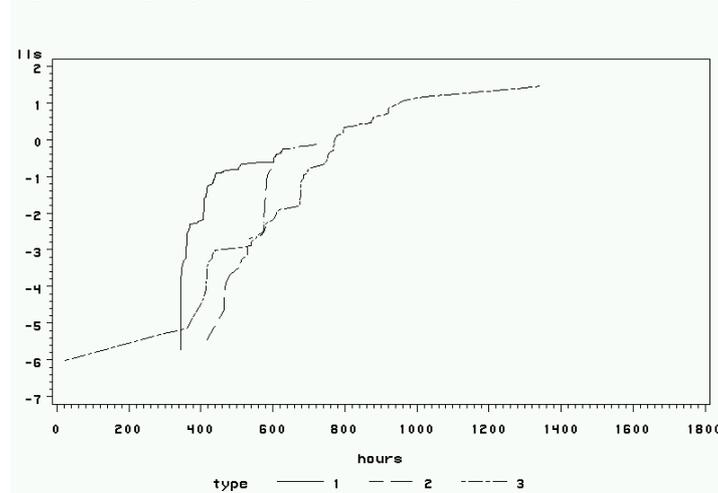
The hypothesis that the three survivor functions are equal is rejected ⁽¹²⁾, in other words the three groups have different survival behaviour. While exits after one or two solicitations might have similar determinants, exits after three solicitations can be considered a distinct phenomenon.

3.3. The time specific hazard functions

The time specific hazard functions are not the same for all event types. Figure 2 shows the behaviour of each log-log survivor function $\log(-\log(S_{ij}(t)))$ where $S_{ij}(t)$ is the type specific survival function at time t for firm i . The time is measured in hours from the sending (receiving) of the e-mail of invitation to participate in the survey. If the hazard is proportional, the log-log survivor functions should be parallel.

The curves for types 1 and 2 exits are virtually similar. The curve for type 3 exit is lower than the other two curves and there is evidence for the last curve to move closer to the other two in later times, which is an evidence of non-proportionality. The hypothesis of proportional hazard can be tested with a multinomial logit model for the relation between the type specific hazard and time. The results of Table 1 show that the effect of time is highly significant, indicating a rejection of the proportionality assumption.

Figure 2: Log-log survival plot for the three types of self-interview



Moreover parameter 3 is the coefficient of the contrast between type 1 (exit after 1 solicitation) and type 3 (exit after three solicitations): type 1 hazard decreases much more

⁽¹²⁾ The tests of equality of the survivor functions over the three groups are significant at $p < 0.0001$ level (Log Rank: Chi square 323.0337 df 2, Wilcoxon: Chi square 254.5206 df 2).

rapidly with time than the hazard for type 3 event. The same conclusion made on the basis of parameter 4 for the contrast between types 2 and 3. Hazards for types 1 and 2 exits decrease much more rapidly with time than the hazard for type 3 exit. There is evidence that the hazard for self-interview after three solicitation messages is lower than for the other two exit types (even if the behaviour is not the same in the whole period, see Figure 1), and that it increases less rapidly with time ⁽¹³⁾.

Table 1: Multinomial model: maximum likelihood analysis of variance

Source	DF	Chi-Square	Pr> ChiSq
Intercept	2	110.72	<.0001
Time	2	104.75	<.0001
Likelihood Ratio	610	386.65	1.0000

Table 2: Multinomial model: analysis of maximum likelihood estimates

Effect	Parameter	Estimate	Standard Error	Chi-Square	Pr>ChiSq
Intercept	1	17.0835	1.6345	109.25	<.0001
	2	5.5694	1.0531	27.97	<.0001
Time	3	-0.0314	0.00307	104.75	<.0001
	4	-0.00860	0.00165	27.14	<.0001

3.4. Covariate effects on the hazard functions

The effect of covariates on each type-specific hazard can be tested by fitting a Cox model (equation (2)) to each type. The goal is to test whether the effect of the covariates is the same or different across event types.

The model included the following variables.

- Time (ore): time spent before the self-interview, measured in hours, minutes and seconds from the sending (receiving) of the invitation to participate in the survey.
- Solicitations (s): s=0, no solicitation messages, s=1, self-interview after the 1st solicitation message, s=2, self-interview after the 2nd solicitation message, s=3, self-interview after the 3rd solicitation message.
- Age (eta): age of the firm in years at the time of the survey.
- Size of the firm (small): size of the firm in term of employees: small=1, number of employees greater or equal to 20, small=0 otherwise.
- Legal form of the firm (di): di = 1 if the firm is run by only the entrepreneur, di=0 otherwise.
- Frequency of Internet connections (fuicf): frequency of Internet use for communication with clients and suppliers fuicf=1, every day connection in Internet, fuicf=0 otherwise.
- E-commerce (q6ind): q6ind=1 if the firm practice e-commerce, q6ind=0 otherwise.

The following tables show the cross tabulation of the number of solicitations and the independent variables included in the model.

⁽¹³⁾ The solicitations do not occur at the same time of the day: this explains the time shift between the three curves.

Independent variables indicators by number of solicitations

Solicitations	Q6ind		fuicf		small	
	0	1	0	1	0	1
0 (120)	36	84	53	67	59	61
1 (124)	37	87	46	78	74	50
2 (95)	37	58	41	54	61	34
3 (88)	38	50	34	54	43	45
Total (n=427) ⁽¹⁴⁾	148	279	174	253	237	190

Solicitations	eta			di	
	1-10	11-20	21+	0	1
0 (120)	29	33	58	104	16
1 (124)	28	39	57	110	14
2 (95)	19	30	46	88	7
3 (88)	23	28	37	80	8
Total (n=427)	99	130	198	382	45

The time when the invitation e-mail is sent is considered proxy of the time when the recipient notices and reads the e-mail message. This time varies among the four different groups of the previous table.

1.1.1.3 Average individual response time (days)

Solicitations	Mean	Median	Mode	Std dev
0 (120)	3.68	2.50	1	3.99
1 (124)	17.39	17	17	2.20
2 (95)	24.26	24	24	1.95
3 (88)	29.65	31	28	8.77
Total (n=427)	17.59	18	17	10.81

The fitted models are the following.

- Base model: this model treats all event types the same. The model is included for comparison; here the censoring is on firms who completed the self-interview with at least one solicitation.
- Model 1: this is the model for type 1 exit (self-interview after the first solicitation), types 2 and 3 are treated as censoring.
- Model 2: this is the model for type 2 exit (self-interview after the second solicitation), types 1 and 3 are treated as censoring.
- Model 3: this is the model for type 3 exit (self-interview after the third solicitation), types 2 and 1 are treated as censoring.

There is an implicit stratification by province: the data were collected in five different provinces of the Lombardy region.

⁽¹⁴⁾ The data set size reduces to 427 observations because we restrict the analysis only to the firms with no missing values in the independent variables.

The proportional hazards assumption may not be realistic for all the provinces. So we have performed a stratified analysis allowing that the base risk changes in each province. In the application of the competing risk models we have proceeded as follows: the covariate effect on the likelihood of the model have been tested with the likelihood ratio test; 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 non-nested models). The effect of the covariates has been tested with a forward selection procedure. All the fittings have been done by SAS procedures for survival analysis. In the following we report and discuss only the effect of the covariates on the hazard of responding after 0, 1, 2, 3 solicitations. Table 2 shows what happens when no distinction is made among the different kinds of exits (Base model).

Table 2: Analysis of exits after at least a solicitation (Base model)

Testing global null hypothesis: BETA=0						
	Test	Chi-square	DF	Pr > Chi-sq		
	Likelihood ratio	7.6784	3	0.0531		
	Score	7.5728	3	0.0557		
	Wald	7.5376	3	0.0566		
Analysis of maximum likelihood estimates						
Variable	DF	Parameter estimate	Standard error	Chi-square	Pr > Chi-sq	Hazard ratio
SMALL	1	-0.20013	0.12179	2.7001	0.1003	0.819
fuicf	1	0.22183	0.12493	3.1527	0.0758	1.248
q6ind	1	-0.14167	0.12290	1.3286	0.2490	0.868

The strongest effect is the frequency of the Internet connection: the effect of a solicitation seems to be higher for firms which are used to checking frequently their mail boxes. Frequent Internet users are 1.2 times as likely as other firms to exit by the effect of at least one solicitation message.

Table 3: Model 2: Analysis of exits after two solicitations

Testing global null hypothesis: BETA=0						
	Test	Chi-square	DF	Pr > Chi-sq		
	Likelihood ratio	7.5699	3	0.0558		
	Score	7.4423	3	0.0591		
	Wald	7.2684	3	0.0638		
Analysis of maximum likelihood estimates						
Variable	DF	Parameter estimate	Standard error	Chi-square	Pr > Chi-sq	Hazard ratio
SMALL	1	-0.54451	0.21893	6.1857	0.0129	0.580
fuicf	1	0.08926	0.21794	0.1677	0.6821	1.093
q6ind	1	-0.16730	0.21553	0.6025	0.4376	0.846

Exits after two solicitations (Model 2) show a different pattern. Frequency of Internet use for communication with clients and suppliers is not important, nor is the practice of e-commerce. We see an effect of the size of the firm in terms of employees (see Table 3): the small size of the firm yields a.52 decrease in the risk of exit after two solicitations.

For exits after three solicitations (Model 3, see Table 4), e-commerce is the only significant variable: firms who practise e-commerce have a 66 % lower risk of leaving the survey after three solicitation messages. The picture changes when we focus on the hazard for exits after one solicitation (Model 1): the previous model is not significant (see Table 5).

Table 4: Model 3: Analysis of exits after three solicitations

Testing global null hypothesis: BETA=0						
	Test	Chi-square	DF	Pr >	Chi-sq	
	Likelihood ratio	14.6499	3	0.0021		
	Score	15.6206	3	0.0014		
	Wald	14.9201	3	0.0019		
Analysis of Maximum Likelihood Estimates						
Variable	DF	Parameter estimate	Standard error	Chi-square	Pr > Chi-sq	Hazard ratio
SMALL	1	0.11694	0.23369	0.2504	0.6168	1.124
fuicf	1	0.39719	0.25775	2.3746	0.1233	1.488
q6ind	1	-0.89506	0.23952	13.9645	0.0002	0.409

Table 5: Model 1: Analysis of exits after one solicitation

Testing global null hypothesis: BETA=0						
	Test	Chi-square	DF	Pr >	Chi-sq	
	Likelihood ratio	5.8430	3	0.1195		
	Score	5.7221	3	0.1259		
	Wald	5.6728	3	0.1287		
Analysis of maximum likelihood estimates						
Variable	DF	Parameter estimate	Standard error	Chi-square	Pr > Chi-sq	Hazard ratio
SMALL	1	-0.10567	0.18992	0.3096	0.5779	0.900
fuicf	1	0.27305	0.19134	2.0363	0.1536	1.314
q6ind	1	0.33820	0.19838	2.9064	0.0882	1.402

4. Concluding remarks

The method proposed to investigate the respondent behavior in web surveys and timeliness of the data collection has allowed us to measure the times of reaction to the solicitation e-mails and to model the permanence in the survey process.

The timeliness characteristics of the list-based web surveys are largely due to the ability of the survey organisation to capture the attention of the respondent. The solicitation e-mails seem to be a good stimulus. The permanence (survival) in the survey process of the respondents is affected by the solicitation messages. It should be observed that the effect of the stimulus is active over a short period of time and that the empirical evidence is that after three stimuli the effect is decreasing, after four stimuli the effect is irrelevant.

The time of reaction to the each stimulus is generally short (no more than two days) with a differential behaviour depending on the firm characteristics. The frequent Internet user for business purposes is more likely than the others to need at least one solicitation. The size of the firms seems related to the risk of exiting after two solicitations and the practise of e-commerce reduces the risk of being solicited three times.

Small firms have a decreased risk of responding after two solicitations compared with big firms. This is coherent with other results obtained for the same survey. We have had evidence (see contributions 4 and 5) that small firms (less than 20 employees) tend to react sooner than the others to electronic mailing. Our hypothesis of explanation, suitable for Italian employees, is that only the entrepreneur often runs small firms and he personally checks the electronic mail and decides about his participation in surveys. This might not be the case in greater firms where staff members not in charge of the participation in surveys may check the e-mail address. In this case the message might be redirected many times before the competent respondent notices and reads it. The practice of e-commerce, proxy of the high level of Internet usage of the firm, tends to make the follow-up with solicitation e-mails less relevant in order to capture the attention of the firm.

Taking into account the behavioural results, it appears that gains in timeliness of data collection process results could be achieved by a good balancing of the reaction time to stimulus and stimulus periodicity. Modelling the response times with the tools of survival analysis is crucial in order to find out which respondent needs to be solicited many times but further empirical evidence is required in order to personalise the follow-up according to the characteristics of the respondent.

5. Acknowledgement

This paper describes part of the results of the research activity of the local research unit 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)' and of the activity supported by the 2000 research grant of the University of Bergamo 'Modelli statistici per la probabilità di contatto e di risposta nelle indagini web' (Statistical models for the probability of contact and response in web surveys).

6. References

- [1] Allison P.D., 1998, Survival analysis using the SAS system. A practical guide, SAS Institute.

- [2] Batagelj Z., Lozar K. and Vehovar V., 1998a, *Who are non-respondents in web surveys?*, Paper presented at the 9th International Workshop on household survey non-response, Bled, September 1998.
- [3] Batagelj Z., Lozar K. and Vehovar V., 1998b, *Technical and methodological issues in WWW surveys*, Meeting of the American Association for Public Opinion Research, St. Louis, May.
- [4] Biffignandi S., Pratesi M. (2000a), 'Le indagini sulle imprese: aspetti metodologici e un'analisi degli intervistati', in *Tecnologie informatiche e fonti amministrative nella formazione dei dati economici*, C. Filippucci ed., F. Angeli, Milano.
- [5] Biffignandi S., Pratesi M., 2000b, *Modelling firm response and contact probabilities in web surveys*, International conference on establishment surveys, Buffalo NY, June 2000, Proceedings.
- [6] Biffignandi S., Pratesi M., 2002, The respondent's profile in a web survey on firms in Italy, *Metodoloski zvezki*, n.18 Ljubljana, Slovenia.
- [7] Bosnjak M., Tuten T.L., 2001, 'Classifying response behaviour in web-based surveys', *Journal of Computer –Mediated Communication*, 6 (3).
- [8] Cox D. R. Oakes D., 1984, *Analysis of survival data*, Chapman & Hall, London.
- [9] Clayton R.C., Werking G.S., 2000, 'Business surveys of the future: the world wide web as a data collection methodology', in Couper M. et al., *Computer Survey Information Collection*, J. Wiley, New York.
- [10] Couper M.P., Blair J., Triplett T., 1999, 'A comparison of mail and e-mail for a survey of employees', *Journal of Official Statistics*, vol. 15, n. 1.
- [11] Couper M.P., 2000, *Web surveys. A review and guide*, Public Opinion Quarterly, 64, 4.
- [12] Dillman D.A., 1999, 'Mail and other self-administered surveys in the 21st Century: the beginning of a new era', revisited version in *Gallup Research Journal*.
- [13] Elvers E., Rosén B., 1997, *Encyclopaedia of statistical sciences, Quality Concepts for Official Statistics*, Wiley & Sons.
- [14] Metha R., Sividas E., 1995, 'Comparing response rates and response content in mail versus electronic mail surveys', *Journal of Marketing Research Society*, vol. 37.

- [15] Schaefer D., Dillman D.A., 1998, 'Development of a standard e-mail survey methodology: results of an experiment', Paper presented at the American Association for Public Opinion Research, revised version available from *Public Opinion Quarterly*, 62, 378–397.
- [16] Smith C., 1997, 'Casting the net: surveying an Internet population', *Journal of Communication -Mediated Communication* 3 (1).