Sociological Methods & Research

http://smr.sagepub.com

Question Order and Interviewer Effects in CATI Scale-up Surveys

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

Sociological Methods Research 2009; 38; 287 originally published online Oct 8, 2009;
DOI: 10.1177/0049124109346163

The online version of this article can be found at:
http://smr.sagepub.com/cgi/content/abstract/38/2/287

Published by:
SAGE
http://www.sagepublications.com

Sociological Methods & Research
38(2) 287–305

DOI: 10.1177/0049124109346163
http://smr.sagepub.com

Question Order and Interviewer Effects in CATI Scale-up Surveys

Silvia Snidero[1,2], Federica Zobec[2], Paola Berchialla[3], Roberto Corradetti[1], and Dario Gregori[4]

Abstract

The scale-up estimator is a network-based estimator for the size of hidden or hard to count subpopulations. Several issues arise in the public health context when the aim is the estimation of injuries occurring in a certain population, where two common problems are present: (a) Small injuries are usually difficult to observe and rarely reported in the official data and (b) people are not always compliant in giving information about some specific injuries, in particular when children are involved. This study checked the methodological issues arising from using a computer-assisted telephone interview (CATI) survey using the scale-up methodology for detecting the number of injuries due to choking in children ages 0 to 14 in Italy. For this purpose, 1,000 CATI interviews were conducted during a week using a questionnaire based on 33 questions about populations of known size according to census data. Then, each respondent was asked about other questions related to the main target population (e.g., number of children

[1]Department of Statistics and Applied Mathematics, University of Torino, Italy
[2]S&A S.r.l. - Surveys & Analyses, Torino, Italy
[3]Department of Public Health and Microbiology, University of Torino, Italy
[4]Laboratory of Epidemiological Methods and Biostatistics, Department of Environmental Medicine and Public Health, University of Padova

Corresponding Author:
Dario Gregori, MA, PhD, Laboratory of Epidemiological Methods and Biostatistics, Department of Environmental Medicine and Public Health University of Padova,
35121 Padova, Italy, Phone: +39 049 8215384, Fax: +39 02 700445089
E-mail: dario.gregori@unipd.it

known to suffer from a choking accident). A sensitivity analysis was conducted for estimating the effect of varying subpopulations, order of the questions, and interviewer effects on the resulting estimates. For the interviewer effect, no particular differences were observed in the overall estimates of injuries. The conclusion is the scale-up estimator in association with CATI methodology shows 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

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 et al. (1989, 1991) 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 et al. 1989, 1991; Johnsen et al. 1995; Killworth, Johnsen, et al. 1998; Killworth, McCarty et al. 1998). In the previous equation, the only unknown variable is c, the average social network size; thus, it should be substituted with a good estimate (Bernard et al. 1989, 1991; Killworth, Johnsen et al. 1998; Killworth, McCarty et al. 1998; 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 et al. 2003; McFarland 1981; Moy et al. 2001; Pickery and Loosveldt 2000, 2001, 2004; Ramirez and Straus 2006; Rimal and Real 2005; van Tilburg 1998). Among these works, only a few dealt with the problems of question order and interviewer effects in the social network field: The van Tilburg (1998) and Marsden (2003) 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 ego-centered network data collected via the Web. The scale-up method case is quite different from the name generator estimator because 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 does not suffer 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 that 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 usually comes 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 et al. 2005). Although this method gained a big advantage from indirect questioning, avoiding speaking about one's own sensitive matters, some questions could in any case make respondents feel uncomfortable (Bernard et al. 2001; Johnsen et al. 1995; Kadushin et al. 2006; Killworth, Johnsen, et al. 1998; Killworth, McCarty, et al. 1998; Moody 2006; Snidero et al. 2004). Therefore, the need emerged to understand where it is better to place the sensitive questions (e.g., after some nonsensitive questions that may make respondents feel more comfortable and avoid therefore some possible bias?) and how the questionnaire should be administered (e.g., Are more gentle and less determined interviewers more likely to obtain answers on sensitive questions? Is giving more time to respondents to recall the number of people known in a certain population better than quickening the delivery of questions? Is it helpful to avoid distortion to remind respondents several times during the interview of 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 ages 0 to 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 younger than 4 years of age (Zigon et al. 2006).

In the next section, we briefly describe the scale-up method and the social network estimator. Then, we introduce the study design and the questionnaire. Finally, we present the results and a discussion.

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 and colleagues (Killworth, Johnsen, et al. 1998; Killworth, McCarty, et al. 1998) developed a maximum likelihood estimator that was very simple to calculate:

$$\hat{e}_0 = t\frac{\sum_i m_{i0}}{\sum_i c_i}, \tag{1}$$

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

The scale-up estimator is unbiased, and its standard error is (Killworth, Johnsen, et al. 1998; Killworth, McCarty, et al. 1998)

$$s.e.(\hat{e}_0) = \sqrt{\frac{t \cdot \hat{e}_0}{\sum_i c_i}}. \tag{2}$$

In the past years, several social network size estimators were studied (Bernard et al. 1989, 1991; Bernard, Johnsen, et al. 1990; Bernard, Killworth, et al. 1990; Freeman and Thompson 1989; Johnsen et al. 1995; Killworth, Bernard, and McCarty 1984; Killworth et al. 1990; McCarty et al. 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 chose an estimator belonging to a class of estimators that underlies the same idea of the scale-up method, which is the proportional estimator:

$$\hat{c}_i = t \cdot \frac{\sum_{j=1}^{L} m_{ij}}{\sum_{j=1}^{L} e_j}, \tag{3}$$

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

The standard error of the social network size estimator is

$$s.e.(\hat{c}_i) = \sqrt{\frac{tc_i}{\sum_{j=1}^{L} e_j}}. \tag{4}$$

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: (a) Each subject in the general population T should have the same probability of knowing a subject in the subpopulations, (b) everyone in T should know all about his or her acquaintances, and (c) 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 (a) barrier effect, (b) transmission effect, and (c) 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 refer to fields not friendly to respondents (illnesses, crime, etc.). Snidero et al. (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's occupations than their illnesses).

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

The Study Design

The survey was computer-assisted telephone interview (CATI) based and 1,081 Italian women aged 18 to 50 were interviewed during one week of March 2005. We chose 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 that 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,'' namely, ''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, et al. 1990; Killworth, Johnsen, et al. 1998; Killworth, McCarty, et al. 1998).

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, Fowler, and Louis (1992:296), 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 of a friend from cancer, etc.).

The third subset of questions was formed by target questions that were (a) ''How many children do you know that have had an injury due to the ingestion of a foreign body?'' (b) ''Of those, how many children went to the emergency services or were hospitalized?'' and (c) ''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 six questionnaires (see Table 2). For example, the questionnaire called LHT asked first the low sensitive questions (L), then the high sensitive questions (H), and finally the target questions (T), while the questionnaire THL asked first the

Table 1. List of the Subpopulations Employed in the Questionnaire (the Subpopulations of Known Size Are Classified in Sensitive and Nonsensitive Questions)

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 basketball? players members of FIP	169
5	Families with five or more components	1,635
6	People went in business during 2004	426
7	Families with three or more children	1,276
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 who had a child in 2004	528
13	People who bought a new car	2,249
14	People owning a BMW car	630
15	People owning a car running on gas or methane	1,356
16	Teachers	707
17	Teachers of primary school	236
18	People who volunteer in nonprofit associations	3,481
19	Families with two children	4,436
20	People with a postgraduated specialty (master, medical specialization, etc.) or a PhD	644
High sensitive questions		
21	Children adopted in 2004	6
22	People who had a heart transplantation in 2004	0.3
23	People who had a kidney transplantation in 2004	1.7
24	People who committed suicide in 2004	3
25	People dead from cancer in 2004	164
26	People currently detained in prison	57
27	People killed in 2004	0.7
28	People who reported a rape in 2004	2
29	Widows younger than 60 years	506
30	Families with only one parent living alone with children	2,101
31	People who voted for Casa delle Libertà (Italian right party) in 2001 elections	18,300

(continued)

Table 1. (continued)

ID	Subpopulations of known size	Absolute size in thousands
32	People who voted for Rifondazione Comunista (Italian left party) in 2004 elections	1,972
33	People who reported a robbery	1,303
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

target questions, then the high sensitive questions, and finally the low sensitive questions, and so on. Each subsample counted about 175 people (Table 2).

The questionnaires were administered one at a time, and because interviewers work 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 to 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 interviews with incomplete fields as the interviews where at least one question was unanswered. Interviewers were identified through the ID number they use to log into the survey system.

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

To assess the interviewer effect, we calculated the probability of having interviews interrupted, refused, and with missing fields, and then we

Table 2. The Composition of the Six Different Questionnaires and the Size of Each Subsample

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

The six questionnaires are formed by mixing in all the possible ways the three groups of questions. The LHT questionnaire (low sensitive, high sensitive, and target questions) is the basic questionnaire.

compared each interviewer's probability with the probability averaged across all the interviewers.

The statistical analyses were carried out with R ver. 2.3.1. and the libraries Hmsic and Design.

Results

A total of 13 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, whereas after eliminating these 13 subpopulations through a graphical analysis of residuals, the resulting R^2 was 0.79. Of the 13 nonselected subpopulations, 8 were classified as nonsensitive (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 who suffered a foreign body injury, children who 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, even though not statistically significant, compared to the other questionnaires. (see Table 4).

There is not a significantly different risk of having interviews interrupted for each questionnaire as compared to the basic questionnaire (LHT

Table 3. Subpopulations Eliminated by the Selection Algorithm

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

The sizes of these subpopulations were not in linear relation with the mean number of people in each respective subpopulation recalled by respondents.

questionnaire; see Table 5). Regarding 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

Table 4. Mean Social Network Estimate, the Estimate of the Number of Children That Suffered a Foreign Body Injury, the Estimate of the Number of Hospitalized Children, and the Estimate of the Number of Injuries due to Toys per Questionnaire Type and in the Overall Survey (with Standard Errors)

Questionnaire	Mean social network estimate	*SE*	Foreign body injury	*SE*	Hospitalizd	*SE*	Toys	*SE*
Overall	218.3	38.7	15,829.0	741.3	12,844.0	667.7	4,061.0	375.5
1–LHT	203.2	38.0	14,974.6	1,857.4	11,519.0	1,629.0	4,607.6	1,030.3
2–HLT	256.0	41.2	17,286.9	1,783.0	16,000.0	1,715.3	3,126.4	758.3
3–THL	243.1	40.1	18,603.4	1,765.8	15,419.0	1,607.5	6,871.5	1,073.2
4–TLH	205.5	38.1	11,225.0	1,603.6	6,643.4	1,233.6	1,374.5	561.1
5–HTL	213.8	38.7	15,249.0	1,796.0	12,911.0	1,653.2	4,233.3	946.6
6–LTH	184.9	36.0	16,359.0	2,029.1	12,835.0	1,797.3	3,271.8	907.4

LHT (low sensitive, high sensitive, and target questions) is the basic questionnaire.

Table 5. Interrupted Interviews by Questionnaire Type

Questionnaire type	Interrupted (*N* = 317)		Noninterrupted (*N* = 1,081)		Odds ratio	95% confidence interval
	%	*n*	%	*n*		
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

Ref is the reference category and LHT (low sensitive, high sensitive, and target questions) is the basic questionnaire.

Table 6. Refused Interview by Questionnaire Type

Questionnaire type	Refused (*N* = 8,507)		Accepted (*N* = 1,398)		Odds ratio	95% confidence interval
	%	*n*	%	*n*		
1–LHT	17	1,441	16	226	Ref	
2–HLT	17	1,460	16	225	0.98	0.81, 1.20
3–THL	16	1,357	19	259	1.22	1.00, 1.48
4–TLH	17	1,449	15	213	0.94	0.77, 1.15
5–HTL	16	1,395	17	238	1.09	0.89, 1.32
6–LTH	17	1,405	17	237	1.08	0.88, 1.31

Ref is the reference category and LHT (low sensitive, high sensitive, and target questions) is the basic questionnaire.

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 ($n = 1{,}398$) is 25.9 percent. Of the interviewers, 10 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 percent, and 9 interviewers have a lower probability of having an interview refused (Figure 2). The mean probability of having an interview

Table 7. Incomplete Fields by Questionnaire Type

Questionnaire type	Completed fields ($N = 385$)		Incompleted fields ($N = 696$)		Odds ratio	95% confidence interval
	%	*n*	%	*n*		
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

Ref is the reference category and LHT (low sensitive, high sensitive, and target questions) is the basic questionnaire.

with incomplete fields is 59 percent; 12 operators have a lower probability compared to the mean (Figure 3).

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 one that was not accepted in the introductory phase, when the interviewer asks respondents if they want to participate in 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. Because some interviewers participated with just 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, by the voice tone, and by the different levels of work experience. The van Tilburg's (1998) and Marsden's (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.

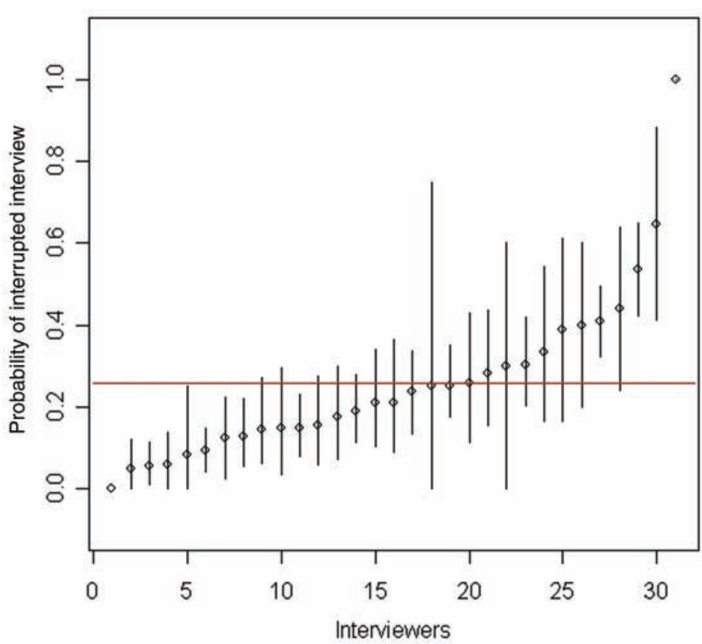

Figure 1. Probability for an interviewer to have an interrupted interview
Overall probability, represented by the horizontal line: 0.259; vertical lines represent 95 percent confidence intervals.

A lot of work in the literature was also aimed at understanding the influence of question order in survey results: 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 on 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 and Straus (2006) analyzed two 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

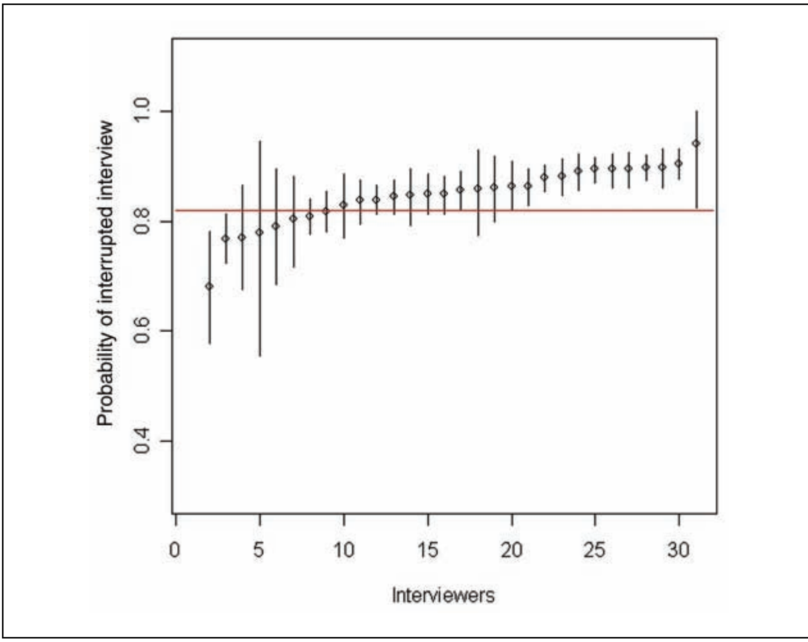

Figure 2. Interviewers' probability to have a refused interview
Overall probability, represented by the horizontal line: 0.82; vertical lines represent 95 percent confidence intervals.

emerged that respondents had a significantly higher disclosure when administered 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 3,126 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 accidents and this could make respondents feel more uncomfortable than high sensitive questions.

Mangione et al.'s (1992) study focused on the characteristics of questions that produced interviewer effects. They found that sensitivity and difficulty

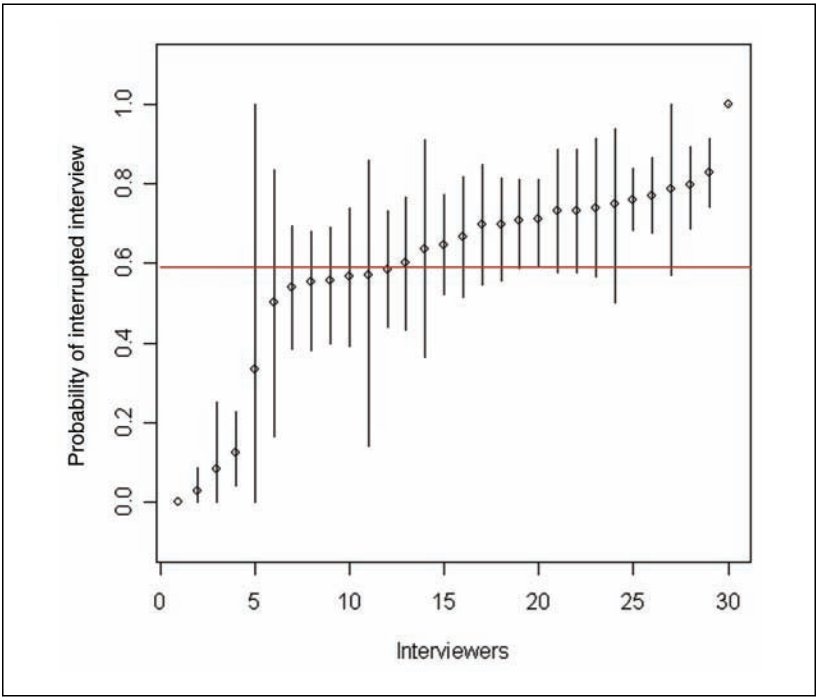

Figure 3. Interviewers' probability to have interviews with incompleted fields
Overall probability, represented by the horizontal line: 0.59; vertical lines represent 95 percent confidence intervals.

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

We feel that the scale-up questionnaire could not be compared to Mangione et al.'s (1992) 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 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 nonlinearity between the relative size of such subpopulations and the mean size of acquaintance in the respective subpopulations could not be held 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 unsuitable way in which the interviewers put the questions. Excluded subpopulations could also suffer of transmission and barrier effect. Nevertheless, if a clear tendency to discard subpopulations of bigger size is observed, 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 three 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 that might be worth further investigation.

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 because we are aware that the results could change with different target questions. The interactions among questionnaire type and operator were not analyzed as the interviewers did not work or administer 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.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

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. Johnsen, P. D. Killworth, and S. Robinson. 1989. ''Estimating the Size of an Average Personal Network and of an Event Population.'' Pp. 159-79 in *The Small World*, edited by M. Kochen. Norwood, NJ: Ablex.

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* 20:109-21.

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* 28:209-31.

Crespi, I. and D. Morris. 1984. ''Question Order Effect and the Measurement of Candidate Preference in the 1982 Connecticut Elections.'' *Public Opinion Quarterly* 48:578-91.

Crossley, T. F. and S. Kennedy. 2002. ''The Reliability of Self-assessed Health Status.'' *Journal of Health Economics* 21:643-58.

DeMoranville, C. W. and C. C. Bienstock. 2003. ''Question Order Effects in Measuring Service Quality.'' *International Journal of Research in Marketing* 20:217-31.

Freeman, L. C. and C. R. Thompson. 1989. ''Estimating Acquaintanceship Volume.'' Pp. 147-58 in *The Small World*, edited by M. Kochen. Norwood, NJ: Ablex.

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* 17:149-61.

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* 17:167-87.

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* 6:417-40.

Killworth, P. D., H. R. Bernard, and C. McCarty. 1984. ''Measuring Patterns of Acquaintanceship.'' *Current Anthropology* 23:381-97.

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. 1998. ''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. 1998. ''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* 41:777-90.

McFarland, S. G. 1981. ''Effects of Question Order on Survey Responses.'' *Public Opinion Quarterly* 45:208-15.

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(2). Retrieved August 7, 2009 (http://repositories.cdlib.org/imbs/socdyn/sdeas/vol1/iss2/art9/).

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* 31:2230-55.

Pickery, J. and G. Loosveldt. 2000. ''Modeling Interviewer Effects in Panel Surveys: An Application.'' *Survey Methodology* 26:189-98.

Pickery, J. and G. Loosveldt. 2001. ''An Exploration of Question Characteristics that Mediate Interviewer Effects on Item Nonresponse.'' *Journal of Official Statistics* 17:337-50.

Pickery, J. and G. Loosveldt. 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-populations.’’ *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. 3586-91 in *ASA Proceedings of the Joint Statistical Meetings*. Minneapolis: American Statistical Association.

van Tilburg, T. 1998. ‘‘Interviewer Effects in the Measurement of Personal Network Size—A Nonexperimental Study.’’ *Sociological Methods & Research* 26:300-28.

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* 101:409-23.

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* 26:154-61.

Bios

Silvia Snidero is a PhD student in Risk Management at the University of Torino, Department of Statistics and Applied Mathematics, working in the field of injury prevention and risk assessment.

Federica Zobec is a supervisor of the CATI system and interview conducting management.

Paola Berchialla is a researcher at the Department of Public Health and Microbiology at the University of Torino, working in the field of risk assessment

Roberto Corradetti is an associate professor in Economic Statistics, working in the field of risk management.

Dario Gregori is an associate professor of biostatistics, coordinator of the European Registry of Foreign Bodies Injuries (Susy Safe)