The Transition from University to Work: Web Survey Process Quality

Claudio Quintano, Rosalia Castellano, and Antonella D'Agostino[1]

Abstract

The biggest advantage of the web survey method is that in many cases it is easier, faster and cheaper than any conventional methods. However, quality issues in web surveys are of interest because improving the quality of the survey process in terms of timeliness, response rate and accuracy is a basic requirement. The aim of this paper is to analyse several dimensions of quality of data collection through the web, adopting a mixed-mode contact and follow-up plan using both telephone and e-mail. In this way, we can compare differences between e-mail and telephone contact modes in response patterns on the occasion of both the first contact and the call-backs. Different quality indicators referring to data collection timeliness, the relationship between structural variables and quality of data collection process are computed. The data used on this empirical investigation concern university-to-work transition – an issue of great relevance in Italy considering its high levels of youth unemployment.

1 Introduction

Web survey is a relatively new method for collecting data in social research and its use has been growing because it is in many cases easier, faster and cheaper than any conventional methods. However, quality issues in social web surveys have to be investigated in-dept in order to improve quality of the survey process in terms of response rate, timeliness and accuracy. In particular, the contact modes and their combination need to be studied in order to know the proper strategy to attract co-operation in a web survey.

[1] Department of Statistics and Mathematics for Economic Research, University of Naples "Parthenope", Via Medina 40, 80133 Naples, Italy; claudio.quintano@uniparthenope.it, rosalia.castellano@uniparthenope.it, antonella.dagostino@uniparthenope.it.

In the last few years, a great deal of social empirical researches by using web surveys has been conducted and published with the objective to assess data quality resulting from the web survey mode (http://www.websm.org). Among others, Couper (2000) discusses survey errors and sampling approaches in web surveys, Quintano et al. (2000) suggest an electronic questionnaire for an Italian survey of ISTAT on value-added provisional estimate; Klassen and Jacobs (2001) report comparisons among three survey technologies; Vehovar et al. (2002) give a significant picture of validation problems related to web surveys and describe in detail the non response process; Biffignandi and Pratesi (2002) examine the timeliness in an internet survey on firms; and Healey et al. (2005) present an empirical evaluation of three web survey design principles. The discussion on methodological problems in web surveys is an open one and concerns different aspects of the survey process.

Our purpose in this paper is to examine different aspects of the quality of the web survey process when a mixed-mode contact is used both for the first contact and for the follow-up. The main goal of this study is to answer the practical question about whether and to what extent the mechanism underlying the graduates' response behaviour in terms of the response rate and timeliness depends on the data collection mode used for the first contact.

After a brief discussion of surveys about university-to-work transition (Section 2), the paper presents the design of a survey on graduates in Economics at the University of Naples Parthenope (Section 3). Then, the participation process in terms of exogenous and endogenous factors influencing graduates cooperation in the web survey is considered by: i) the definition and the computation of some outcome rates as quality indicators at different analysis steps (Section 4); ii) the application of some statistical models to assess the graduate cooperation behaviour (Section 5); and iii) the use of survival analysis in order to measure the speed of data collection (Section 6).

2   Surveys on university-to-work transition

In the European Union and elsewhere, the transition from education to work is a matter of current policy and research interest. Policy-makers and researchers need to be able to monitor trends in the different processes and outcomes of the transition, in order to identify policy needs and to assess the effectiveness of alternative policy interventions. The interest in studying university-to-work transition has been growing during the last years in consideration of its relevance both for the users (students) and for the agents (universities). For these purposes, data on the labour market outcomes for university graduates have to be constantly collected. In Italy in the last ten years, many sample surveys on the subject have been conducted (Crocetta and D'Ovidio, 2003; Balbi and Balzano 2000; Chiandotto, 2001), and some projects composed of different research units - such

as ALMALAUREA established in 1994 (Almalaurea, 2004), or the recent survey conducted by ISTAT in 2002 - have been co-ordinated. Most of the above mentioned surveys have been conducted by traditional methods (such as by telephone, mail or face-to-face interview). Recently, new methodologies based on Internet have been tested, for example by the universities of Padua and Florence (Fabbris and Giusti, 2001) and by the University of Pisa (http://www.studenti.unipi.it and http://www.diogenet.net). This experience shows that web-based surveys present many advantages in terms of cost compared with the telephone surveys. However, the main conclusions have been that the use of the telephone is systematically preferable in terms of completeness of the information obtained.

In 2005 the Department of Statistics and Mathematics for the Economic Research at Parthenope University of Naples carried out a project in order to monitor the transition from university-to-work through a web survey. The project results make a new contribution towards defining and illuminating in-depth issues linked with the web survey data collection mode in the context of the transition from university to work.

Table 1: Description of the survey design.

	Wave 1	Wave 2		
Units	Respondents	Respondents with e-mail		Respondents without e-mail
n	651	555		96
% row	100%	85%		15%
Sample units definition	**W1R** (Wave 1 Respondents)	**SMU** (Survey Mail Units)	**STU** (Survey Telephone Units)	
n	651	444	111	96
% row	100%	68%	17%	15%
% column	100%	80%	20%	100%
First contact mode	Telephone	E-mail	Telephone	Telephone
Follow-up plan	-	E-mail and then if necessary telephone	-	-

3   The survey design

This survey aims to monitor the transition from university to work of university graduates in Economics in a longitudinal perspective. In fact, the panel design offers the possibility to follow graduates' working life across time, and hence to analyse the factors influencing their work histories. The first survey wave was conducted in Winter 2003 by a telephone interview. The sample size was

composed of 813 units (Quintano, Castellano and D'Agostino, 2004, 2005) and the response rate was 80%. Since internet penetration rate among graduates is high at wave 1 a high percent of graduates gave their e-mail address (85%). Graduates therefore are an optimal population for studying quality of the web participation process. The main characteristics of the web survey in wave 2 are: i) the availability of a probabilistic sample; ii) the mixed-mode strategy used for the first contact and for the follow-up plan; iii) the panel design. The first feature allows one to make statistically valid inference – something quite unusual in web surveys. The second allows us to measure the effect of the kind of the contact. The third helps to evaluate the marginal effect of the web mode of data collection on survey statistics.

The survey design is rather complex and it is summarized in Table 1.

In wave 2 sample units are composed of two groups:

i)   wave 1 respondents with e-mail address (555/651)=85%;
ii) wave 1 respondents without e-mail address (96/651)=15%.

As we can observe in Table1, the sample units are randomly assigned to a mode of the first contact as follows:

a) The group of the first contact by e-mail is composed of 444 units randomly chosen from the wave 1 respondents with e-mail, and are defined as Survey Mail Units (SMU);
b) The group of the first contact by telephone is composed of 111 units randomly chosen from the wave 1 respondents with e-mail address and all the 96 units from the wave 1 respondents without e-mail address; they are defined as Survey Telephone Units (STU).

The first step of the survey process is different for the two groups a) and b). For a), the first step consists of sending an e-mail requesting participation and a reply giving contact is required before an access password and the link questionnaire is sent to the respondent. For group b), the access password and link questionnaire are sent by e-mail, after they give a positive reply to the request for cooperation by telephone.

Then, a follow-up plan has been defined only for the SMU because we have focused our main interest on the quality process of the survey made by e-mail. In particular, two different modes were used one after another: first by e-mail and then by telephone if necessary.

In summary, the survey design in wave 2 allows us to make different analyses in terms of survey process quality: we can compare SMU and STU in terms of the first contact mode (e-mail or telephone), and analyse SMU depending on the follow-up plans.

4   Outcome rates as quality indicators

4.1   Definition of outcome rates

The definition of the outcome rates as quality indicators is necessary in order to evaluate the quality of any data collection process.

In web surveys in particular, outcome rates are very important because empirical evidence (Kiesler and Sproull, 1986; Sheehan and Hoy, 1997) shows that this data collection mode can suffer from high non-response rates, that are related to the first contact mode used. For this reason, in this paper we also analyse the effect of the first contact mode (e-mail or telephone) on the response rates achieved.

4.1.1 Definition of outcome rates for SMU

In order to simplify the definition of the outcomes rates as quality indicators for the SMUs, they are separated according to their possible outcome after the first contact by e-mail. In this way we obtain the distribution of SMU that is described in Figure 1.
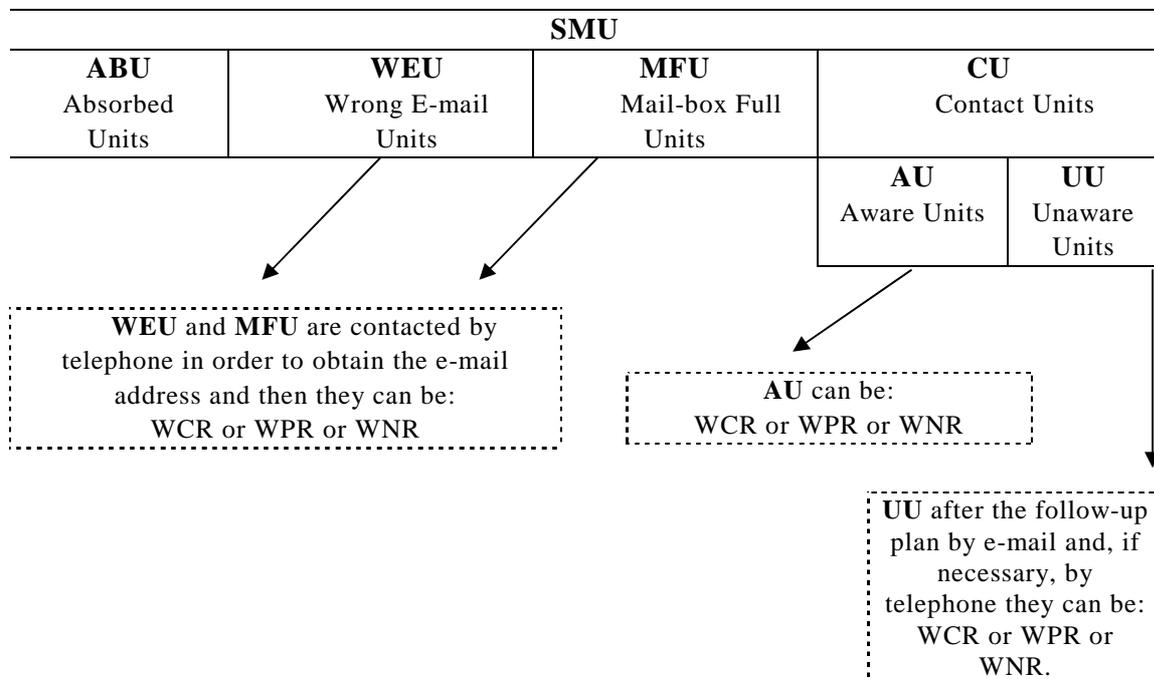

Figure 1: Distribution of SMU.

First of all, it is important to note that when the first contact is made by e-mail, it is difficult to identify who the units actually contacted for the survey are. In fact,

this contact mode is more complex than the one by telephone because it is based on a e-mail list. Consequently, it can happen that some wrong e-mail addresses do not send a return receipt to the sender but are absorbed by the network. Effectively we cannot estimate such "Absorbed Units" (ABU), and therefore the actual contact rate has to be estimated on the basis of the assumption that all SMU received the first contact by e-mail.

The non-coverage rate can be estimate as the sum of the wrong e-mail addresses and the e-mails rejected for people whose mail-box was full. In Figure 1, we define, respectively WEU as Wrong E-mail Units and MFU as Mail-box Full Units and consequently the total non-coverage rate is given by:

Total non-coverage rate= (WEU+MFU)/SMU

Table 2: Definition of outcome rates for SMU.

List of indicators	Computation
Contact quality indicators	
Gross contact rate	CU/SMU
Net contact rate	AU/SMU
Hypothetical contact rate	UU/SMU
Reply non-response rate given the contact	UU/CU
Survey participation quality indicators	
Response rate given the initial agreement	(WCR+ WPR)/AU
Complete response rate given the initial agreement	(WCR)/AU
Not complete response rate given the initial agreement	(WPR)/AU
Response rate given the contact	(WCR+ WPR)/CU
Total response rate	(WCR+ WPR)/SMU
Survey follow-up indicators	
E-mail follow-up effect rate	new AU after first reminder/
Telephone follow-up effect rate	UU-new AU after second reminder/ (UU-new AU after first reminder)

Further in Figure 1, we define CU as Contact Units, e.g. people informed of the survey. Logically, the CU are composed of two groups: the Aware Units (AU), giving an e-mail positive answer and the Unaware Units (UU) not replying to the same e-mail invitation[2].

---

[2] In fact, we define UU as people who do not reply to the first contact by e-mail without having explicitly refused to co-operate in the survey by e-mail.

The survey design at this stage, as described in Section 3, provides that only the AU receive the access password and the link questionnaire by e-mail. Essentially, not all AU have the same behavior with respect to the web questionnaire and for this reason we define three new groups of sample units that derive from AU as follows.

People not giving an answer to the web questionnaire are defined as WNR (Web Non Respondents); the ones giving a complete response to the web questionnaire are defined as WCR (Web Complete Respondents); and those giving a partial response to the web questionnaire are called WPR (Web Partial Respondents).
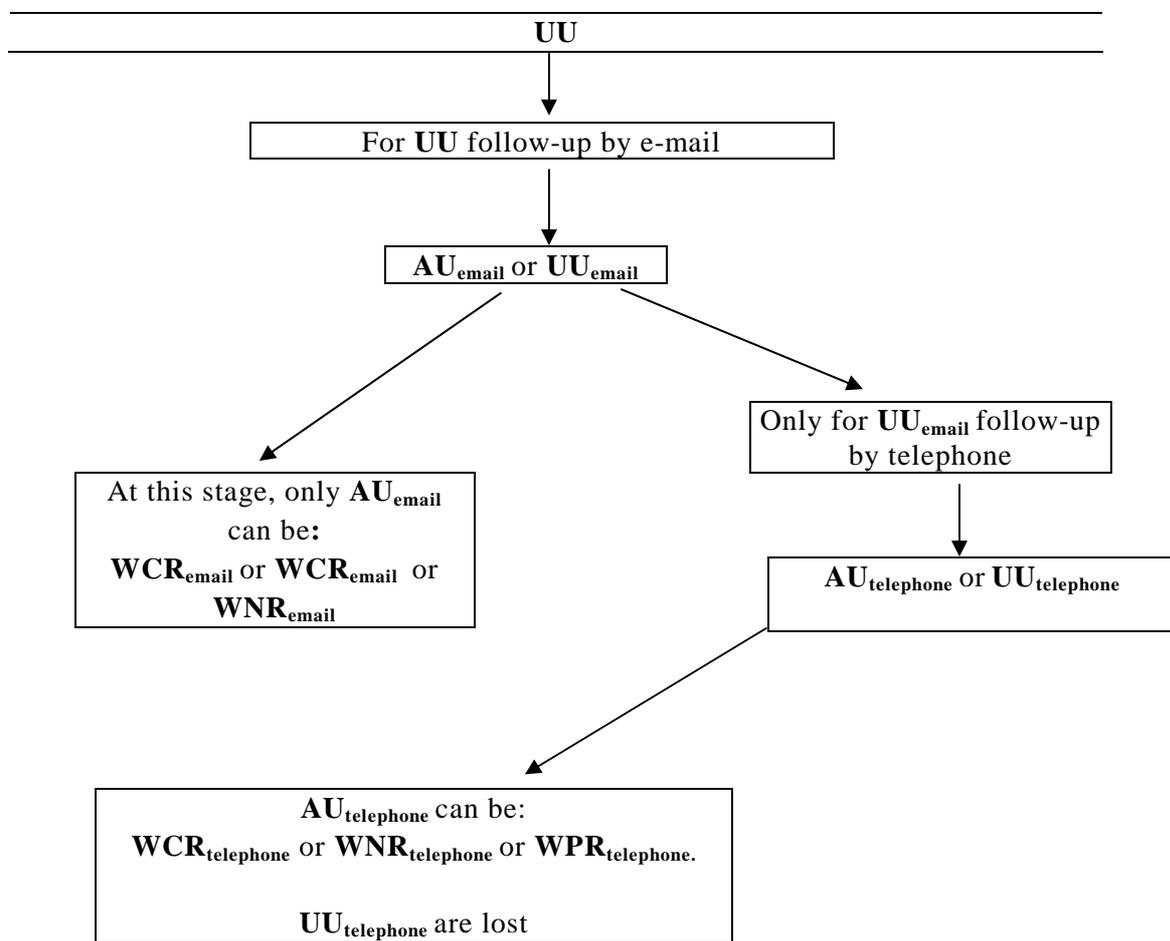

Figure 2: Distribution of UU after the follow-up plan.

These same groups can be defined for UU, but after the follow-up plan (by e-mail and then, if necessary, by telephone); and for WEU and MFU after the contact by telephone[3].

Using the groups defined in Figure 1, we can evaluate the process quality at each step of the data collection operation by computing different outcome rates as indicators of quality. We describe these outcome rates in Table 2.

We define in Table 2:

  i)   four outcome rates in order to evaluate the impact of the contact, named contact quality indicators;

  ii)   five outcome rates in order to measure the effect of the contact in terms of the web response rate, named survey participation quality indicators;

  iii)   two outcome rates in order to evaluate the effect of the follow-up mode (type of reminder), named survey follow-up indicators.

In order to clarify the discussion it is also important to give a definition of the new groups of units that derive from the follow-up plans. Figure 2 describes these groups.

Figure 2 shows how the follow-up plans work and how the number of UU reduces each time. Logically, the outcome rates described in Table 2 can be also computed after the follow-up plan and their computations are presented in Table 3.

4.1.2 Definition of outcome rates for STU

In the case of the first telephone contact, the discussion about outcome rates as quality indicators is easier because it is based on the telephone providing a direct contact with the statistical unit concerned. So, the CU consist of all graduates contacted by telephone. Moreover, the AU are all graduates contacted by telephone and declaring their availability and interest in participation in the survey. Logically, the definitions of WCR, WPR and WNR are the same as those given for SMU. Therefore, the definition of outcome rates for the STU, presented in Table 4 is made easier, even if it is necessary to define two new outcome rates that take into account people who refuse to collaborate and/or are unreachable.

4.2   Empirical evaluation of outcome rates

After the first contact by e-mail we can observe that the total non-coverage rate is 24%, with 18% of it due to wrong e-mail address and only 6% due to full mail-

---

[3] The telephone is made in order to have the e-mail for sending a password and the link questionnaire.

box. This quite high percentage of total non-coverage rate is plausible because the e-mail list was completed at wave 1.

Table 3: Definition of outcome rates for SMU after the follow-up plan.

List of indicators	Computation (e-mail)	Computation (e-mail then if necessary telephone)
Contact quality indicators		
Net contact rate	$(AU+AU_{email})/$ SMU	$(AU+AU_{email}+AU_{telephone})/$ SMU
Refusal rate	-	number of graduates that refuse to collaborate/ $(UU-AU_{email})$
Unreachable rate	-	number of unreachable graduates/ $(UU-AU_{email}))$
Survey participation quality indicators		
Response rate given the initial agreement	$(WCR+WPR+WCR_{email}+WPR_{email})/$ $(AU+AU_{email})$	$(WCR+WPR+WCR_{email}+WPR_{email}+WCR_{telephone}+WPR_{telephone})/$ $(AU+AU_{email}+AU_{telephone})$
Complete response rate given the initial agreement	$(WCR+WCR_{email})/$ $(AU+AU_{email})$	$(WCR+WCR_{email}+WCR_{telephone})/$ $(AU+AU_{email}+AU_{telephone})$
Not complete response rate given the initial agreement	$(WPR+WPR_{email})/$ $(AU+AU_{email})$	$(WPR+WPR_{email}+WPR_{telephone})/$ $(AU+AU_{email}+AU_{telephone})$
Response rate given the contact	$(WCR+WPR+WCR_{email}+WPR_{email})/$ CU	$(WCR+WPR+WCR_{email}+WPR_{email}+WCR_{telephone}+WPR_{telephone})/$ CU
Total response rate	$(WCR+WPR+WCR_{email}+WPR_{email})/$ SMU	$(WCR+WPR+WCR_{email}+WPR_{email}+WCR_{telephone}+WPR_{telephone})/$ SMU

In order to reduce the non-coverage rate, a telephone contact has made. In this way we retrieved 94% of these units. The coverage error is, therefore, very much reduced. At the end of this section we describe the survey participation indicators including this set of graduates. For now, these units are not considered.

We start by analysing the behaviour of CU by the mode of first contact (e-mail or telephone). In Table 5, the outcome rates described respectively in Table 2 and Table 4 are computed, and an approximate z-test for the differences between two binomial parameters obtained with the two different first contact modes is performed.

Table 4: Definition of outcome rates for STU.

List of indicators	Computation
Contact quality indicators	
Gross contact rate	CU/STU
Net contact rate	AU/STU
Refusal rate	number of graduates who refuse to co-operate, divided by STU
Unreachable rate	number of graduates who are unreachable, divided by STU
Survey participation quality indicators	
Response rate given the initial agreement	(WCR+ WPR)/AU
Complete response rate given the initial agreement	(WCR)/AU
Not complete response rate given the initial agreement	(WPR)/AU
Response rate given the contact[4]	(WCR+ WPR)/CU
Total response rate	(WCR+ WPR)/STU

Looking at the second column of the Table 5, the non-response rate given successful contact by e-mail is only 69%. This suggests that simply an e-mail contact is not sufficient to attract graduates' co-operation. Analysis of survey participation indicators shows that graduates participation in the survey is good when graduates respond to the initial contact. That's why the complete response rate is 85%. In any case, at this stage the response rate, given the contact by e-mail, is only 28% and the total response rate is consequently also lower (21%). We evaluate the effect of follow-up plan shortly; for the moment we concentrate on the differences with the first contact by telephone.

Looking at the third column of Table 5, the net contact rate by telephone is, as expected, much higher (82%) than the e-mail net contact rate (24%): telephone contact is usually more convincing as concerns participation in the survey. In fact, the refusal rate is very low (3%). On the other hand, the unreachable rate is about 15%, and this percentage is plausible considering that telephone number list is two years old.

The response rate given the contact by telephone (59%) is, as we expected, much higher than the corresponding rate for e-mail (28%). However, the total response rate with telephone contact is still only 50%. We expect a higher cooperation from graduates contacted by telephone. A possible explanation to this phenomenon is that among STU units, there are some graduates who subsequently did not give their e-mail address at wave 1, and they are the same units who refuse to answer the web questionnaire, as we describe in next section.

---

[4] It is important to note that in this case CU=AU+ people who refuse to collaborate in the survey.

Table 5: Outcome rates for SMU and STU.

List of indicators	Values % (first contact by e-mail)	Values % (first contact by telephone)	Approximate* z-test for the difference between two binomial parameters
Contact quality indicators			
Gross contact rate	76%	85%	-2.62
Net contact rate	24%	82%	-13.9
Hypothetical contact rate	52%	-	-
Reply non response rate given the contact	69%	-	-
Refusal rate	-	3%	-
Unreachable rate	-	15%	-
Survey participation quality indicators			-
Response rate given the initial agreement	89%	61%	4,62
Complete response rate given the initial agreement	85%	58%	4,41
Not complete response rate given the initial agreement	4%	3%	0,22
Response rate given the contact	28%	59%	-7,88
Total response rate	21%	50%	-9,83
Survey follow-up quality indicators			
E-mail follow-up effect rate	15%	-	-
Telephone follow-up effect rate	84%	-	-

*Theoretical value for the one-side z test is equal to -1.64 or 1.64 depending on the alternative hypothesis.

The results obtained for the response rate, given the initial agreement, is surprising: we have a higher value for e-mail (89%) than for telephone (61%). Once more we think that this can be explained by the fact that among 15% of graduates contacted by telephone (see Table 1) had not indicated their e-mail address at wave 1.

The results of the statistical test (fourth column of Table 5) show that in almost all cases the null hypothesis of equality between the two rates (e-mail and telephone) is rejected at 5% significance level. This means that the first contact mode has a significant effect on the outcome rates. In particular, first contact by telephone works better than first contact by e-mail for all contact quality indicators, and also, for "response rate given the contact" and for "the total response rate". Otherwise it has a negative effect for "response rate given the initial agreement" and for "complete response rate given the initial agreement". The only non-significant effect is that for "not complete response rate given the initial agreement".

Table 6: Outcome rates for SMU after the follow-up plan.

List of indicators	Values % (e-mail)	Values % (telephone)
Contact quality indicators		
Net contact rate	32%	69%
Refusal rate*	-	7%
Unreachable rate**	-	9%
Survey participation quality indicators		
Response rate given the initial agreement	90%	86%
Complete response rate given the initial agreement	86%	78%
Not complete response rate given the initial agreement	4%	8%
Response rate given the contact	37%	78%
Total response rate	28%	59%

An evaluation of the effect of the follow-up plans can be made by observing the last two rows of Table 5. The e-mail follow-up effect rate is only 15%; by contrast, the telephone follow-up effect rate is 84%. Therefore, it seems that the follow-up by telephone works much better than by e-mail. In order to measure the impact of the two follow-up modes on the other outcome rates, in Table 6 we report indicators presented in Table 3 that we consider crucial for the analysis.

The net contact rate after the e-mail reminder does not increase very much (from 24% to 32%). It is quite interesting to observe that in this case as well, if graduates reply to the contact their participation in the survey is very likely.

In fact, the "complete response rate given the initial agreement" after the e-mail follow-up is 86% and the not complete response rate is 4%. The "response rate given the contact" is increased to 37%, from the "response rate given the contact" of 28% presented in Table 5.

Otherwise the "net contact rate" after the reminder by telephone increases a lot (from 32% to 69%). At this stage we also observe a "refusal rate" equal to 7% and a "unreachable units rate" equal to 9%. These graduates are lost.

Concerning the participation in the web survey we have: 78% of units complete the questionnaire, an additional 8% give a not complete response. The "response rate given the contact" is now 78%. In summary, results confirm that the telephone reply can considerably improve the survey participation rates.

As noted above, we also reduce the coverage error using a telephone contact. for this group of people we observe the complete responses to the web questionnaire equal 70%, with additional 12%[5]. partial responses.

---

[5] These outcome rates are calculated by considering the response to the web questionnaire relative to the 107 sample units who are composed of WEU plus MFU. Moreover, we do not report all indicators of this group of analysis to simplify the discussion.

Therefore, also in this case, we note that the telephone contact is highly effective.

In summary, we put together results of different groups of analyses that derive from the 444 SMU and obtain the final total response rate equal to 78%, composed of 70% complete responses and 8% partial responses. If again we join results from SMU and STU we obtain a total response rate of 70%, composed of 63% complete responses and only 7% of partial responses[6].

5   Studying the probability of survey co-operation

The survey structure offers us the possibility to study survey co-operation under its different aspects, in particular considering the various endogenous (contact mode, follow-up plan) and exogenous (individual characteristics) factors explaining it. For this reason four different *logit* analysis (Cox and Snell, 1989; Amemiya, 1985) are performed in order to assess if the survey participation depends on: i) co-operation at wave 1; ii) the first contact mode; iii) characteristics of the graduates. The results of these analyses are presented in Table 7.

In particular, we estimate a logit model considering the STU that, as noted earlier, are composed of 15% graduates who do not give their e-mail address at wave 1 and of 85% graduates who do (see Figure 1 and question *(1)* in Table 7).

We observe that to give the e-mail address at wave 1 has a positive effect on the probability to be a respondent. So this information can be useful in the future in order to estimate the survey participation rate. And this can also explain in-depth the evidence already discussed in the previous section for STU. In fact, the high non-response rate is obviously connected to graduates who did not give their e-mail address and are unwilling to collaborate.

For measuring the effect of the first contact mode, we put together SMU and STU (see Figure 1). We can affirm that the probability to be a respondent increases if the first contact is made by telephone, even if some other structural factors are taken into account (see *(2)* and *(3)* logit analysis in Table 7). This result suggests the use of a mixed mode strategy in the survey process in order to attract more graduates participation. It can be also observed that the only factor with effect significantly different from zero is the age at graduation: younger graduates are more available to participate in the survey.

The last logit estimate (see *(4)* logit analysis in Table 7), made again on SMU and STU together, shows that the response rate depends on two covariates: i) the membership cohort (it decreases if the length of time to graduation increases); and ii) the individual age at graduation (it decreases if the individual age at graduation

---

[6] The last mentioned total response rate has been calculated on the 651 sample units that are SMU+STU.

increases)[7]. This result is quite convenient since it means that the non-response mechanism is not related to many variables and, consequently, the observed data quality is good. The fact is that only variables related to time influence the response: indeed it is interesting to link these results with the great ease of the younger generations with new technologies.

Table 7: Results of logit analysis.

(1) Does the probability to be a respondent at wave 2 increase if the graduate gave an e-mail at wave 1?	
Covariates	**Estimates (std)**
Intercept	0.34 (0.20)
X (1 if graduate gave e-mail at wave 1)	0.61 (0.28)
Likelihood ratio=4.71 p-value=0.03, n=207	
(2) Does the probability to be a respondent increase if the first contact is by telephone?	
Covariates	**Estimates (std)**
Intercept	-0.009 (0.13)
X (1 if first contact is by e-mail)	-1.305 (0.18)
Likelihood ratio= 52.84 p-value<.0001, n=651	
(3) Does the probability to be a respondent increase if the first contact is by telephone, controlling by individual factors?	
Covariates	**Estimates (std)**
Intercept	2.319 (1.03)
X_1 (1 if first contact is by e-mail)	-1.352 (0.18)
X_2 (age at the graduation)	-0.086 (0.04)
Likelihood ratio=58.24 p-value<.0001, n=651	
(4) Does the probability to be a web respondent depend on some individual factors?	
Covariates	**Estimates (std)**
Intercept	3.307 (0.93)
X_2 (age at graduation)	-0.075 (0.03)
X_3 (1 if cohort 1999)	-0.655 (0.28)
X_4 (1 if cohort 2000)	-0.666 (0.27)
X_5 (1 if cohort 2001)[8]	-0.580 (0.22)
Likelihood ratio=14.86 p-value=0.005, n=651	

---

[7] Both logit models (3) and (4) presented in Table 7 have been estimated only with significant covariates. Anyway, different extended models have been estimated earlier with many covariates including other individual characteristics but none of them had an effect significantly different from zero.

[8] The reference category is the cohort 2002.

6   Studying timeliness of co-operation

One of the very desirable characteristics of any survey is the timeliness of data collection. A measure of the data collection timeliness can be estimated considering the total length of people in the survey process (Pratesi *et al*.2004).

In this web survey, this duration is defined with respect to two different stages:

i)   between the first e-mail contact and receiving the reply, (in this case we analyze the group of AU);

ii)   between the sending of the web questionnaire link and receiving the reply (in this case we analyze all respondents to the web questionnaire[9]).

In order to model the duration of the total "stay" in the survey process in cases i) and ii), we use a non-parametric survival analysis model (Blossfeld *et al*., 1989; Cox and Oakes, 1984). However, it is important to note that the aim of survival analysis in this context is somewhat different from the conventional one. Usually the researcher is concerned with survival of the units and therefore with the factors that influence the survival in particular spells (for example the factors that prolong the persistence in a unemployment spell).

In studying the survey co-operation process instead, the researcher is interested in drop-outs of the process, among people who cooperated in the survey. And for this reason, we are looking for the factors that can influence the drop-outs.

In case i), we have 106 units who are the AU of the SMU subset. Time is expressed in hours. E-mails were sent on Thursday March 7 at about 1 p.m.

In Figure 3 Kaplan-Meyer survival curve is plotted. The median time is about 21 hours, and after 100 hours (about 4 days) only 20% of graduates survive.

This information can be useful in terms of the decision regarding the "wait time" before the first follow-up is undertaken. In fact, results suggest that it is desirable to quickly send e-mails without waiting for more than two days. This confirms the suggestion made by Schaefer and Dillman (1998) about compressing the time intervals between the follow-up contacts when e-mail remainders are used (Schaefer *et al*., 1998).

The estimated hazard[10,] presented in Figure 4, shows some peaks at particular dates; the level first decreases but then increases at about 300 hours. This means that the probability to reply has a negative duration dependence at first and then it becomes positive. It seems that there are some particular days when the reply possibility is higher, probably during the weekend.

---

[9] In this analysis we consider 397 units who are less than the effective respondents to the web questionnaire because for some units we can not calculate the durations.

[10] From the theory of survival analysis the hazard function gives the instantaneous potential per unit time for the event to occur among the units have survived up to time *t*.
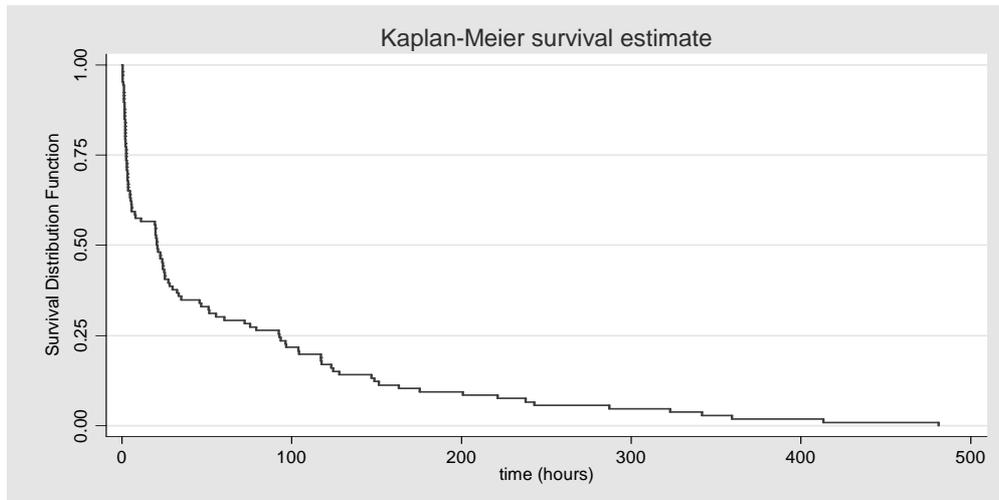
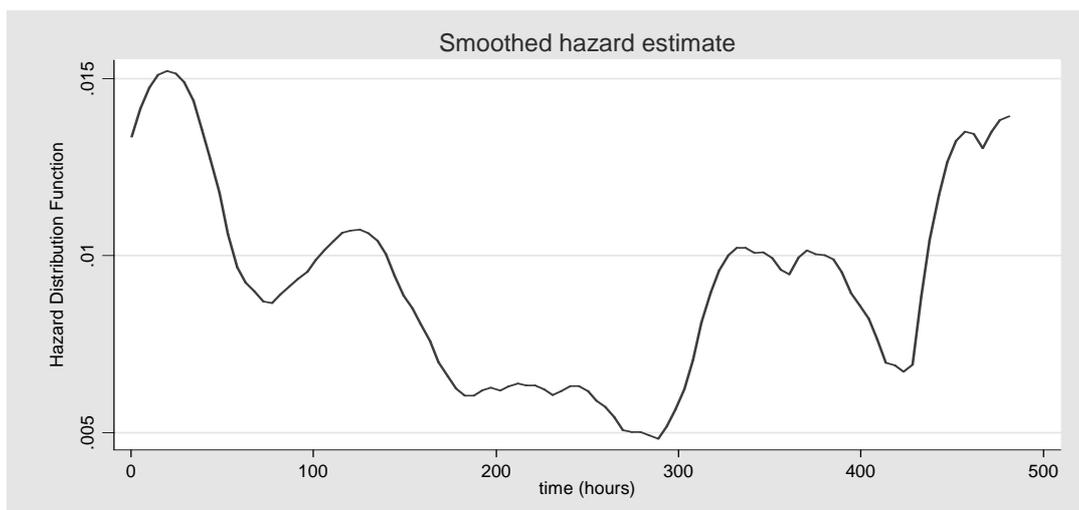**Figure 3**: Kaplan-Meier survival estimate – case i).

Figure 4: Smoothed hazard estimate – case i).

In case ii) we observe 397 durations. Time is expressed in days. The aim of this analysis is to study the effect of the first contact mode (e-mail or telephone). For this reason two survival curves are plotted in Figure 5. The first one is relative to the telephone contact and the other to the e-mail contact.

The estimated median duration is higher if the first contact is made by e-mail (about 8 days versus 2 days) and the survival curve decreases more rapidly for the first telephone contact. This means that if the e-mail is used for the first contact we observe a longer delay before receiving the answer to the web questionnaire.
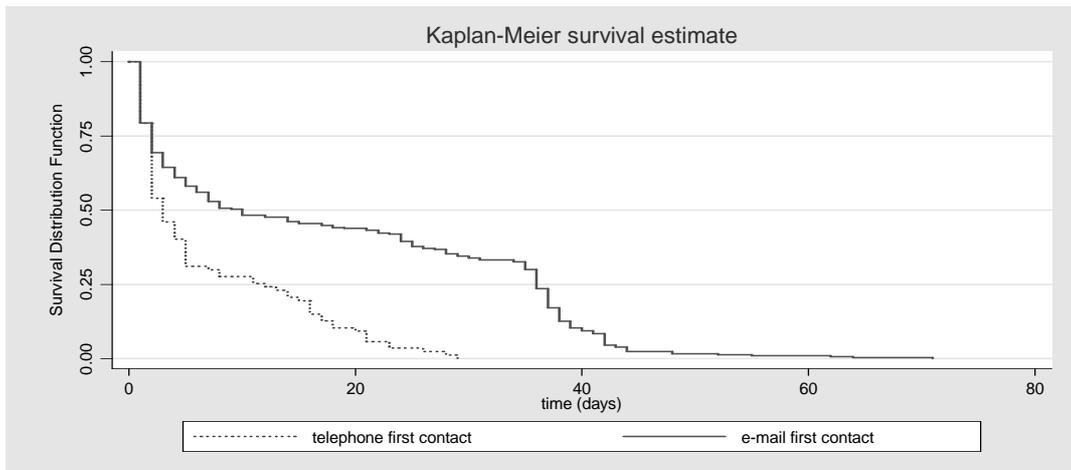
Figure 5: Kaplan-Meier survival estimate by first contact mode – case ii).

The smoothed hazard estimate in Figure 6 is always higher for telephone contact than for the e-mail one, so the "risk" to obtain a fast answer is higher with telephone contact.

In conclusion, the results have shown that in order to obtain timely information it is certainly better to perform the first contact by telephone. This does not require an excessive increase in survey costs and resources[11].
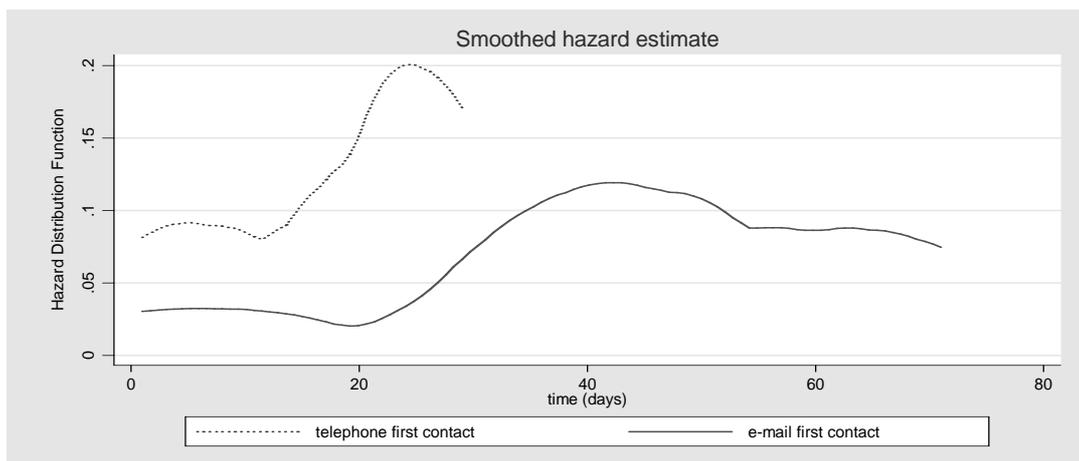

Figure 6: Smoothed hazard estimates by first contact mode – case ii).

---

[11] It is important to note that, logically, individual covariates can be added in the model specification, but at this stage we are only interested in studying how endogenous factors related to the survey process can influence duration.

7   Concluding remarks

The paper deals with quality issues in web surveys. In particular, a web survey for studying transition from university to work has been described and analysed, and empirical results have been discussed. We have paid particular attention to the definition of quality process indicators in order to evaluate the best strategy for attracting cooperation among graduates.

Empirical results in this framework are not widely available in Italy because surveys on transition from university to work in our country are usually conducted by traditional methods. For this reason, our contribution can be useful in order to investigate this new data collection mode both on its merits and its shortcomings.

Empirical results suggest that the adoption of a mixed-mode strategy, based on the use of telephone just for the first contact, can be the appropriate method for achieving acceptable response rates.

Anyhow, the population of graduates seem to react well to the web survey mode, and its reaction seems to confirm the suitability of this methodology for a longitudinal survey to collect data across time without an excessive waste of time and money. Some papers regarding the timeliness of web surveys show that even if surveys can theoretically compress the time of the data collection process, often in practice the average time of response can still amount to several days (Biffignadi and Pratesi, 2002; Bosnjak *et al*., 2001). By contrast, our present empirical study does not confirm this view: the total data collection duration of our web survey is very similar to that registered with the telephone survey that we conducted two years ago on the same units. Hence, at this stage we still believe that web survey can be considered a competitive method of data collection.

Of course other aspects have to be evaluated in-dept – such as the marginal effect of the web method on survey statistics - in order to have a more complete and exhaustive view. These issues can be the starting point for a future research.

Acknowledgement

References

[1]   Almalaurea (2004): *Relazione sulla condizione occupazionale dei laureati in Italia,* Bologna.

[2]   Amemiya, T. (1985): *Advances Econometrics*, Harvard University Press, Cambridge, MA.

[3] Balbi, S. and Balzano, S. (2000): *Controlli logici per indagini ripetute: indagine sugli sbocchi occupazionali dei laureati di Economia a Napoli* ("Federico II"), CAPTOR 2000, Padova.

[4] Biffignandi, S. and Pratesi, M. (2002): Internet surveys: the role of time in Italian firms response behaviour, *Journal of Research in Official Statistics*, **2**, 53-66.

[5] Blossfeld, H.P., Hamerle, A., and Mayer, K.U., (1989): *Event History Analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] Bosnjak, M., Tuten, T.L., and Bandilla, W. (2001): Participation in Web surveys – A typology in ZUMA Nachrichten, **48**, 7-17.

[7] Chiandotto, B. (2001): *Profilo e condizione occupazionale dei laureati dell'Ateneo fiorentino ad uno, due e tre anni dal conseguimento del titolo*, Università degli Studi di Firenze: http://www.unifi.it/aut_dida/rapporto-condoccup-092001.pdf.

[8] Cox, D.R. and Oakes, D. (1984): *Analysis of Survival Data*, London: Chapman and Hall.

[9] Cox, D.R. and Snell, E.J. (1989): *The Analysis of Binary Data*, Second Edition, London:Chapman and Hall.

[10] Couper, M.P. (2000): Web Survey: A review of issues and approaches. *Public Opinion Quarterly*, **64**, 464-494.

[11] Crocetta, C. and D'Ovidio, F. (2003): La valutazione dell'inserimento lavorativo dei laureati dell'università di Foggia attraverso un'analisi di segmentazione, in M. Civardi (a cura di), transizione Università-lavoro: la definizione delle competenze, collana, Determinazione e previsione di rischi sociali e sanitari, n. 4, Cleup, Padova, 111-132.

[12] Healey, B., Macpherson, T., and Kuijten, B. (2005): An Empirical Evaluation of Three Web Survey Design Principles. *Marketing Bulletin*, **16**, Research Note 2.

[13] Fabbris, L. and Giusti, A. (2001): Il progetto EXPERTUM: Sperimentazione di sistemi *computer assisted* per la rilevazione della valutazione della didattica universitaria da parte degli studenti e dell'inserimento lavorativo e professionale dei laureati e dei diplomati, *Atti del convegno intermedio della S.I.S. 2001 "Processi e metodi statistici di valutazione"*, Roma 4-6/6/2001.

[14] Klassen, R.D. and Jacobs, J. (2001): Experimental comparison of web, electronic and mail survey technologies in operations management. *Journal of Operations Management*. **19**, 713-728.

[15] Kiesler, S. and Sproull, L.S. (1986): Response effect in the electronic survey. *Public Opinion Quarterly*. **50**, 402-413.

[16] Pratesi, M., Lozar Manfreda, K., Biffignandi S., and Vehovar V. (2004): List-based web survey: quality, timeliness and nonresponse on the steps of the participation flow. *J. off. Stat*., **3**, 451-465.

[17] Quintano, C., Castellano, R., and Romano, A.A. (2000): La costruzione di un questionario elettronico per una indagine economica. Il caso della Stima provvisoria del Valore Aggiunto dell'ISTAT, Workshop "Metodi statistici e informatici per la rilevazione di dati assistiti da computer", Napoli 19 december 1997, in Fabbris L. Eds., *Il questionario elettronico*. CLEUP, Padova.

[18] Quintano, C., Castellano, R., and D'Agostino, A. (2004): The Determinants of the Transition from the University to Work: the Case of the Graduates in Economics at the University of Naples Parthenope, Atti della *XLII Riunione Scientifica della Società Italiana di Statistica.*

[19] Quintano, C., Castellano, R., and D'Agostino, A. (2005): "Moving from university to work: the case of a Faculty of Economics in the South of Italy", forthcoming in Quintano C. (a cura di), *Scritti di Statistica economica*, Istituto di Statistica e Matematica, Università degli Studi di Napoli "Parthenope", Napoli.

[20] Schaefer, D.R. and Dillman, D.A. (1998): Development of a standard e-mail methodology: Results of an experiment, *Public Opinion Quarterly*, **62**, 378-397.

[21] Sheehan, K.B. and Hoy, M.G. (1997): E-mail survey: response patterns, process and potential. *Proceedings of the 1997 Conference of American Academy of Advertisers.*

[22] Vehovar, V., Batagelj, Z., Manfreda, K.L., and Zaletel, M. (2002): Non response in Web surveys. In Groves R.M et al. (Eds): *Survey Nonresponse,* Wiley & Sons, Inc.