Question order and interviewer effects in CATI scale-up surveys

Silvia Snidero, Federica Zobec, Paola Berchialla, Roberto Corradetti,
Dario Gregori

Working paper: 01/2008

Abstract

The Scale-Up estimator (Bernard, Social Science Research, 1991) is network-based estimator for the size of hidden or hard-to-count subpopulations. Several issues are arising in the public health context when the aim is the estimation of injuries occurring in a certain population, where two common problems are present: (i) the fact that small injuries are usually difficult to observe, and a rarely reported in the official data, and (ii) people are not always compliant in giving information about some specific injuries, in particular when children are involved.
The aim of this study is to check the methodological issues arising from the usage of a CATI survey using the scale-up methodology for detecting the number of injuries due to chocking in children age 0-14 years in Italy. For this purpose, 1000 CATI interviews were conducted in Italy during a week using a questionnaire based on 33 questions about populations of known size according to the ISTAT Census data. Then, each respondent was asked about other questions, related to the main target population (e.g.: number of children known to suffer for a chocking accident). A sensitivity analysis was conducted for estimating the effect of varying sub-populations, the order of the questions and the interviewer effects on the resulting estimates. But for the interviewer effect, no particular differences were observed in the overall estimates of injuries. Our conclusion is that the scale-up estimator in association with CATI methodology is showing a high potential in the field of injury prevention, being accurate and robust, but particular attention should be given to the training of the interviewers to improve stability of the estimates.

Keywords:

Scale-up method, interviewer effect, question order, foreign body injuries, population size estimation.

1. Introduction

A lot of emphasis has been given in recent years to research, both methodological and applied, on the scale-up estimators. The scale-up method was first introduced by Bernard (Bernard, Johnsen, Killworth, and Robinson 1991; Bernard, Johnsen, Killworth, and Robinson 1989) and relies on the reasoning that the proportion of people known by respondents in a target subpopulation, averaged over all respondents, should give an approximation of the proportion (and hence the number) of people in the general population who are members of the subpopulation. This can be expressed as $m/c = e_0/t$ where m is the average number of people known in the target subpopulation E_0 by respondents, c is the mean social network size, e_0 is the target subpopulation size and t the size of the general population T .(Bernard, Johnsen, Killworth, and Robinson 1991; Bernard, Johnsen, Killworth, and Robinson 1989; Johnsen, Bernard, Killworth, Shelley, and McCarty 1995; Killworth, Johnsen, McCarty, Shelley, and Bernard 1998a; Killworth, McCarty, Bernard, Shelley, and Johnsen 1998b). In the previous equation, the only unknown variable is c , the average social network size; thus it should be substitued with a good estimate (Bernard, Johnsen, Killworth, and Robinson 1991; Bernard, Johnsen, Killworth, and Robinson 1989; Killworth et al. 1998a; Killworth et al. 1998b; Snidero, Corradetti, and Gregori 2004).

Quite surprisingly, even though the precision of the scale-up relies on the precision of the answers given regarding the sizes of the specific subpopulations, no dedicated research has been found in the literature on the stability of answers across variations: interviewer effect and question order. Issues like the effect of the interviewer and the

order of the questions (should we ask first about the target or about the other populations?), are still open and need to be addressed.

More general results are of course available on question order and interviewer effects, which have been deeply studied with reference to traditional surveys (Crespi and Morris 1984; Crossley and Kennedy 2002; DeMoranville and Bienstock 2003; Fushs 2000; Gao 2001; Marsden 2003; McColl, Eccles, Rousseau, Steen, Parkin, and Grimshaw 2003; McFarland 1981; Moy, Scheufele, Eveland, and McLeod 2001; Pickery and Loosveldt 2001; Pickery and Loosveldt 2000; Pickery and Loosveldt 2004; Ramirez and Straus 2006; Rimal and Real 2005; van Tilburg 1998). Among these works only a few papers dealt with the problem of question order and interviewer effects in the social network field: Van Tilburg and Marsden (Marsden 2003; van Tilburg 1998) studies were aimed at evaluating the interviewer effects in measuring the network size using name generators, and Coromina's paper (Coromina and Coenders 2006) focused on egocentered network data collected via the web. The scale-up method case is quite different from the name generator estimator as in the first one respondents give just a number and in the second one respondents elicit the names of their social network. For this reason it could seem that the scale-up method do not suffer of interviewer effects, but eliciting names gives respondents more time to recall the people effectively belonging to a certain group than answering with just a number including all the people known in the same group. In this sense a different interviewer approach can facilitate the disclosure of respondents.

Another issue, which is specific to the scale up, is the fact that the subpopulations of known size on which the questions of the scale-up method are based, comes usually from institutional and official sources (e.g.: medical statistics, census data and governmental statistics like suicide, injuries, homicides, etc.), thus several questions

concern sensitive subjects and people might not be willing to answer certain kinds of questions (Johnsen et al. 1995; Snidero, Morra, Corradetti, and Gregori 2005). Although this method took a big advantage from indirect questioning, avoiding speaking about own sensitive matters, some questions could in any case make respondents feel uncomfortable (Bernard, Killworth, Johnsen, Shelley, and McCarty 2001; Johnsen et al. 1995; Kadushin, Killworth, Bernard, and Beveridge 2006; Killworth et al. 1998a; Killworth et al. 1998b; Moody 2006; Snidero, Corradetti, and Gregori 2004). Therefore, the need emerged to understand where it is better to place the sensitive questions (e.g. after some non sensitive questions that may feel respondents more comfortable and avoid therefore some possible bias?) and how the questionnaire should be administered (e.g. more gentle and less determined interviewers are more likely to obtain answers on sensitive questions? Giving more time to respondents to recall the number of people known in a certain population is better than quicken the questions? Is it helpful to avoid distortion remember to respondents several times during the interview the social network definition?).
Things are further complicated when the research is focused on public health sensitive issues. In our work, we applied the scale-up to the estimation of the number of self-resolved foreign body injuries in children 0-14. Foreign body injury consists of the insertion, ingestion or aspiration of an object in the upper air-digestive ways. In public health, foreign body injury represents a rare event but not negligible as airway obstruction represents the most common cause of death in children under 4 years of age (Zigon, Gregori, Corradetti, Morra, Salerni, Passali, and Passali 2006).
In the next chapter we will briefly describe the scale-up method and the social network estimator. Then we will introduce the study design and the questionnaire. Section 4 presents the results and finally there is a discussion.

2. The scale-up method and the social network estimator

From Bernard's idea, consisting of the proportionality among the mean number of people known in a certain subpopulation and the size of this subpopulation, Killworth (Killworth et al. 1998a; Killworth et al. 1998b) a maximum likelihood estimator was developed that was very simple to calculate:

$$\hat{e}_0 = t \frac{\sum_i m_{i0}}{\sum_i c_i} \tag{0.1}$$

where m_{i0} is the number of people known in the target subpopulation E_0 by the i-th respondent, t is the general population size and c_i is the social network size of the i-th respondent (Killworth et al. 1998a; Killworth et al. 1998b).

The scale-up estimator is unbiased and its standard error is (Killworth et al. 1998a; Killworth et al. 1998b):

$$s.e.(\hat{e}_0) = \sqrt{\frac{t \cdot \hat{e}_0}{\sum_i c_i}}. \tag{0.2}$$

In the past years several social network size estimators were studied (Bernard, Johnsen, Killworth, McCarty, Shelley, and Robinson 1990; Bernard, Johnsen, Killworth, and Robinson 1991; Bernard, Johnsen, Killworth, and Robinson 1989; Bernard, Killworth, McCarty, and Shelley 1990; Freeman and Thompson 1989; Johnsen et al. 1995; Killworth, Bernard, and McCarty 1984; Killworth, Johnsen, Bernard, Shelley, and McCarty 1990; McCarty, Killworth, Bernard, Johnsen, and Shelley 2000; Zheng, Salganik, and Gelman 2006) and some of these use the known size of a certain number of subpopulations to measure the social network sizes.

To get an unbiased estimate of the social network size we choose an estimator belonging to a class of estimators that underlies the same idea of the scale-up method, which is the proportional estimator:

$$\hat{c}_i = t \cdot \frac{\sum_{j=1}^{L} m_{ij}}{\sum_{j=1}^{L} e_j} \tag{0.3}$$

where m_{ij} is the number of people known by the i-th respondent in the j-th subpopulation of known size, e_j is the size of the subpopulation j and t the size of the general population T.

The standard error of the social network size estimator is:

$$s.e.(\hat{c}_i) = \sqrt{\frac{tc_i}{\sum_{j=1}^{L} e_j}}. \tag{0.4}$$

Selection of the known size subpopulations

The scale-up class of estimators suffers from some heavy assumptions that can lead to several problems in the estimates: (i) each subject in the general population T should have the same probability of knowing a subject in the subpopulations , (ii) everyone in T should know all about his/her acquaintances and (iii) the difficulty to recall in a short time all the people known in a certain subpopulation should be assumed to be negligible. The violation of these assumptions lead respectively to (i) barrier effect, (ii) transmission effect and (iii) estimation effect (Johnsen et al. 1995) which may affect the assumption of linear proportionality among e_j (the size of E_j) and $\bar{m}_j$ (the mean number of people known by respondents in the same subpopulation E_j).

Therefore, great attention should be paid to the choice of the subpopulations of known size, but often this is quite difficult because this choice depends on the availability of institutional data that usually refers to fields not friendly to respondents (illnesses, crime, etc.). Snidero et al. (Snidero, Morra, Corradetti, and Gregori 2005) assessed this linearity when the survey was already carried out using a regression model aimed at eliminating the subpopulations where the scale-up predictions for population size and the actual size of the populations are not correlated.

The exclusion of some population could be due to the barrier and transmission effects that can affect the estimates (e.g. it is most likely that a man knows the car brand than a woman and it is less likely to know people occupations than their illnesses).

In any case in principle, the exclusion these subpopulations could depend also on their position in the questionnaire and on how interviewers broach the subject (e.g.: voice tone, manners, etc.).

3. The Study Design

The survey was CATI based and 1081 Italian women aged 18-50 were interviewed during one week of March 2005. We choose to administer the questionnaire just to women because in Italy women spend much more time with their children and at least in principle it is more likely that they remember injuries happened to their children or to learn about injuries of other children. The CATI system randomly selected the telephone numbers which are also randomly assigned to interviewers.

The sample was stratified by age and province of residence. The definition of social network that we employed was the “active network”, i.e.: “Mutually recognize each other by sight or name, can be contacted, and have had contact within the last two

years, either in person, by phone or mail" (Bernard, Killworth, McCarty, and Shelley 1990; Killworth et al. 1998a; Killworth et al. 1998b).

In order to estimate the social network size we selected 33 subpopulations of known size from Census and other official sources. These subpopulations were classified in two subsets: the questions with low sensitive impact and those of high sensitive impact (see Table 1). We classified high sensitive questions following the definition given by Mangione (Mangione, Fowler, and Louis 1992), who defined sensitive items as those for which "giving a particular answer would make the respondent look better or would be more socially acceptable". As the questions referred to one's social network, we extended this definition also to items that touch directly the personal life of respondents (e.g.: the death for cancer of a friend, etc.).

The third subset of questions was formed by target questions which were: (i) "How many children do you know that have had an injury due to the ingestion of a foreign body?" (ii) "Of those, how many children went to the emergency services or were hospitalized?" (iii) "Of those, how many children ingested a toy?".

Then, the three subsets of questions (low sensitive, high sensitive and target questions) were used in all possible orders to form 6 questionnaires (see Table 2). For example, the questionnaire called LHT asked first the low sensitive questions, then the high sensitive questions and finally the target questions, while the questionnaire THL asked first the target questions, then the high sensitive questions and finally the low sensitive questions, etc. Each sub-sample counted about 175 people (Table 2). The questionnaires were administered one at a time and as interviewers works on three different shifts, some of the operators could differ from questionnaire to questionnaire.

Once the data were collected first we calculated the social network sizes of respondents (c_i) which were then used to get the estimates of subpopulations sizes.

The next step was to assess if the linearity among the average recalled number of people for each subpopulation and the subpopulation was held and eventually drop out the populations with low correlation.

The effects of the different interviewers and the questionnaire types on the interrupted and refused interviews and on incomplete fields were assessed. We considered as interviews with incomplete fields the interviews where at least one question was unanswered. Interviewers were identified through the id number they use to log into the survey system.

To measure the risk of having interrupted and refused interviews, and incomplete fields for each typology of questionnaire compared to the basic questionnaire, that should facilitate the disclosure of respondents and that is composed first by low sensitive questions, then by high sensitive and finally by target questions (LHT questionnaire), (see Table 2) we used a logistic regression.

To assess the interviewer effect we calculated the probability of having interviews interrupted, refused and with missing fields and then we compared each interviewer's probability with the probability averaged over all the interviewers.

The statistical analyses were carried out with R ver. 2.3.1. (The R Development Core Team 2006) and the libraries Hmsic and Design (Harrell 2005a; Harrell 2005b).

4. Results

Thirteen subpopulations of known size were eliminated for the absence of linear relation with the mean number of people recalled by respondents in the respective subpopulation. The R^2 of the regression model with all the subpopulations was 0.21,

while eliminating these 13 subpopulations through a graphical analysis of residuals, the resulting R^2 was 0.79. Eight out the 13 non selected subpopulations were classified as non sensitive (see Table 3).

The six different questionnaires' settings did not give statistically different results on the estimates of the respondents' network size (see Table 4). Also the estimates of the target subpopulations sizes (children that suffered a foreign body injury, children that were hospitalized consequently to a foreign body ingestion and injuries due to a toy) did not give significantly different results except for the TLH questionnaire (see Table 2). On the other hand the questionnaire THL, also starting with target questions but followed by high sensitive and low sensitive questions gave higher estimates compared to the other questionnaires even though not statistically significant (see Table 4).

There is not a significantly different risk of having interviews interrupted for each questionnaire as compared to the basic questionnaire (LHT questionnaire) (see Table 5). As regards the risk of having an interview refused, it emerged that there is a significantly, even though quite limited, higher probability of having interviews accepted in the questionnaire THL as compared to the basic questionnaire (Table 6).All the questionnaires have an increased risk of having incomplete fields as compared to the basic questionnaire (Table 7).

The mean probability of having interviews accepted for each operator among the accepted interviews (n=1398) is 25.9%. Ten interviewers have a higher probability compared to the mean of having an interview interrupted (Figure 1). The mean probability of having an interview refused is 82% and 9 interviewers have a lower probability of having an interview refused (Figure 2).The mean probability of having

an interview with incomplete fields is 59%. Twelve operators have a lower probability compared to the mean (Figure 3).

5. Discussion

Our analyses show that the risk of having incomplete fields is greater for all the questionnaire typologies as compared to what we defined as a basic questionnaire (low sensitive, high sensitive, target questions).

Only the survey with the questionnaire TLH (see Table 2) has a decreased risk of having interviews refused. We considered as refused an interview that was not accepted in the introductory phase, when the interviewer asks respondents if they want to participate to the survey. In the introductory phase, the core survey questions were not asked and therefore it is not possible that their order influences the answers and consequently the results. As some interviewers participated just to some questionnaires, we guess that interviewers' characteristics should represent the principal reason for the refusal of the interview.

The probability of having incomplete fields, interrupted and refused interviews for each interviewer is quite different among interviewers. This could be explained by the difference in approaching respondents by telephone, the voice tone and the different level of work experience. Van Tilburg (van Tilburg 1998) and Marsden's (Marsden 2003) studies were aimed at evaluating the interviewer effect in measuring the network size using name generators. Van Tilburg found a strong relationship between interviewers and the respondents' answers, while Marsden found a smaller influence.

A lot of work in the literature was also aimed at understanding the influence of question order in survey results: DeMoranville (DeMoranville and Bienstock 2003) studied the effects of question order on service quality measurement. In this study, as

in a large number of other studies, order effects were expressed as differences in means and correlations for specific and general questions. The results were focused in the changes of the placement of specific questions relative to general questions in the survey. Question order was found to influence service quality measurements: general questions had higher means when asked after specific questions than when asked before.

Ramirez (Ramirez and Straus 2006) analyzed two different approaches to question order in research on sensitive or criminal behavior: one approach had the questions ordered by topic and the other had the same items but two of the very severe violence items were moved from the beginning of the questionnaire. It emerged that respondents had a significantly higher disclosure administering the second questionnaire.

These studies highlighted the influence of question order in survey results. Our study shows that question order does not greatly influence the estimates except when administering the TLH (target, low sensitive, high sensitive questions) questionnaire which gave significantly lower estimates on the target subpopulation sizes. Interestingly, the effect of question order is not homogeneous between the case of estimating network sizes and estimating the target subpopulations. Indeed, the biggest network size is estimated by the sequence HLT, which corresponds to a toy population size of 3126 injuries, which is far from being the biggest estimate (obtained with the order LHT). In our study target questions refer to a particular typology of children accident and this could make feel respondents more uncomfortable than high sensitive questions.

Mangione's study (Mangione, Fowler, and Louis 1992) focused on the characteristics of questions that produced interviewer effects. He found that sensitivity and difficulty

of questions were not significantly related to interviewer effects. Another finding of his study was that in order to reduce interviewer effects, a researcher should design questions that minimize the need for interviewers to probe in order to produce a usable answer.

We feel that the scale-up questionnaire could not be compared to Mangione's situation due to the fact that the question composition is quite "rigid" in such a setting: the questions are of the type: "How many people do you know that…?" and it is difficult to change this question structure.

Therefore, even though some help could be given to them during the interview through the questionnaire design (e.g.: writing the social network definition in each screenshot, writing suggestions on what to say when putting the most sensitive questions, etc.), in these kind of CATI interviews we think that the interviewers should be well trained and learn how to probe in order to obtain the correct answer and then to have unbiased estimates.

Moreover, an advantage of scale-up interviews consists of the absence of difficulty for the operators in coding answers and/or entering open ended questions, because answers consist only of figures (i.e.: the size of respondents' network belonging to each subpopulation).

Some subpopulations were eliminated by the subpopulation selection algorithm. The non linearity between the relative size of such subpopulations and the mean size of acquaintance in the respective subpopulations could not be held both for the objective difficulty in recalling the correct number of people belonging to those subpopulations, for the barrier and transmission effect and for the wrong and/or not suitable way in which the interviewers put the questions. Excluded subpopulations could also suffer of transmission and barrier effect. Nevertheless, if a clear tendency is observed to

discard sub-populations of bigger size, this is not a sufficient requirement: indeed, to be discarded, a subpopulation should combine the characteristics of being related to a commonly observable social aspect and its perceived precision. This is for instance the case of question 33 (people reporting a robbery) and question 7 (families with 3 or more children), having almost the same size but being different in terms of perceived precision (number of children is intuitively known at a higher level of precision than the knowledge of an episode of robbery). The relationship between perceived precision of the question and its impact on reliability of the scale-up method is an unexplored aspect which might be worth further investigations.

In conclusion, in this study, the need emerged to deeply train the operators when they are requested to administer a questionnaire in a scale-up setting.

Study limitations

This study has several limitations. First, the results could be only generalized to CATI surveys about scale-up questionnaires in the field of injuries, since we are aware that the results could change with different target questions. The interactions among questionnaire type and operator where not analyzed as the interviewers did not work or administered all the questionnaires. Moreover, it would be of great interest to analyze also the influence of the characteristics (e.g.: age, gender, education, etc.) of each interviewer, which were not collected in this study due to refusal from the majority of interviewers to provide such data.

6. References

Bernard, H. R., E. C. Johnsen, P. D. Killworth, C. McCarty, G. A. Shelley, and S. Robinson. 1990. "Comparing four different methods for measuring personal social networks." *Social Networks* 12:179-215.

Bernard, H. R., E. C. Johnsen, P. D. Killworth, and S. Robinson. 1991. "Estimating the size of an average personal network and of an event subpopulation: some empirical results." *Social Science Research* 20:109-121.

Bernard, H. R., E. Johnsen, P. D. Killworth, and S. Robinson. 1989. "Estimating the size of an average personal network and of an event population." in *The small world*, edited by M. Kochen. Norwood, NJ: Ablex Publishing.

Bernard, H. R., P. D. Killworth, E. C. Johnsen, G. A. Shelley, and C. McCarty. 2001. "Estimating the Ripple Effect of a Disaster." *Connections* 24:18-22.

Bernard, H. R., P. D. Killworth, C. McCarty, and G. A. Shelley. 1990. "Estimating the size of personal networks." *Social Networks* 23:289-312.

Coromina, L. and G. Coenders. 2006. "Reliability and validity of egocentered network data collected via web - A meta-analysis of multilevel multitrait, multimethod studies." *Social Networks* 28:209-231.

Crespi, I. and D. Morris. 1984. "Question order effect and the measurement of candidate preference in the 1982 Connecticut elections." *Public Opinion Quarterly* 48:578-591.

Crossley, T. F. and S. Kennedy. 2002. "The reliability of self-assessed health status." *Journal of Health Economics* 21:643-658.

DeMoranville, C. W. and C. C. Bienstock. 2003. "Question order effects in measuring service quality." *International Journal of Research in Marketing* 20:217-231.

Freeman, L. C. and C. R. Thompson. 1989. "Estimating acquaintanceship volume." Pp. 147-158 in *The Small World*, edited by M. Kochen.

Fushs, M. 2000. "Screen design and question order in a CAI instrument results from a usability field experiment." *Survey Methodology* 26:199-207.

Gao, S. 2001. "Estimation of interviewer effects on multivariate binary responses in a community based survey." *Journal of Official Statistics* 17:149-161.

Harrell, F. E. 2005a. "The Design Package."

—. 2005b. "The Hmisc Package."

Johnsen, E. C., H. R. Bernard, P. D. Killworth, G. A. Shelley, and C. McCarty. 1995. "A social network approach to corroborating the number of AIDS/HIV victims in the US." *Social Networks* 17:167-187.

Kadushin, C., P. Killworth, H. R. Bernard, and A. Beveridge. 2006. "Scale-Up Methods as Applied to Estimates of Heroin Users." *Journal of Drug Issues* 6:417-440.

Killworth, P. D., H. R. Bernard, and C. McCarty. 1984. "Measuring patterns of acquaintenceship." *Current Anthropology* 23:381-397.

Killworth, P. D., E. C. Johnsen, H. R. Bernard, G. A. Shelley, and C. McCarty. 1990. "Estimating the size of personal networks." *Social Networks* 12:289-312.

Killworth, P. D., E. Johnsen, C. McCarty, G. A. Shelley, and H. R. Bernard. 1998a. "A social network approach to estimating seroprevalence in the United States." *Social Networks* 20:23-50.

Killworth, P. D., C. McCarty, H. R. Bernard, G. A. Shelley, and E. C. Johnsen. 1998b. "Estimation of seroprevalence, rape and homelessness in the U.S. using a social network approach." *Evaluation Review* 22:289-308.

Mangione, T. W., F. J. Fowler, Jr, and T. A. Louis. 1992. "Question characteristics and interviewer effects." *Journal of Official Statistics* 8:293-307.

Marsden, P. V. 2003. "Interviewer effects in measuring network size using a single name generator." *Social Networks* 25:1-16.

McCarty, C., P. D. Killworth, H. R. Bernard, E. C. Johnsen, and G. A. Shelley. 2000. "Comparing two methods for estimating network sizes." *Human Organization* 60:28-39.
McColl, E., M. P. Eccles, N. S. Rousseau, I. N. Steen, D. W. Parkin, and J. M. Grimshaw. 2003. "From the generic to the condition-specific? Instrument order effects in quality of life assessment." *Medical Care* 41:777-790.
McFarland, S. G. 1981. "Effects of question order on survey responses." *Public Opinion Quarterly* 45:208-215.
Moody, J. 2006. "Fighting a Hydra: A Note on the Network Embeddedness of the War on Terror." *Structure and Dynamics: eJournal of Anthropological and Related Sciences* 1:9.
Moy, P., D. A. Scheufele, W. P. Eveland, and J. M. McLeod. 2001. "Support for the death penalty and rehabilitation: Question order or communication effect?" *Journal of Applied Social Psychology* 31:2230-2255.
Pickery, J. and G. Loosveldt. 2001. "An exploration of question characteristics that mediate interviewer effects on item nonresponse." *Journal of Official Statistics* 17:337-350.
Pickery, J. and G. Loosveldt. 2000. "Modeling interviewer effects in panel surveys: An application." *Survey Methodology* 26:189-198.
—. 2004. "A simultaneous analysis of interviewer effects on various data quality indicators with identification of exceptional interviewers." *Journal of Official Statistics* 20:77-89.
Ramirez, I. L. and M. A. Straus. 2006. "The effect of question order on disclosure of intimate partner violence: An experimental test using the conflict tactics scales." *Journal of Family Violence* 21:1-9.
Rimal, R. N. and K. Real. 2005. "Assessing the perceived importance of skin cancer: How question-order effects are influenced by issue involvement." *Health Education & Behavior* 32:398-412.
Snidero, S., R. Corradetti, and D. Gregori. 2004. "The network scale-up method: a simulation study in case of overlapping sub-popualtions." *Metodoloski zvezki* 1:395-405.
Snidero, S., B. Morra, R. Corradetti, and D. Gregori. 2005. "Estimating the number of foreign bodies injuries in children with the scale-up method " Pp. 3586-3591 in *ASA Proceedings of the Joint Statistical Meetings*. Minneapolis: American Statistical Association (Alexandria, VA).
The R Development Core Team. 2006. "R."
van Tilburg, T. 1998. "Interviewer effects in the measurement of personal network size -- A nonexperimental study." *Sociological Methods & Research* 26:300-328.
Zheng, T., M. J. Salganik, and A. Gelman. 2006. "How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks
" *Journal of the American Statistical Association* 101:409-423.
Zigon, G., D. Gregori, R. Corradetti, B. Morra, L. Salerni, F. M. Passali, and D. Passali. 2006. "Child mortality due to suffocation in Europe (1980-1995): a review of official data." *ACTA Otorhinolaryngologica Italica* 26:154-161.

Id	Subpopulations of known size	Absolute size in thousands
	Low sensitive questions	
1	Bought a motorcycle	409
2	Own a Mercedes car	956
3	Competitive athletes members of FIDAL	127
4	Competitive basket players members of FIP	169
5	Families with 5 or more components	1635
6	People went in business during 2004	426
7	Families with 3 or more children	1276
8	People 100 or more years old	6
9	People doing a temporary job	119
10	People working in hotels and restaurants	859
11	People sentenced for driving under the influence of alcohol in 2004	426
12	Women had a child in 2004	528
13	People bought a new car	2249
14	People owing a BMW car	630
15	People owing a car with gas or methane	1356
16	Teachers	707
17	Teachers of primary school	236
18	People volunteer in non-profit associations	3481
19	Families with 2 children	4436
20	People with a post graduated speciality (master, medical specialization, etc.) or a PhD	644
	High sensitive questions	
21	Children adopted in 2004	6
22	People had a heart transplantation in 2004	0.3
23	People had a kidney transplantation in 2004	1.7
24	People committed suicide in 2004	3
25	People dead for a cancer in 2004	164
26	People currently detained in prison	57
27	People killed in 2004	0.7
28	People reported a rape in 2004	2
29	Widows younger than 60 years	506
30	Families with only one parent living alone with children	2101
31	People voted for Casa delle Libertà (Italian right party) in 2001 elections	18300
32	People voted for Rifondazione Comunista (Italian left party) in 2004 elections	1972
33	People reported a robbery	1303
	Target questions	
34	How many children do you know that had an injury due to the ingestion of a foreign body?	Unknown
35	Of those, how many children went to the emergency service or were hospitalized?	Unknown
36	Of those, how many children ingested a toy?	Unknown

Table 1 List of the subpopulations employed in the questionnaire. The subpopulations of known size are classified in sensitive and non sensitive questions.

Questionnaire name	First group	Second group	Third group	Sample size
1 – LHT (basic questionnaire)	Low sensitive	High sensitive	Target	175
2 - HLT	High sensitive	Low sensitive	Target	174
3 - THL	Target	High sensitive	Low sensitive	201
4 - TLH	Target	Low sensitive	High sensitive	174
5 - HTL	High sensitive	Target	Low sensitive	176
6 - LTH	Low sensitive	Target	High sensitive	181

Table 2 The composition of the six different questionnaires and the size of each sub-sample. The six questionnaires are formed by mixing in all the possible way the three groups of questions. The LHT questionnaire (low sensitive, high sensitive and target questions) is the basic questionnaire.

Id	Excluded subpopulations	Subpopulation type
1	People bought a new car	Non sensitive
2	People owing a BMW car	Non sensitive
3	People owing a car with gas or methane	Non sensitive
4	Teachers	Non sensitive
5	Teachers of primary school	Non sensitive
6	People volunteer in non-profit associations	Non sensitive
7	People with a post graduate speciality (master, medical specialization, etc.) or a PhD	Non sensitive
8	Widows younger than 60 years	Sensitive
9	Families with only one parent living alone with children	Sensitive
10	Families with 2 children	Non sensitive
11	People voted for Casa delle Libertà (Italian right party) in 2001 elections	Sensitive
12	People voted for Rifondazione Comunista (Italian left party) in 2004 elections	Sensitive
13	People reported a robbery	Sensitive

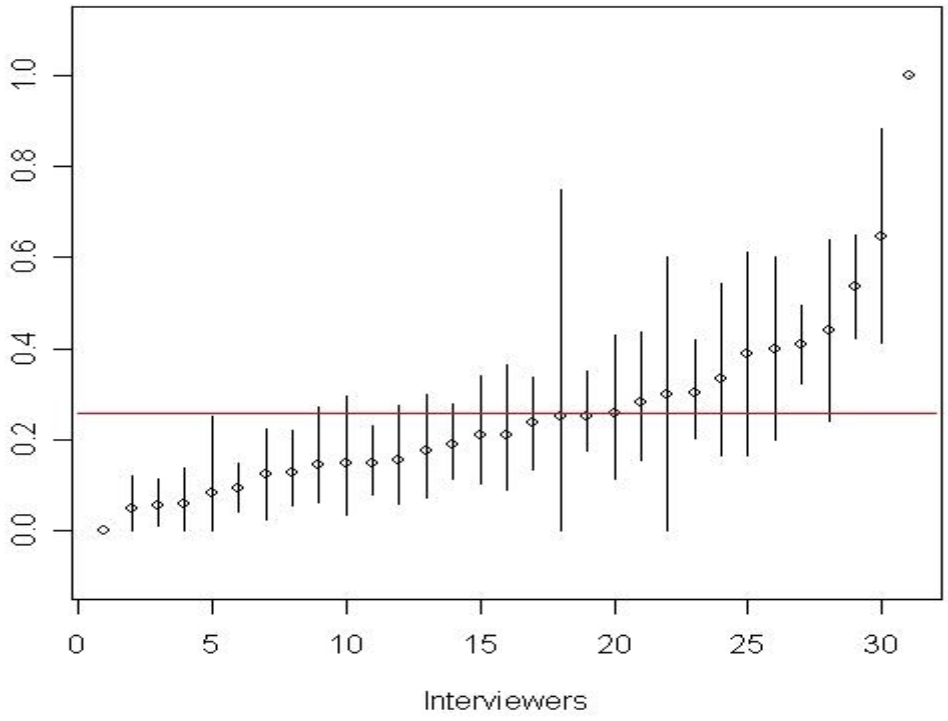
Table 3 Subpopulations eliminated by the selection algorithm. The sizes of these subpopulations were not in linear relation with the mean number of people in each respective subpopulation recalled by respondents

Q	Mean ĉ (SE)	FB injury (SE)	Hospitaliz. (SE)	Toys (SE)
Overall	218.3 (38.7)	15829.0 (741.3)	12844.0 (667.7)	4061.0 (375.5)
1 – LHT	203.2 (38.0)	14974.6 (1857.4)	11519.0 (1629.0)	4607.6 (1030.3)
2 – HLT	256.0 (41.2)	17286.9 (1783.0)	16000.0 (1715.3)	3126.4 (758.3)
3 – THL	243.1 (40.1)	18603.4 (1765.8)	15419.0 (1607.5)	6871.5 (1073.2)
4 – TLH	205.5 (38.1)	11225.0 (1603.6)	6643.4 (1233.6)	1374.5 (561.1)
5 – HTL	213.8 (38.7)	15249.0 (1796.0)	12911.0 (1653.2)	4233.3 (946.6)
6 – LTH	184.9 (36.0)	16359.0 (2029.1)	12835.0 (1797.3)	3271.8 (907.4)

Table 4 Mean social network estimate (standard error), the estimate of the number of children that suffered a foreign body injury (standard error), the estimate of the number of hospitalized children (standard error) and the estimate of the number of injuries due to toys (standard error) per questionnaire type and in the overall survey. LHT is the basic questionnaire.

Questionnaire type	**Interrupted** (N=317)	**NonInterrupted** (N=1081)	**OR (95% C.I.)**
1 – LHT	16% (51)	16% (175)	Ref
2 – HLT	16% (51)	16% (174)	1.01 (0.65, 1.56)
3 – THL	18% (58)	19% (201)	0.99 (0.65, 1.52)
4 – TLH	12% (39)	16% (174)	0.77 (0.48, 1.23)
5 – HTL	18% (57)	17% (181)	1.08 (0.70, 1.66)
6 - LTH	19% (61)	16% (176)	1.19 (0.78, 1.82)

Table 5 Interrupted interviews by questionnaire type. Percentage (numbers), OR (95% confidence interval). Ref is the reference category and LHT is the basic questionnaire.



ity,

Questionnaire Type	Refused (N=8507)	Accepted (N=1398)	OR (95% C.I.)
1 – LHT	17% (1441)	16% (226)	ref
2 – HLT	17% (1460)	16% (225)	0.98 (0.81, 1.20)
3 – THL	16% (1357)	19% (259)	1.22 (1.00, 1.48)
4 – TLH	17% (1449)	15% (213)	0.94 (0.77, 1.15)
5 – HTL	16% (1395)	17% (238)	1.09 (0.89, 1.32)
6 - LTH	17% (1405)	17% (237)	1.08 (0.88, 1.31)

Table 6 Refused interview by questionnaire type. Percentage (numbers), OR (95% confidence interval). Ref is the reference category and LHT is the basic questionnaire.

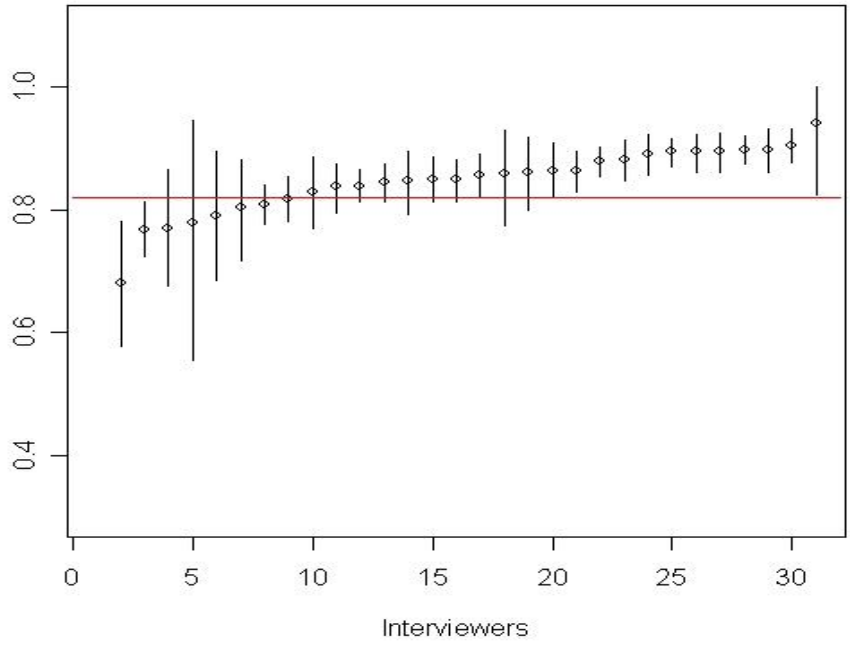


Figure 2 Interviewers' probability to have a refused interview ((overall probability, represented by the horizontal line: 0.82 vertical lines represents 95% confidence intervals))

Questionnaire Type	Completed Fields (N=385)	Incompleted Fields (N=696)	OR (95% C.I.)
1 – LHT	23% (88)	12% (87)	ref
2 – HLT	16% (61)	16% (113)	1.87 (1.22, 2.88)
3 – THL	14% (55)	21% (146)	2.69 (1.75, 4.12)
4 – TLH	15% (59)	17% (115)	1.97 (1.28, 3.04)
5 – HTL	14% (52)	19% (129)	2.51 (1.62, 3.89)
6 – LTH	18% (70)	15% (106)	1.53 (1.00, 2.34)

Table 7 Incomplete fields by questionnaire type. Percentage (numbers), OR (95% confidence interval). Ref is the reference category and LHT is the basic questionnaire.

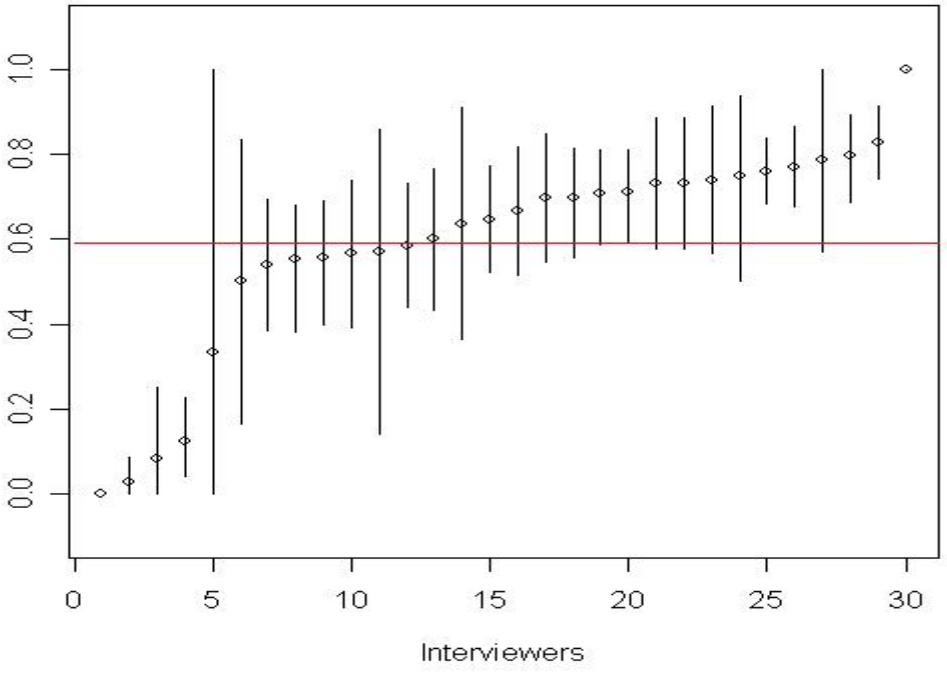


Figure 3 Interviewers' probability to have interviews with incompleted fields (overall probability, represented by the horizontal line: 0.59 vertical lines represents 95% confidence intervals))