

Interviewer's Effect on Refusal Risk in the Italian Multipurpose Survey: a Multilevel Approach⁽¹⁾

L'Effetto Rilevatore sul Rischio di Rifiuto nelle Indagini Multiscopo: un Approccio Multilevel

Barbara Baldazzi

Romina Fraboni

Sante Orsini

ISTAT

Via Adolfo Ravà, 150

Roma

e-mail: baldazzi@istat.it

e-mail: fraboni@istat.it

e-mail: orsini@istat.it

Alessandro Rosina

Università Cattolica del Sacro Cuore

Largo Gemelli, 1

20123 – Milano

e-mail: rosina@stat.unipd.it

Riassunto: Il rifiuto di collaborare all'intervista è il motivo principale di mancata risposta totale nelle indagini multiscopo ISTAT sulle famiglie, e risulta in tendenziale aumento nel tempo. E' pertanto importante studiare tale fenomeno, soprattutto al fine di acquisire informazioni che consentano di migliorare la qualità non campionaria delle indagini future. Presentiamo qui una analisi condotta con un modello logistico multilevel, con unità di primo livello le famiglie e di secondo gli intervistatori. I risultati ottenuti evidenziano un maggiore rischio di rifiuto dove è maggiore la percezione di insicurezza (grandi centri del Nord, persone anziane) e un effetto di contenimento del rischio nei rilevatori che (per esperienza, età ruolo istituzionale) ispirano fiducia.

Keywords: refusal risk, multilevel modelling, non sampling errors.

1. Introduction

Istat has been currently conducting "Multipurpose" integrated system of social surveys since 1993. Seven sample surveys on resident households, one annual on living conditions and other six on specific themes, rotate in a time slot of five years covering most important topics of social statistics. They constitute a system because they share methodological approaches, data collection and treatment techniques, analysis and dissemination strategies. Five of the seven sample surveys are based on PAPI technique. Istat PAPI (Paper And Pencil Interview) survey data collection is based on local administrative offices spread all over Italian territory at regional and municipal level. From Istat central office to the local network of interviewers two main steps are performed: at first, Istat central office contacts the 21 Istat regional offices to provide information on the survey (questionnaires, instructions, time schedule, etc.); then, every regional office contacts selected municipalities (first stage sampling units) to co-

⁽¹⁾ The authors share contents and views expressed in this paper. However S. Orsini drew up the section 1; A. Rosina section 2; B. Baldazzi section 3 and R. Fraboni section 4.

ordinate second stage sampling selection (households), interviewers' recruiting and fieldwork performance. The co-ordination and monitoring of such an articulated network of operators are not easy, because many of them are not experienced in statistical surveys and because there is not a direct link between the place where the survey is planned and the place where fieldwork is performed. This framework has its pro and cons: on one hand it guarantees a low non response rate due to the fact that respondents tend to co-operate with local administrative officers and personnel; on the other hand it makes more difficult to check and dominate fieldwork process (interviewers' training, contact rules, face to face or self administered interview, etc.). Fieldwork phase can hide many traps causing non sampling errors. Such errors can have an even stronger impact on estimates' quality as compared to probabilistic sampling ones. Then, the quality of the survey network is a crucial aspect in the strategy of non sampling error prevention. The annual survey on living conditions is one of the PAPI surveys and it is the pivot of the multipurpose system. Information about data quality has been collected repeatedly and for a long time. For every selected household data on non response by cause, municipality and interviewer are available. Thanks to this characteristics this survey represents a good benchmark also to analyse and test new techniques to prevent non sampling errors. Aim of this paper is to use this information by the application of multilevel models to analyse deeply the relationships between interviewer's features and the refusal risk. In particular, refusal risk has recently increased over time: in the 1999 Annual survey refusal risk has reached on average 6.3%, beyond 10% in metropolitan areas. The results of the analysis can be used to ensure a better quality of the survey by improving the selection criteria of interviewers, the training methods and testing the efficiency and effectiveness of the local network.

2. Strategies of analysis and model

Households are not randomly assigned to interviewers. In comparing interviewers' performances, both taking into account the territorial dimension (such as macroregion and municipality typology where each interviewer operates) and some basic features of assigned households (such as age and sex of the householder, number of family members) are then necessary. Territorial and household characteristics are included in the model as control variables. Interviewer's status in occupation and his/her experience in survey fieldwork are explicative variables of major interest for us. In particular, we are interested in evaluating if young, occasionally recruited (especially among students) interviewers' performances are better, in terms of risk refusal containment, than long trained interviewers recruited among local administrative officers and personnel. Let us suppose that the refusal to collaborate to the interview is to large extent due to the fear to open the door to a stranger. If that is true, beside a higher refusal risk in northern towns and among elderly (where the unsafety feeling is higher), we expect to obtain a higher propensity to co-operate when the interviewer belongs to the municipal police than in other cases. To take into account within-interviewer correlation the model we adopt is a multilevel one with households as first level units and interviewers as second level units. The analysis has been conducted by using *nlmixed* procedure.

3. Dataset construction.

The dataset used for this analysis comes from a linkage of different data sources.

At the beginning of survey, we arrange the following information for every sampled household: household's code (Province's code + Municipality's code + household's progressive number), interviewer's code (Province's code + Municipality's code + interviewer's progressive number). At the end of survey we arrange two datasets: one for not interviewed households and one for interviewed households.

The first dataset is created with information derivate by family certificates issued by Municipality: number of family members and age, sex, citizenship and married status of each family's head (one record per household). Due to lack of uniformity in family certificates coming from different municipalities, this dataset contains some missing data (about 2,7%). The second dataset is created with information derivate by family's interview. At the end of the survey we arrange a dataset with one record per every family member. From this dataset we create a file containing information only for one family member, the householder. Information recorded are the same of the first dataset, but, in this case, come directly by interview. This dataset is free of missing data.

Joining the two dataset (interviewed and not interviewed households files), we perform a linkage to interviewer's information derivate by ISTAT interviewer's archives (through interviewer key).

4. Results and discussion

In table1 we present the outcomes of three models: in the first one we only control for the territorial dimension, while in the second one we also add the family characteristics. In the third model we evaluate if and to what extent different characteristics of the interviewers recruited in each municipality allow to explain the differences in terms of refusal risk. Previous analyses on the same data set showed that especially fieldwork experience and occupational status of the interviewer are very important (Rosina et al., 2001). In order to satisfy the parsimony criterion we therefore limit the inclusion of our covariates only to these ones. Model 1 shows a higher refusal risk for households resident in large towns (especially centre and ring of the metropolitan areas) and in the North of Italy. The null model without covariates (that we do not show here) has a component variance equals to 2.10: the introduction of the territorial dimension allows to reduce this variance to 1.46 (see model 1), i.e. a reduction of about 30%.

In model 2 we estimate refusal risk net of relevant family characteristics: municipalities with high risks of refusal are confirmed also when the family characteristics are included (especially when the householder is older than 75 years and when s/he lives alone). The last model analyses the importance of the role of the interviewer, controlling for territorial dimension and family characteristics. When interviewer has been long trained in fieldwork (more than 6 years of experience) s/he is more likely to succeed in obtaining a collaboration to the survey. Moreover, low (although not significant) refusal risks are associated with interviewers working for the municipal police while particularly high (and significant) refusal risks are associated with students or employees somewhere but the municipality setting. When interviewer profile is included the variance component (1.2347) is reduced of about 14% with respect to model 2, controlling for the characteristics of the family.

Table 1: Main results of the multilevel model. Refusal risk in the 1999 Multipurpose Household Survey.

Covariate	MODEL 1		MODEL 2		MODEL 3	
	Param.	(s.e.)	Param.	(s.e.)	Param.	(s.e.)
<i>Territorial dimension:</i>						
Metropolitan area	2.1913 ***	(0.1479)	2.1457 ***	(0.1473)	2.3414 ***	(0.1468)
Metropol. area ring	1.1921 ***	(0.1734)	1.2173 ***	(0.1728)	1.2869 ***	(0.1684)
Large municipality	0.8817 ***	(0.1168)	0.8722 ***	(0.1164)	0.9151 ***	(0.1139)
Small municip. (ref.)	0		0		0	
North	0.7983 ***	(0.1170)	0.7778 ***	(0.1165)	0.6554 ***	(0.1134)
Centre	0.1463	(0.1249)	0.1375	(0.1244)	0.1176	(0.1207)
South (ref.)	0		0		0	
<i>Family characteristics:</i>						
<i>Householder's age:</i>						
<=54 yrs			-0.1003	(0.06821)	-0.09430	(0.06812)
55-74 yrs (ref.)			0		0	
>=75 yrs			0.2443 **	(0.08931)	0.2547 **	(0.08911)
<i>Family dimension:</i>						
1 member			0.2970 **	(0.09135)	0.2995 **	(0.09114)
2-3 members			0.1792	(0.07806)	0.1782	(0.07797)
more than 3 members (ref.)			0		0	
<i>Interviewer characteristics:</i>						
<i>Experience</i>						
<6 yrs (ref.)					0	
6 yrs and more					-0.3652 **	(0.1151)
<i>Occupational status</i>						
Official statistic (ref.)					0	
Municipal police					-0.4487	(0.1677)
Other municipal employee					0.08725	(0.1314)
Student					0.8732 ***	(0.1303)
Non municipal employee					0.6982 **	(0.2239)
Unemployed					0.3412	(0.1466)
Intercept	-4.2822 ***	(0.1394)	-4.4077 ***	(0.1334)	-4.5107 ***	(0.1441)
Variance component	1.4602		1.4358		1.2347	

Therefore our hypotheses regarding the fear to open the door to a stranger are confirmed by higher refusal risks in the case of elderly, living alone, in large towns, in the North of Italy, contacted by students, with short experience. The multilevel approach allows to diminish the variance component observed in each successive model. In particular, its major reduction is due to the introduction of the territorial dimension as covariates. The variance component is furtherly reduced when the characteristics of the interviewer are introduced.

References

- Camporese R., Orsini S. (2001) "Non response rates as indicators of field work quality in the Italian experience of Multipurpose Annual Survey on Living Conditions".
- Camporese R., Quattrocioni L., Sabbadini L.L. (2001) "Strategie di prevenzione delle mancate risposte e degli errori non campionari nelle indagini Multiscopo", SIEDS.
- Camporese R., Romano M.C., Rosina A., Vitaletti S. (2001) "Una valutazione della rete di rilevazione comunale: l'esperienza delle indagini multiscopo sulle famiglie", Convegno SIS "Processi e metodi statistici di valutazione", Roma 2001.
- Hox J.J. (1995) *Multilevel Statistical Models*, TT-Publikaties, Amsterdam.
- Rosina A. (2001) "Analisi della dipendenza del rischio di non risposta totale dalle caratteristiche delle famiglie e del rilevatore", ISTAT, Roma.
- SAS Institute Inc. (1999) *The NL MIXED Procedure*, SAS/STAT User's Guide. Version 8, Cary, NC, pp 2421-2504.