Estimating the Number of Foreign Bodies Injuries in Children with the Scale-up Method

Silvia Snidero[1], Bruno Morra[2], Roberto Corradetti[1], and Dario Gregori[3]
Dept. of Statistics and Applied Mathematics, University of Torino[1]
Dept. of Otorhinolaryngology, "S. Giovanni Battista" Hospital, Torino[2]
Dept. of Public Health and Microbiology, University of Torino[3]

Abstract

The foreign body injury in the upper aero-digestive ways is a rare but not negligible event. Available data come from official discharge records and death certificates. Thus, self-resolved injuries are usually not recorded, resulting in lost at observation.
The scale-up method estimates the size of hidden or hard to count subpopulations. The estimates are obtained asking respondents the number of people they know in subpopulations of known size and in unknown target subpopulations. The aim of this study is to assess if this method is suitable for estimating the number of all FB injuries. Moreover, we set up an algorithm for the choice of subpopulations to use in estimates.

Keywords: Scale-up method, non-probabilistic sampling, small prevalence, foreign body, children

1. Introduction

Suffocation due to foreign bodies (FB) is a rare but not negligible event since a recent estimated number of choking accidents per year in children aged 0-14 is in the European Union of about 50000, 10% of which are fatal (Zigon, Gregori et al. 2004).
These estimates are based on hospital discharge records and death certificates. Actually, the major part of FB injuries is self-resolved (e.g. with the parents' help) and therefore lost at observation. Due to the importance of this phenomenon we tried to get a reasonable estimate also of the non hospitalized part of it. Common methods of probabilistic sampling are quite inadequate for this aim and better results are obtained from non probabilistic sampling schemes (Morrison and Stone 2000; Guard and Gallagher 2005). Our proposal is to apply the scale-up estimator to this issue. The scale-up method is a methodology capable of estimating the size of subpopulations of unknown size with smaller samples and perhaps with lower costs.
We do not have any empirical way to assess directly if the application of scale-up estimators to the number of self reported injuries would lead to reliable results. Therefore, we decided, for the purposes of understanding the correctness and efficiency of the estimator, to apply it to the estimation of hospitalizations for such injuries. Thus, the otorhinolaryngologists of the Piemonte region –a North-West Italian region- were interviewed about the number of injuries they have observed having as a consequence an hospitalization. The estimates obtained were then compared with those derived from the official discharge records of the Piemonte region referring to the same period of time.
In the next section, the scale-up method and the social network size estimator are introduced, in the 3rd section it is described the study design, in the 4th section the estimates are presented and the results are described.

2. The Scale-Up Method Estimator

The estimators belonging to the scale-up method is a novel class of estimators born from an idea of Bernard and his collaborators in the 90s (Bernard, Johnsen et al. 1989; Bernard, Johnsen et al. 1991; Johnsen, Bernard et al. 1995; Killworth, Johnsen et al. 1998; Killworth, McCarty et al. 1998; Bernard, Killworth et al. 2001), aimed at estimating the size of hidden or hard to count populations. This method is based on the concept of social network, which is the set of people known by each person; obviously the width of the network is strictly dependent from the definition used for "knowing" someone. Besides the difficulties on defining exactly this concept, this estimator has two main advantages: (i) people are asked indirectly about problems; (ii) and consequently the samples used are smaller as compared to those obtained using common estimators for events with small prevalence (Killworth, Johnsen et al. 1998; Snidero, Corradetti et al. 2004).
Respondents are not asked directly about problems but on how many people he/she knows with the characteristic under study. This method is very useful when you manage sensitivity arguments, and it was used for the first time to estimate the number of HIV+ people in U.S.A., a population with a small prevalence,

difficult to reach and for which do not exist official data (Killworth, Johnsen et al. 1998).

This method assumes a total population T of size t and a subpopulation E of T with size e; the basic assumption underlying the scale-up method is

$$\frac{m}{c} = \frac{e}{t} \quad (1)$$

where m is the mean number of persons known in E and c is the mean social network size of the members of T. Hence, the proportion of subjects in E known to each member of T is the same as the proportion of members of E belonging to general population T, i.e. $\frac{e}{t}$ (Bernard, Johnsen et al. 1989; Bernard, Johnsen et al. 1991; Johnsen, Bernard et al. 1995; Killworth, Johnsen et al. 1998; Killworth, McCarty et al. 1998).

Besides the size of target subpopulation also the social network size of respondets is unknown. Therefore we used one of the several estimators for the social network size. Some of these propose to estimate c using the number of persons known by each respondent in several subpopulations of known size (Bernard, Johnsen et al. 1989; Freeman and Thompson 1989; Bernard, Johnsen et al. 1990; Bernard, Killworth et al. 1990; Bernard, Johnsen et al. 1991; Killworth, Johnsen et al. 1998; McCarty, Killworth et al. 2000). Thus, the basic idea is to ask to respondents how many people they know in the target subpopulation of unknown size and how many people they know in a certain number of subpopulation of known size. For example, the respondent could be asked: "How many people do you know who is seropositive?" (the unknown size subpopulation) and "How many people do you know owing a swimming pool?"(the known size group).

The scale-up method has also several strong assumptions which can lead to several problems in estimations: (i) each subject T has the same probability to know a person in subgroup E, (ii) everyone in T knows all about his/her acquaintance and (iii) the difficulty to recall in short time people known in certain subpopulation is negligible (Bernard, Johnsen et al. 1989; Bernard, Johnsen et al. 1991; Johnsen, Bernard et al. 1995; Killworth, Johnsen et al. 1998).

The violation of these assumptions can conduct to some problems called the barrier, transmission and estimation effects. The barrier effect is due to some social and geographical characteristics that create a barrier in knowing some specific groups of persons. The transmission effect faces us when the information about a person is not transmitted with the same probability to his/her social network and finally the difficulty of recalling people belonging to a subpopulation can lead to the estimation effect (Bernard, Johnsen et al. 1989; Bernard, Johnsen et al. 1991; Johnsen, Bernard et al. 1995; Killworth, Johnsen et al. 1998).

Concluding, the goal of this study was to test the scale-up method for a future use for estimating the total amount (i.e. also the cases not hospitalized) of foreign body injury in Italy.

The second goal was to find a technique for avoiding the problems of barrier and transmission effect that affect the estimates.

3. Scale-up Method Estimator

The scale-up estimator is an extensions of the basic model described before.

Assuming that E_0 is a subpopulation of T of unknown size e_0, then the scale-up estimator (Killworth, Johnsen et al. 1998; Killworth, McCarty et al. 1998) of e_0 is a maximum likelihood estimator obtained by maximizing the probability:

$$\text{Prob}(c_i, m_{i0}, e_0) = \prod_i \binom{m_{i0}}{c_i} p_0^{m_{i0}} (1-p_0)^{c_i - m_{i0}}$$

(where $p_0 = \frac{e_0}{t}$), by varying e_0, and we find:

$$\hat{e}_0 = t \frac{\sum_i m_{i0}}{\sum_i c_i} \qquad (0.2)$$

This estimator is unbiased, i.e.: with $\mathrm{E}(\hat{e}_0) = e_0$ and standard error given by

$$s.e.(\hat{e}_0) = \sqrt{\frac{t \cdot \hat{e}_0}{\sum_i c_i}}. \qquad (0.3)$$

The scale-up estimator requires simply computing the ratio of the sum of m_{i0} over all respondents and the sum of c_i social networks size over all respondents.

4. Estimation of Social Network Size

As the respondent's social network size is unknown we have to substitute c_i with a good estimate of it.

Two of the several methods aimed at estimating the social network sizes employ subpopulations of known size (Killworth, Johnsen et al. 1998). The first is the *subgroup estimator*, which can be obtained by maximizing:

$$\text{Prob}\left(i \text{ knows } m_{ij},\ j = 1,2,...,L\right) = P_{c_i} = \prod_{j=1}^{L} \binom{m_{ij}}{c_i} p_j^{m_{ij}} q_j^{c_i - m_{ij}} \qquad (0.4)$$

where m_{ij} are the number of persons known by respondent i in the j-th subpopulation of known size

$(j = 1,...,L)$ and $p_j = \frac{e_j}{t}$ is the fraction of subpopulation E_j in T.

The network size estimator we used in our analysis is the *proportional estimator*

$$\hat{c}_i = t \cdot \frac{\sum_{j=1}^{L} m_{ij}}{\sum_{j=1}^{L} e_j} \quad (0.5)$$

with standard error

$$s.e.(\hat{c}_i) = \sqrt{\frac{tc_i}{\sum_{j=1}^{L} e_j}}. \quad (0.6)$$

Killworth et al. (Killworth, 1998a) proved that for small values of $\frac{m_{ij}}{c}$ and p_j, the estimate of c_i is almost the same of that obtained by the *subgroup estimate*, result also corroborate by our simulations (Snidero, Corradetti et al. 2004).

Each estimate of c_i has an error that is transmitted to the unknown subpopulation size estimate. It is proven (Killworth, Johnsen et al. 1998; Killworth, McCarty et al. 1998) that the effect of this error on estimation is negligible if $\sum_{ij} m_{ij}$ is sufficiently large.

Moreover, some previous simulations (Snidero, Corradetti et al. 2004) show that unbiased estimates and low standard errors do not depend from the number of subpopulations of known size.

5. The Study Design

The aim of this study is to estimate the number of foreign body (FB) injuries in children aged 0-14 in Piemonte -an Italian region- for the years 1999-2000-2001 with the scale-up estimator and then to compare the results with the official discharge records.

Therefore, all otorhinolaryngologists of the Piemonte Region were requested to fill in a questionnaire about:

- the number of people they know in 25 subpopulations of known size
- the number of children they remember were hospitalized in their institutions for choking injuries in the years 1999-2000-2001.

The target subpopulation question was: "How many hospitalizations were made in your institution for foreign bodies in the aero-digestive ways in the years 1999, 2000 and 2001?". The definition of social network used in this survey is the so called *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, Johnsen et al. 1990; Killworth, Johnsen et al. 1998).

We carried out a pre-test pilot study in order to choose those subpopulations of known size, which are, among the other, most suitable for the target of the interviews (ORLs). Thus, 8 names and 12 pairs of subpopulations from Census were chosen for the final questionnaire by a psychologist, who interviewed 10 people and eliminated one subgroup for each pair of Census subpopulations. In Table 1 are shown all the included 25 subpopulations on which the otorhinolaryngologists were interviewed. The names subpopulations were used because it is simpler recall the number of people by name than by other characteristics. Therefore, the names subpopulations are less affected by transmission and barrier effects.

6. Results

In June 2004 seventy-two questionnaires were collected. The missing nine items (out of a total number of 1872 items) have been replaced with the median value of the same questions.

In Table 2 are shown the characteristics of the interviewed ORLs: the median age was 47 years (40, 51 I and III quartile) and 86.1% were male. The median number of years at work was 18, the median number of years worked in the current Institution was 9 and in the current Department 8. To avoid the possible multiple counting of each injury occurred in the same Department, responses for all cases in the same department were weighted for the number of ORLs working in the same department.

7. Estimation of the Unreported Injuries

The total number of injuries recalled by the otorhinolaryngologists was 174. Table 3 shows the number of ORLs working in each Institution and the mean number of children having a hospitalization for a foreign body injury recalled by physicians in each Department. These data provided the information for weighting the cases by the number of ORLs working in each Institution. The adopted weighting scheme consisted in dividing the number of cases observed by each physician by the number of otorhinolaryngologists in each Department. In three centers no cases of hospitalization for foreign body injuries in children were observed, whereas the maximum number of children hospitalized was observed in the hospital of Tortona (17.5 weighted cases). Using ISTAT official data (Sistema statistico nazionale - Istituto nazionale di statistica 1992-1997), the population of Piemonte was reported as 4303000 people and the number of children under 15 years of age as 546000 children.

8. The Choice of Subpopulation of Known Size

The choice of the suitable subpopulations to use in the estimates of the social network sizes is the most important and sensitive operation to carry out in this methodology. A not careful choice of the subgroups could lead to problems of transmission and barrier effects.

In the scale-up method the mean number of people known in a subpopulation should be linearly proportional to the size of the same subpopulation (Killworth, McCarty et al. 2002).

Sometimes this linearity is not hold and the reason could be due to the problems of barrier and transmission effect.

Moreover, several works were published trying to explain (i) how respondents reply to these kinds of questions and (ii) how the mean number of people belonging to a subpopulation recalled by respondents and the subgroup size are related (Johnsen, Bernard et al. 1995; McCarty, Killworth et al. 2000; Johnsen, Killworth et al. 2002; Killworth, McCarty et al. 2002; Killworth, Johnsen et al. 2003).

Therefore, we decided to use in the final estimates only the subpopulations for which this linearity was hold. We assessed the linear relation between the subpopulation sizes and the mean number of people recalled by respondents for each subgroup with a regression model. The dependent variable was the mean number known in each subpopulation and the independent the relative size of subpopulations.

The regression model with all the 25 subgroups accounted only for the 32% of the variability (adjusted $R^2 = 0.32$), showing that some subpopulations were not in linear relation. Therefore, using a graphical analysis of residuals we eliminated five subpopulations and the resulting adjusted R^2 was 0.79.

The five eliminated subpopulations were: “People living in a rented houses”, “Self employed”, “Families with 7 or more components”, “Farmers” and “Unlettered”.

With the chosen 20 subgroups we estimated the social network size of each ORLs involved in the study ($\hat{c}_i$); the mean network size was 947.3 (95% C.I. 810.7-1083.9) persons.

These estimates were then used for estimating our target subpopulation. So the estimated number of children under 15 years that were hospitalized for a choking injury in years 2000-2002 was 281 (95% C.I. 188.5-374.6).

Official data indicates that the number of foreign body injuries in children, as recorded in the Hospital official discharge database, was for the years 1999-2000-2001 equal to 218 injuries (Ministry of Health data (Ministero della Salute 2004)).

Moreover, a previous estimate of the mean social network size of an American study was of 286 persons (Killworth, Johnsen et al. 1998).

9. Discussion

The estimate of the number of injuries is matching the estimate based on the discharge records, the latter being within the 95% confidence bounds. The estimate of the social network size for the otorhinolaryngologists is bigger than the one of other studies (286 persons) and probably this is due to the big amount of people that a physician usually knows and meets.

Eventually, our results show that the scale-up method estimates are an efficient and precise way to estimate the size of unknown subpopulations in the context of injury prevention with very selected populations (like the one represented by the otorhinolaryngologists), with the clear condition that the appropriate subpopulations are chosen in advance to obtain in addition stability of the estimates.

For choosing the subpopulations of known size we used an algorithm based on the regression model that is simply to implement and seems useful for our purposes. Clearly, these kinds of selection algorithms have to be specified before the analysis is started and at the moment it represents the weakest part of the overall estimation procedure, due to its data-driven nature.

Several limitations are however characterizing both the method and the study. It is reasonable to think that the number of people belonging to the eliminated subpopulations was difficult to count for a physician due his/her particular social position, which makes very difficult to meet some kind of people (e.g. unlettered or farmers) or help the acquaintance of other kinds of people (e.g. self employed).

Unfortunately the problems due to the barrier and transmission effects are only reduced with the selection algorithm adopted in the present work but not solved. Therefore the best solution is choosing the subpopulations we think that are suitable for the selected sample, i.e. subgroups that we suppose are known by our sample with a higher or lower probability than the general population (i.e. a physician can easily know a self-employed and difficultly an unlettered).

Barrier and transmission effect could affect also the estimates of the number of injuries. Obviously some kinds of accidents are more common in some regions and countries than in others but if we focus the studies on single and precise areas these types of injuries could be nevertheless estimated with a sufficient degree of accuracy. Moreover the barrier and the transmission

effect are also due to the social characteristics of people interviewed, but probably information about all kinds of injuries (omitting the violence) is well transmitted to the acquaintance. The major problem is recalling the injuries of minor severity: some studies show that even parents tend to forget less serious accidents occurred to their children (Scheidt, Brenner et al. 2000; Cummings, Rivara et al. 2005). Moreover, physicians recall in a more accurate way the injuries they treat than those they are just aware of, within their working social network.

The intent of this work was understood the capability of the scale-up method to estimate in an efficient and correct way the number of foreign body injuries in the Italian population. The scale-up method estimates precisely the size of the unknown subpopulations, but for avoiding the problems of transmission and barrier effects the choice of the subpopulation of known size have to be made with accuracy and putting particular attention on the severity level of the injury.

References

Bernard, H. R., E. Johnsen, et al. (1989). Estimating the size of an average personal network and of an event population. The small world. M. Kochen. Norwood, NJ, Ablex Publishing.

Bernard, H. R., E. C. Johnsen, et al. (1990). "Comparing four different methods for measuring personal social networks." Social Networks **12**: 179-215.

Bernard, H. R., E. C. Johnsen, et al. (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., P. D. Killworth, et al. (2001). "Estimating the Ripple Effect of a Disaster." Connections **24**(2): 18-22.

Bernard, H. R., P. D. Killworth, et al. (1990). "Estimating the size of personal networks." Social Networks **23**: 289-312.

Cummings, P., F. P. Rivara, et al. (2005). "Ability of parents to recall the injuries of their young children." Inj Prev **11**(1): 43-47.

Freeman, L. C. and C. R. Thompson (1989). Estimating acquaintanceship volume. The Small World. M. Kochen**:** 147-158.

Guard, A. and S. S. Gallagher (2005). "Heat related deaths to young children in parked cars: an analysis of 171 fatalities in the United States, 1995-2002." Inj Prev **11**(1): 33-37.

Johnsen, E. C., H. R. Bernard, et al. (1995). "A social network approach to corroborating the number of AIDS/HIV victims in the US." Social Networks **17**: 167-187.

Johnsen, E. C., H. R. Bernard, et al. (1995). "Who knows your HIV status? Waht HIV+ patients and their network members know about each other." Social Networks **17**: 189-217.

Johnsen, E. C., P. D. Killworth, et al. (2002). "Recalling numbers of alters in target subpopulation: an explanatory model for a power law."

Killworth, P. D., E. Johnsen, et al. (2003). "Attempting to quantify transmission and barrier errors in scale-up methods."

Killworth, P. D., E. Johnsen, et al. (1998). "A social network approach to estimating seroprevalence in the United States." Social Networks **20**: 23-50.

Killworth, P. D., C. McCarty, et al. (2002). "Two interpretations of reports of knowledge of subpopulation sizes." Social Networks **25**: 141-60.

Killworth, P. D., C. McCarty, et al. (1998). "Estimation of seroprevalence, rape and homelessness in the U.S. using a social network approach." Evaluation Review **22**: 289-308.

McCarty, C., P. D. Killworth, et al. (2000). "Comparing two methods for estimating network sizes." Human Organization **60**: 28-39.

Ministero della Salute (2004). Dati di ricovero - classificazione ICD-9 CM, Ministero della Salute. **2005**.

Morrison, A. and D. H. Stone (2000). "Capture-recapture: a useful methodological tool for counting traffic related injuries?" Inj Prev **6**(4): 299-304.

Scheidt, P. C., R. A. Brenner, et al. (2000). "Parental attitudes regarding interviews about injuries to their children." Inj Prev **6**(1): 51-55.

Sistema statistico nazionale - Istituto nazionale di statistica (1992-1997). 13. censimento generale della popolazione e delle abitazioni : 20 ottobre 1991, ISTAT.

Snidero, S., R. Corradetti, et al. (2004). "The network scale-up method: a simulation study in case of overlapping sub-popualtions." Metodoloski zvezki **1**(2): 395-405.

Zigon, G., D. Gregori, et al. (2004). "Children's mortality for suffocation in Europe (1980-1995): a review of the official data." submitted.

	Subpopulations	*Absolute size*
1	Maria	29326
2	Anna	13489
3	Rosa	8184
4	Giuseppina	8108
5	Angela	8033
6	Giovanna	7654
7	Giuseppe	59865
8	Antonio	40996
9	Giovanni	37738
10	Francesco	32433
11	Mario	27508
12	Luigi	27356
13	Graduates in physical education	3715
14	Women with one child	180870
15	Houses with more than 6 inhabitants	137671
16	People living in a rented houses	1305153
17	Self employed	325040
18	Families with 7 or more components	40019
19	Women with 3 or more children	32897
20	Graduated at school of arts	3544
21	Sport trainers	558
22	Farmers	73177
23	Architects	3415
24	People living in assistance institute	24240
25	Unlettered	37253

Table 1 The subpopulations of known size

	Median	*1st percentile*	*3rd percentile*	*N.*	*%*
Age	47.0	40.0	51.0		
Total nr. years of work	18.0	10.5	24.0		
Nr. years of work in the institution	9.0	6.0	15.8		
Nr. years of work in department	8.0	5.0	15.0		
Gender					
Male				62	86.1%
Female				9	12.5%
Missing				1	1.4%

Table 2 Characteristics of the sample of otorhinolaryngologists

Institution	Nr. of otorhinolaryngologists	***Mean nr. of cases***
Gradenigo	3	0.0
Molinette 2	4	3.5
Molinette 3	10	1.9
Alba	5	2.4
Asl 19	6	0.7
Ivrea	1	0.0
Ciriè	5	1.8
Maria Vittoria	6	3.0
Mauriziano	9	2.0
S. Croce Cuneo	5	0.4
S. Giovanni Bosco	6	0.0
SS. Annunziata	4	1.0
Tortona	4	17.5
Vercelli	4	1.0

Table 3 Mean number of cases for each department.