

A Competing Risks Model for Nonresponse in CATI Surveys

Un Modello a Rischi Concorrenti per la Non Risposta nelle Indagini CATI

Simona Drovandi

Dipartimento di Statistica “G. Parenti”
Viale G. B. Morgagni n. 59
Firenze
drovandi@ds.unifi.it

Monica Pratesi

Dipartimento di Matematica, Statistica,
Informatica e Applicazioni
Via dei Caniana, 2 I-24127 Bergamo
pratesi@unibg.it

Riassunto: Il contributo ha come oggetto l'applicazione di modelli di durata a rischi concorrenti per la nonrisposta nelle indagini CATI. Ogni candidato all'intervista telefonica ha una “biografia” che inizia con la prima telefonata e si conclude con un rifiuto, una risposta, un appuntamento mancato, o un mancato contatto. Al termine del periodo di rilevazione, il tasso globale di nonrisposta è il risultato dell'esposizione individuale ai rischi concorrenti di concludere l'indagine con uno degli esiti finali elencati. Il tempo tra la prima telefonata e l'esito finale della “biografia” definisce il periodo individuale di esposizione al rischio di partecipare all'indagine. I modelli proposti sono applicati allo studio della nonrisposta in tre indagini CATI condotte presso il Dipartimento di Statistica dell'Università di Firenze negli anni 2000 e 2001.

Keyword: Nonresponse, Competing risks models

1. Nonresponse in CATI Surveys

The response rate in CATI surveys is the result of the process of participation of each eligible respondent (case) (Groves and Couper, 1998). The process is dissected in two stages: first the telephone contact and then the response or the refusal. It is natural to assume that people who respond soon have an higher propensity to complete the interview than those who do not participate until the end of the survey period. The history of the outcomes of the call attempts done to contact the eventual respondent trace a sort of biography of the case inside the survey period. The biography begins with the first call to the case and ends with a refusal, response, missed appointment or missed contact. The time between the first call and the event that ends the biography defines the individual time of exposure to the risk of participating in the survey.

In this perspective the distribution of non response over refusal and non contacts categories depends on various features of the design of the survey and also on the individual characteristics of the eligible respondents.

The individual exposure to the competing risks of exiting from the survey with one of the defined final outcomes is measured by the unknown individual probability of participating in the survey. This can be estimated specifying a competing risk model that takes into account both the individual characteristics of the cases and the survey conditions.

Among the survey design features a crucial role is played by the calling protocols followed by the call scheduler of the CATI system. The automatic call scheduler is a program of the CATI system: it schedules the call attempts during the survey period following a standardized set of rules common to many CATI centers. For a complete

description of these rules (the calling protocol) we refer to other papers (Groves *et al.* 1988; Giusti and Pratesi, 1996).

2. The Competing Risks Models

The risk of exiting from the survey because of refusal is competing with the risk of responding or of remaining in a category of missed appointment or missed contact. For example, exiting for responding after an appointment inside the survey period is competing with the risk of remaining with an unsolved appointment at the end of the survey period.

Actually, the event that marks the exit of the eligible respondent from one risk set has to be considered as a form of censoring in order to study the other competing risks under the assumption of not informative censoring. This means that, conditional on the covariates, those who are at particularly high (or low) risk of one event type are no more (or less) likely to experience other kinds of events; in other words, given the covariates, the competing risks should be independent.

For this reason the specification of the competing risk model has required two steps. The first step is the test of the dependence/independence between competing risks: we check whether there are individual unobservable factors eventually affecting all the competing risks considered. The second step is the fitting of a proportional hazard model in order to measure the effect of the individual characteristics on the independent competing risks and to estimate the individual hazard of response.

The competing risks model has been fitted on the data collected with three CATI surveys conducted at the Department of Statistics of the University of Florence between 1998 and 2001. The first one is a list based sample survey on 3000 Italian Television Subscribers, interviewed in 2000 about their opinions on the TV service (CATI_ITV in the following). The second one is a survey on the whole population of students who graduated in the 1998 in the University of Florence (CATI_GRA in the following). The last one is a Random Digit Dialling survey on Radio listeners in Tuscany Region during the 2001 (CATI_RDD in the following).

3. The Case Studies

We report here only a subset of the obtained results: we limit the analysis to the biographies that end with an appointment. The types of appointments defines three different competing destination states: event type 1 “best time for the respondent not known”; event type 2 “appointment with respondent”; event type 3 “appointment with respondent parents” (in the following each event type or destination state is indicated with the letter *O*).

The dependence/independence between competing risks has been checked fitting to the data a model with latent variables that allows the dependence of the different event types on the same non parametric unobserved factor (Vermunt, 1996).

We have considered the following covariates: sex (*A*), age 24-29 (*B*), age 30-39 (*C*), residence in then north of Italy(*D*), residence in the middle of Italy (*E*), working level for manager (*F*) and working level for clerk (*G*). The proportional hazard model for the event type *O* (destination state) is the multinomial logit model with a non-parametric latent variable (*W*):

$$l_0(t|a,b,c,d,e,f,g,w) = \frac{\exp(u_0^O + u_{ao}^{AO} + u_{bo}^{BO} + u_{co}^{CO} + u_{do}^{DO} + u_{eo}^{EO} + u_{fo}^{FO} + u_{go}^{GO} + u_{wo}^{WO})}{1 + \exp(u_0^O + u_{ao}^{AO} + u_{bo}^{BO} + u_{co}^{CO} + u_{do}^{DO} + u_{eo}^{EO} + u_{fo}^{FO} + u_{go}^{GO} + u_{wo}^{WO})} \quad (1)$$

From the obtained results the competing risks seem to be independent only for the CATI_GRA survey. Table 1 shows the log-likelihoods of the estimated competing risks for the above model.

Table 1: Test results for the estimated competing risk models for leaving the state of non contact, for the three surveys considered, with different types of appointment as competing destination states

Model	log-likelihood Cati_itv	log-likelihood Cati_gra	log-likelihood Cati_rdd
1. 1 no covariates	30987.23	50089.94	45324.39
2a. 1 class model	30543.12	50045.34	44123.56
2b. 2 class latent model	28324.11	49011.19	42145.66
2c. 3 class latent model	29134.09	49091.14	41234.23
3a. 2 class + AW,BW,CW,DW,EW,FW,GW	28302.97	49101.16	43123.43
3b. 3 class+ AW,BW,CW,DW,EW,FW,GW			42564.78

Model 1.1 contains only the main effect, u_0^O , without covariates, and model 2a contains also the two-variable interactions u_{ao}^{AO} , u_{bo}^{BO} , u_{co}^{CO} , u_{do}^{DO} , u_{eo}^{EO} , u_{fo}^{FO} , and u_{go}^{GO} without the latent variable (1 class model). By comparing these two models without unobserved heterogeneity, it seems that the observed covariates have a significant effect on the probability of leaving non contact state for the different destination states in the three surveys (see the decrease in the log-likelihood from model 1.1 to model 2a). Models 2b-2c contain a latent variable (W) which is postulated to be independent on the observed covariates. Comparing the log-likelihoods of models 2b, and 2c, it can be seen that the two-class solution, that is a latent variable with two categories (model 2b), captures almost all unobserved heterogeneity (for CATI_GRA and for CATI_ITV). The decrease of the log-likelihood by including a third latent class, that is a latent variable with three categories (model 2c), is negligible. The same conclusion is not true for CATI_RDD in which the decrease of log-likelihood by including a third latent class seems to demonstrate that the three-class solution (model 2c) captures almost all unobserved heterogeneity. In model 3a the assumption that the unobserved heterogeneity is independent on the observed covariates is relaxed by including direct effects of A, B, C, E, F and G on W in the model. The two-class model (model 3a) improves not significantly for CATI_ITV, while the same model seems unimproved for CATI_GRA. In other words for CATI_GRA the competing risks seem to be independent given the observed covariates. The same conclusion is not so evident for CATI_ITV.

In order to investigate which individual characteristics of the eligible respondents affect the competing risks the model has been fitted only for CATI_GRA for which the competing risks seem to be independent referring to model 2b.

In Table 2 for event type 2, that is appointment with respondent, we can see that the risk to take an appointment with interviewers declined about 15 percent for men compared with women, and people aged 24-29 and 30-39 have respectively a 77 and 15 percent greater risk to take appointment with interviewers than their 40 years old counterparts.

Moreover people who live in the North and in the Middle of Italy have a higher risk to take an appointment than their Southern counterparts. The same conclusions are true for Manager and clerk compared to workers. Similar consideration are true also for event type 1 that is best time for respondent not known. On the contrary, for event type 3, that is appointment with respondent parents, we have find different results. In fact the risk for interviewers to take an appointment with respondent parents increase about 4 percent when respondent parents are men compared with women, while for respondent parents aged 24-29 and 30-39 we record respectively a 13 and 3 percent lower risk to take an appointment with interviewers. The same conclusions are true for respondents parents who live in the North and in the Middle of Italy. In this case, in fact, the risk for respondent parents to take an appointment with interviewers decline respectively of 60 percent and 54 percent for people resident in the North and in the Middle of Italy than their Southern counterparts. Finally according to working condition of respondent parents we find that manager and clerk ones have a higher risks to take an appointment with interviewers than their worker counterparts.

Table 2: *Estimated hazard ratios for CATI_GRA*

HAZARD-RATIOS	Model 2b CATI_GRA		
	EVENT TYPE 1	EVENT TYPE 2	EVENT TYPE 3
A) MEN VERSUS WOMEN	0.854	0.847	1.041
B) AGE24-29 VERSUS AGE40+	1.768	1.761	0.866
C) AGE30-39 VERSUS AGE40+	1.147	1.14	0.972
D) RESIDENT_NORD VERSUS RESIDENT_SUD	3.093	2.786	0.597
E) RESIDENT_MIDDLE VERSUS RESIDENT_SUD	1.098	0.987	0.539
F) MANAGER VERSUS WORKER	2.263	2.235	1.068
G) CLERK VERSUS WORKER	2.42	2.429	1.15

Event type 1: Best time for the respondent not known; Event type 2: Appointment with respondent; Event type 3: Appointment with respondent parents

References

- Giusti A., Pratesi M. (1996) Nuove tecnologie nel processo di produzione dei dati Statistici, *Atti della XXXVIII Riunione SIS, Rimini*.
- Groves R. M., Biemer P. P., Lyberg L. E., Massey J. T., Nicholls W. L., Waksberg J. (1988) *Telephone Survey Methodology*, J. Wiley.
- Groves R. M., Couper, M. P. (1998) *Nonresponse in Households Interview Surveys*. New York: John Wiley and Sons, Inc.
- Vermunt, J. K. (1996) *Log-linear Event history analysis. A general approach with missing data, latent variables, and unobserved heterogeneity* Tilburg University Press.