International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878
Volume 3 Issue 3, March 2015

Review and Simulation of Different Sampling Methods

Christian Hoops[1], Rahul Pathare[2]

[1]Ipsos Public Affairs, Sachsenstraße 6, 20097 Hamburg (Germany)

[2]Narsee Monjee Institute of Management Studies, University of Mumbai
V.L. Mehta Road, Vile Parle, Mumbai – 400 056, Maharashtra, India

Abstract: ***Obtaining information about hard-to-reach populations is a major challenge in the market research field. In our case of analyzing German minorities in Denmark, only a small fraction of the total population belongs to the target population. Therefore, selecting minorities by generating telephone numbers at random would result in very high costs. Alternative sampling methods have to be used, but there are no practices to identify the best approach. This article tries to fill this gap and creates a comparison of snowball sampling (SS), random digit dialing (RDD), gravity sampling (GS) and facility-based sampling (FBS). Sample data has been extracted by a previous survey (Hoops, Schnapp and Schaefer-Rolffs 2013) and a further model extended by randomly generating and simulating all four sampling methods using bootstrapping procedures. This enabled us to estimate the cardinality of the sample space, the bias and the variance of the inclusion probabilities in the sample for each method. Only GS and RDD create samples which are asymptotically unbiased. The combination of gravity and complete as well as non-overlapping citizens registers produces the highest cardinality of the sample space. But in contrast to RDD, citizens registry-office methods allow no household samples. So gravity analyses help to identify regions with a high prevalence of the target population to create samples with roughly varying inclusion probabilities. Our simulations indicate that gravity sampling methods using official databases produce very high quality samples. For cost reasons this method should be tested in practice to conduct surveys with hard-to-reach populations.***

Keywords: Bootstrapping, gravity sampling, snowball sampling, surveys

1. Introduction

Surveys of hard-to-reach populations have always been a great challenge to scientists. Traditional methods of sampling these populations are ineffective and have many bias problems (Agadjanian and Zotova 2012). Researchers who want to create samples by using established methods like RDD, Marpsat and Razafindratsima (2010) state that the small percentage of the target population will increase the cost of investigation.

Researchers require many respondents for representative statements, which is quite difficult to collect from hard-to-reach populations and, on top of that, the risk of missing important data is slightly higher than in standard research due to marginal group reasons. In addition, respondents might not be able to recommend accurate target samples for further study due to a lack of knowledge or non-willingness to suggest eligible people (Hendricks and Blanken 1992).

However, the term "hard-to-reach populations" has to be delimited from similar terms like "hidden populations" who do not want to reveal their identity, such as HIV-infected persons or prostitutes. It has been seen in many research papers that these hidden populations are not willing to reveal their identity because they might be stigmatized by society and do not wish to recall horrible past events (Faugier and Sargeant 1997; Hendricks and Blanken 1992; Marpsat and Razafindratsima 2010). In contrast, hard-to-reach populations (e.g. immigrants or national minorities) are only defined by a low prevalence. There are several methods of identifying hard-to-reach populations, such as time-location sampling, snowball sampling and facility-based sampling. However, these methods can only generate samples of unknown composition.

This article mainly focuses on national minorities – those who feel that they belong to Germany, but are living in Denmark. With the Bonn-Copenhagen Declarations of 1955 (see Jäckel 1959: 74ff.), it is specified to protect this small group and not to conduct the minority status in official surveys. However, political parties of minority people have strong influence during elections and subsequently in their parliament. Perceptions and views of minority people play an important role during elections, thus it is important to find these minority populations and collect information by interviewing them via appropriate sampling methods (Hoops, Schnapp and Schaefer-Rolffs 2013). However, such groups make up only a small percentage of the country's total population. This means that the prevalence which depends on the sampling method is very low.

There are many approaches available in this research field but, unfortunately, there is no systematic or an established methodology that can accurately suggest which technique to select and apply in the study of low prevalence. Moreover, there are limited researches available which compare survey methods in practice to identify differences between them. This paper tries to fill this research gap.

First, we will give a review of different sampling methods to conduct with hard-to-reach populations. After that we will simulate the method-specific bias, the cardinality of the sample space and the variance of the inclusion probabilities in the sample. The sample space consists of all people with target characteristics who could be included in the sample. From this it follows that the quality of the sample space can be measured as the closeness of the sample space to the target population. The cardinality of the space sample indicates this degree of consistency.

2. Literature Review

2.1 Random Digit Dialing

Telephone method sampling is done by randomly selecting numbers from the telephone directory and making calls to a person (Brunner and Brunner 1971) for an interview. The method was invented in the USA where out of 10 digits the first 3 digits are the area code, then 3 digits are the telephone switching center and 4 digits are the suffix. In RDD, after entering the first six digits in the system, the last four digits are selected randomly.

On the other hand, in Germany, the first 3 to 5 digits (e.g. 040 or 04542) are the regional area code, which is followed by some digits, out of which the last two digits are used to generate a random number. However, unfortunately, regional sampling is not possible for mobile phones.

Table 1: Example of random digit dialing

Area code	*Number block*	*Status*
04542	55832 00	Registered
04542	55832 **01**	Generated
04542	55832 02	Registered
04542	…	…
04542	55832 99	Registered

In random digit dialing, strictly speaking, all non-registered numbers between the lowest and the highest will be determined (see Fuchs 1994). However, this approach is mostly not practical, especially when many generated numbers do not exist. Therefore the randomized last digit method is being performed, where the last digits are replaced by randomly generated digits. Since the shares from registered and generated numbers are not identical in each number block, this method produces strongly varying inclusion probabilities in the sample. Blocks with a large amount of registered numbers are more often included in the sample than blocks with few registered numbers. With appropriate corrections, these selection differences can be compensated.

According to Glasser & Metzger (1972), the telephone survey method (see Table 1) is less expensive and has a better rate of response, with the callback option or monitoring previous calls, as compared to personal contacts or mail survey. However, random digit dialing is limited to people who possess a landline or mobile phone and creates a non-response bias (Glasser and Metzger 1972). The article by Esslemont et al. (1992) compares demographic character-istics between listed and unlisted respondents and found out that, due to telephone directory sampling, differences exist between them.

Previously, usage of cellular sampling was excluded from the RDD sampling because of the lower number of mobile phones users, and for ethical and legal issues. However, in the last decade, mobile phone usage has increased tremendously, and as a result survey companies include mobile sampling in RDD (Frankel et al. 2007). They have further described the significance of weighting telephone surveys that include cellular phones as well as landline telephones.

It is important to apply an accurate sampling and response rate calculation method in survey analysis (Ezzati-Rice et al. 2000). The argument is that the eligibility of respondents is related with non-response of selected interviewees and, to support this statement, Ezzati-Rice et al. (2000) have discussed the advantage of combining the response and coverage rate to calculate the response rate. In RDD, the non-response bias can be calculated by the response rate of a survey. However, while calculating the response rate, the researcher should exclude all non-coverage respondents who are eligible but do not have a telephone or mobile at home (Glasser & Metzger 1972; Frankel et al. 2007).

Massey & O'Connor (1997) have recommended calcu-lating the response rate at every stage of the sample design and also considering new types of incentives for the interviewer and interviewee. They further encouraged the researchers to develop new methodology for a better response rate and to perform research on the effects of questionnaires on respondents.

2.2 Snowball Sampling

It is important to study subjects with sensitive issues, like HIV study (Faugier and Sargeant, 1997; Kendall, et al. 2008), drug addicts (Lopes, Rodrigues and Sichieri 1996), migrants (McKenzie and Mistiaen 2009) etc. To obtain necessary prerequisite data for research and study of hard-to-reach or hidden population, snowball sampling is one of the most efficient method available (Lopes et al. 1996, Berg 1983). The snowball sampling approach suggests that a new respondent can be created by a referral series within a group of people, who know each other, which then forms multiple recommendation waves and develops interesting comparisons in the group.

Berg (1988) posits that if a sampling frame is unavailable to cover the population, respondents' knowledge and connections are necessary to start the referral chain process. The chain referral technique is a self-propelled phenomenon (Biernacki & Waldorf 1981), which, once it gets started, automatically works by itself. During the process the scientist should first search for the respondents as an initial referral chain then verify their eligibility according to the proposed research subject. Furthermore, he must encourage all initial respondents to become research assistants in the project and ask them to recommend the next person for the study. In addition, he should manage, control and limit the chain process in order to avoid excessive, unnecessary respondents, and then motivate them from time to time to maintain the quality of the data (Biernacki & Waldorf 1981; Marpsat & Razafindratsima 2010).

A comparative survey investigated by McKenzie & Mistiaen (2009) to analyze data of Japanese-Brazilian families as well as a HIV study by Faugier & Sargeant (1997) have found out that snowball sampling has a wide scope to conduct qualitative sociological research and has an edge over other

sampling methods when it comes to reaching hidden populations in society. It is a very efficient, economical method which produces results in less time (Atkinson and Flint 2001).

Even though the snowball technique is an effective sampling method for the study of rare population, it has some limitations. Due to social problems like stigmatization from society, hidden population subjects were reluctant to reveal their identity (Faugier and Sargeant 1997; Hendricks and Blanken 1992). Berg (1988) claims, that if respondents are not guaranteed the confidentiality of information about them, ethical issues can arise. Other than social and ethical problems, Biernacki and Waldorf (1981) argue that the snowball technique is an expensive methodology as the interviewer is responsible from the start to end of the research. During the research period the analyzer should be concerned about information confidentiality, for example some people from the same referral chain could meet on the street, in court or in prison, and any disclosure of information may cause a serious problem in the entire research study (Berg 1988). The survey method in an urban populated area could be affected because of sampling error by clustering, interviewers' bias, bad weather and imbalanced interviews of homebound, homeless and unemployed people (Miller et al. 1997).

Illenberger et al. (2008) found out that the overestimation bias problem occurred in the early stage of snowball sampling and the results are limited to mean degree and the clustering coefficient. In their research study they try to find out: How initial respondents affect the subsequent chain process as well as consequences of respondents' refusal to participate in the survey. Furthermore, they encourage other researchers to investigate bias-correction measures of "closeness and betweenness".

The recent paper by Johnston and Sabin (2010) on hard-to-reach population shows that the respondent-driven sampling method by Heckathorn (1997) can be a very good alternative to reduce initial bias problems associated with snowball sampling. They argue that respondents of snowball sampling give unlimited references which create bias issues such as clustering, differential recruitment and accessibility to further samples. This can be eliminated by issuing coded coupons that can be redeemed only at a fixed location (interviewer's office) in a limited time. Respondents are allowed to refer only a limited number (2-3) of target people, which further reduces the overrepresentation bias problem. Moreover, Johnston and Sabin (2010) suggest that respondents should be rewarded with gifts or shopping coupons, rather than cash, to motivate them to participate in the study. The interviewer should estimate the price of the gift in such a way that it should not be overvalued or undervalued. Because, if the gift price is overvalued, then the respondent may sell it to another person and if it is undervalued, then the respondent will not show up for the interview. At least the respondent-driven technique allows to make asymptotically unbiased estimates (Salganik and Heckathorn 2004). Nevertheless, the sample space has a low cardinality due to target people with only a small or without social network.

2.3 Facility-Based Sampling

The success of hard-to-reach sampling depends on available information and knowledge about the target population. Snowball sampling, response-driven sampling, and time location sampling are a few examples of how researchers are able to approach rare population (Sangngam and Suwattee 2010). Researchers can also use the facility-based survey method to find hard-to-reach people, in which the interviewer selects a specific location (e.g. hospital) where he gets information about the target population (e.g. drug addicts) and he can also use different facilities of the location and then the researcher can conduct survey interviews with doctors, nurses or patients. For instance, Marchant et al. (2008) conducted a study on intermittent preventive treatment of malaria in pregnancy (IPTp) in Tanzania, where a researcher interviewed female respondents who had given birth to a child, as well as hospital employees, and then analyzed the coverage of IPTp by using the facilities at an antenatal clinic. During this survey, each piece of equipment and supplier were checked at the facility location to avoid delay during research.

According to Magnani et al. (2005) facilities such as drug treatment centers, HIV specialist hospitals, drug rehabilitation centers and gay bars are some good locations to find rare population involved in drugs and HIV activities. For example, for conducting a facility-based survey of HIV-infected people, first select a hospital with HIV specialist doctors where there is a higher possibility of getting a database of HIV infected people. This means that in the previous case hospitals (facilities) are being used to find and recruit target people. Subsequently, the researcher can conduct interviews at the facility location or can meet target samples personally at their home.

In spite of all of its advantages, the facility-based survey method has some bias problems such as volatile laws of the juridical system, stigmatization by society and police investigations (Turner et al. 2001). In addition, proper pre-survey management and skilled interviewers are needed for the study. Chopra et al. (2005) conducted a survey on intervention in a quality of care in Cape Town in which nurses were trained by local health services with the help of WHO's teaching modules. On top of that, prior juridical permission was always required from the top authority of the hospital or government to perform the entire interview procedure at the facility location, which increases the time and cost of the survey. The number of special clinics for HIV-infected patients, especially in poor and developing countries, is limited, which places restrictions on the research (Shaghaghi, Bhopal and Sheikh 2011).

Furthermore, Turner et al. (2001) pointed out that the major issue for sampling and sample size is the rare occurrence of subjects at the facility place. For instance, target people who are unwilling (because of stigmatization) or unable (because of sickness or family problems) to come to the interview place creates a bias problem in the sampling. To overcome this rare occurrence problem, he suggested a strategy to conduct the entire survey at the facility location for five consecutive days, so that the maximum number of people

would show up and participate in the research study.

2.4 Citizens Registry-Office Sampling

In empirical research, citizen's registry-office sampling is a typical method for surveying immigrants or minority populations. Relevant information fetched from the citizens registry-offices, which includes personal and contact information of the target population, can be used for sampling purposes. The main advantage of this sampling is that it is possible to collect the information of each target person who has registered themselves at the government office. Once the contact information list is available, interviews of the target population would be possible using RDD or direct face-to-face interview. However, with RDD there is a lower probability of the target sample being selected for an interview. Salentin (2007) describes three limitations that may occur in this sampling method:

a) In spite of mandatory regulations, some people do not register at the government office.
b) Sometimes people from minority groups do not update their address after changing their accommodation.
c) If the target population is living in rural areas, it is hard to achieve the expected sample size as the population is limited.

For accurate measurement of minorities in the specific location the researcher can extend our gravity model analysis using city registration methodology. The researcher can get contact information of the target population from schools with minority children. Even though this method is costly and time-consuming, there are better chances of finding people from minority groups near border locations.

2.5 Gravity Sampling

The basic idea behind the gravity sampling method is that the empirically based expectations about the prevalence of the target population at each location in their disseminated territory can be calculated on the basis of the different attractiveness factors and distance to the specific location (Hoops and Schnapp 2013). This method can be used, in particular, when the target population is living in a relatively well-defined area of distribution. Initially the target area of a survey is so limited that such a high prevalence of the target population is achieved in the defined area and then telephone or face-to-face interviews would be beneficial to interview a person from the minority population. The technique is based on the multiplicative competitive interaction model (Nakanishi and Cooper 1974), which was designed as an extension of the univariate model by Huff (1964).
Gravity sampling requires confirmation of two elements. First of all we have to show that the probability of the target population increases at the residences which are located closer to the defined geographical area or respective country's border. Secondly, residences with a high attractive-ness for members of the target population should be selected more often than residences with less attractiveness.

If both hypotheses are correct, it can be considered that the distance to the border and the attractiveness of a place influence the choice of residence of target people. It makes it possible to accurately estimate the relative prevalence of the various destinations. Eventually, it can specify the number of interviews as well as the number of target population in the sample space for each location. In contrast to established methods like RDD, gravity analyses enable substantial cost savings when conducting a low prevalence survey (Hoops, Schnapp and Schaefer-Rolffs 2013).

The gravity probability depends on an empirical parameter called λ. The greater the impact of the distance on the prevalence, the greater is the value. In this paper, gravity analysis is tested with German minorities in Denmark, where both countries have a strong economy and share the same ethnicity. Nevertheless, λ depends on the sampling area and varies from country to country. The parameter that we used for the Danish-German border should be validated for other border region.

3. Methodological Analyses

3.1 Simulation Procedure

Based on an empirical study in Denmark, where information on the frequency of the German minority population was collected and 375 people who were living in Denmark and less than 100 kilometers from the Danish-German border were interviewed, we know that approximately 7.1% of all people feel that they belong to the German minority and around 8.2% of all households have at least one national minority. In the interests of simplification, we concentrated on the 14 most popular locations, which have a total population of 111,300 inhabitants.

Based on these data, we have randomly generated samples with this population size and assigned each case a household number, a household size, an age and the membership to the German minority (or not). Excluding this minority status, we have focused on the official statistics from Denmark (Danmarks Statistik 2013).

With reference to the work of Lu et al. (2009), we estimated the relevant network size of each national minority (see Figure 1).

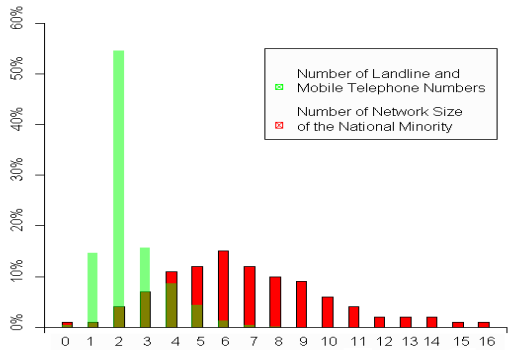

Figure 1: Assumed distributions of numbers and size

We assumed that our relevant size is similarly normally distributed and is slightly right-skewed with the maximum value of 16. About 99 percent of all national minorities should know at least one member who is not a part of their

household. In that context more than half of the total minority is at least connected with six other German minorities.

Based on our Ipsos Democracy Pulse (see Hoops, Glantz and Michael 2013), we have used a RDD-similar technique for creating random mobile and landline numbers in Germany. Since there was no reliable data available before we expected that there would be no differences between these two countries. The assumed distribution of the sum of landline and mobile phone numbers of each person can be seen in Figure 2.

We have used $\lambda = 0.399$ for gravity sampling (Hoops, Schnapp & Schaefer-Rolffs 2013). Moreover, along with distance to the Danish-German border and also considered the election results of the minority party, the density of minority oriented organizations, schools and companies as relevant attractiveness variables. We supposed that our survey will constitute a public interest so we could use citizens registers to select a sample. This registers would be complete, correct and contain no duplicates. For facility-based sampling we restricted ourselves to German schools in Denmark so only people with children that go to school can be included.

We used bootstrapping to approximate the theoretical values of the bias, the cardinality of the sample and the variance of the inclusion probabilities. These statistics are computed by selecting sub-samples of the size of 100 people from the complete sample by different sampling techniques and repeated it 10,000 times. Therefore, the inclusion probability of a person in the sample can be estimated by the number of selections divided through the number of repetitions. While the bias is calculated by the relative medium deviation between sample and target population in some variables (age and number children), the cardinality of the sample space is measured as the number of different cases that are selected at least once divided through the number of target populations.

3.2 Results

In Table 2 you can see that the cardinality of the sample space is very low for facility-based sampling whereas gravity sampling has a complete sample space under our assumption of no missing minority in the register.

Table 2: Measures to compare different methods

Method	*Bias*	*Cardinality of the sample space*	*Variance of the inclusion probability*
FBS	63.1%	38.6%	25553.6
GS	9.5%	100%	127.5
RDD	9.4%	96%	10719.8
SS	17.4%	98.4%	3540.9

In RDD the cardinality of sample space is lower than in the gravity simulation as only people who have a landline or mobile connection can be interviewed. Because 1% of the total target population can't be recommended by any minority, the cardinality of the sample space in the snowball sampling simulation should be 99%. However, the cardinality is about 98.4%, so there is nearly one percent that could be selected, but is not. The facility-based method has the lowest cardinality among all sampling techniques, since only 3,153 out of 7,913 target people have school-going children.

On the contrary, facility-based sampling has the highest variance of the inclusion probabilities as compared to the others (see Figure 2), because there are many people who are eligible but have no chance to be part of the sample. The top-left plot shows only a few data points. Most of the people are selected between 310 and 330 times.

As we can see, gravity sampling produces the lowest variance of the inclusion probability and the selection frequencies are concentrated from 100 to 160 times. In contrast to RDD it is no household sample and enables roughly constant inclusion probabilities in the sample. The simulation of RDD shows a slightly larger variance with selection frequencies, which are lying in the range of 130 to 200 times. The more landline and mobile telephone numbers a minority member has, the greater the probability of that person being selected in the sample space. So one minority has been randomly selected 862 times due to the possession of many telephone numbers in total.

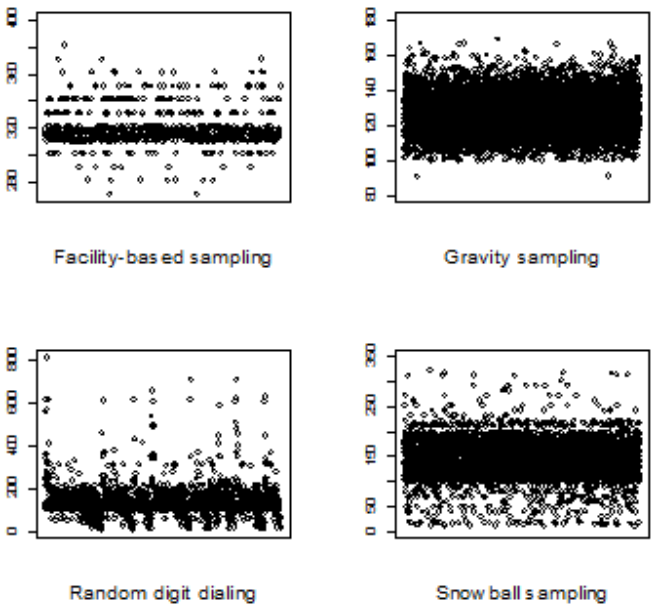

Figure 2: Plot of the inclusion frequencies of each technique

In snowball sampling the selection frequencies are concentrated between 100 to 200 times and a lower total variance of the inclusion probabilities is created. This is surely something surprising and calls our assumptions into question, because we expected a greater variance. The gravity sampling shows values from 91 to 169. Thus there is only less variation due to the expected frequency of about the average number of 126.37 (=(10,000*100)/7913) selections.

However, it is hardly surprising that all sampling methods create biased samples because of the very low sample size of 100. RDD and gravity samples show the lowest differences to the total population. So both methods are recommendable for representative studies. Furthermore, both biases should be asymptotically tending towards zero. Therefore, with

snowball and facility-based sampling there can only be samples produced which differ greatly from the mean of the population. Because we even compute the distortion across the central portion of children in the household, the samples in facility-based sampling are greatly biased since the researcher can only reach target people with children. Snowball sampling is biased because older people tend to have a greater network size and are more often selected.

4. Discussion and Conclusion

Because of low prevalence, traditional survey methods such as random digit dialing and citizens registry-office sampling are very expensive when carrying out surveys of hard-to-reach populations. Therefore, alternative methods such as gravity or snowball sampling with some limitations would be favorable and recommendable.

Our simulations have shown some advantages of gravity sampling. First, this method in combination with citizens registers has the largest cardinality of the sample space under the assumption of correct official data. Second, there is no significant distortion detected and inclusion probability variation is far less. Although we received positive results for gravity sampling, the computations in real studies might behave significantly differently. Therefore, the gravity sampling method should be strongly tested in practice.

It is, however, possible that we are overestimating the variance of the inclusion probabilities in the RDD simulation, because the inclusion probability does not increase proportionally to the amount of telephone numbers. Surely a person who has ten mobile numbers will not always possess ten mobile phones and will not always be equally reachable at each number. This should be taken into account for later simulations.

It is also possible to incorrectly estimate the network size of national minorities. If, for example, a single individual knows more than 16 other members and/or the proportion of the minority without a network is considerably larger, this could lead to an increase in the inclusion probability in random samples. Nonetheless, our conservative assumption should not influence the conclusions.

We must of course admit that the calculation of bias using age and number of children has been established somewhat subjectively. Especially since both variables are obviously correlated with each other. But there was no valid multivariate marginal distribution that we could have used for our simulations. It is also impossible to judge the value of the estimation of loading parameters lambda. So it only remains to be recommended that an appropriately practical method test should be conducted with a large sampling to calculation the deviations from official statistics.

References

[1] Agadjanian, V. and Zotova, N. (2012) ‘Sampling and Surveying Hard-to-Reach Populations for Demographic Research: A Study of Female Labor Migrants in Moscow’, *Demographic Research*, 26(5) 131-150.
[2] Atkinson, R. and Flint, J. (2001) ‘Accessing Hidden and Hard-to-Reach Populations: Snowball Research Strategies’, *Social Research*, 33(1) 1-4.
[3] Berg, S. (1983) ‘Random Contact Processes, Snowball Sampling and Factorial Series Distributions’, *Journal of Applied Probability*, 20(1) 31-46.
[4] Berg, S. (1988) ‘Snowball Sampling’, *Encyclopedia of Statistical Sciences*, 8, 529–532.
[5] Biernacki, P. and Waldorf, D. (1981) ‘Snowball Sampling: Problems and Techniques of Chain Referral Sampling’, *Sociological Methods and Research*, 10(2) 141-163.
[6] Brunner, J. A. and Brunner, A. G. (1971) ‘Are voluntarily Unlisted Telephone Subscribers Really Different?’, *Journal of Marketing Research*, 8(1) 121-124.
[7] Chopra, M., Patel, S., Cloete, K., Sanders, D. and Peterson, S. (2005) ‘Effect of an IMCI Intervention on Quality of Care across four Districts in Cape Town, South Africa’, *Archives of Disease in Childhood*, 90(4) 397-401.
[8] Danmarks Statistik (2013) ‘StatBank Denmark’ [electronic resource]. Copenhagen: Statistics Denmark. http://www.statbank.dk.
[9] Esslemont, D., Petersen, S. and Selvakumar, K. S. (1992) ‘Telephone Directories as Sampling Frames’, *Marketing Bulletin*, 3(5) 38-45.
[10] Ezzati-Rice, T. M., Frankel , M. R., Hoaglin, D. C., Loft, J. D., Coronado, V. G. and Wright, R. A. (2000) ‘An Alternative Measure of Response Rate in Random-Digit-Dialing Surveys that Screen for Eligible Subpopulations’, *Journal of Economics and Social Measurement*, 26, 99-109.
[11] Faugier, J. and Sargeant, M. (1997) ‘Sampling Hard-to-Reach Population’, *Journal of Advanced Nursing*, 26(4) 790-797.
[12] Frankel, M. R., Battaglia, M. P., Link , M. and Mokdad, A. H. (2007) ‘Integrating Cell Phone Numbers into Random Digit-dialed (RDD) Landline Surveys’, *Proceedings of the Survey Research Methods*, 3793-3800.
[13] Fuchs, M. (1994) *Umfrageforschung mit Telefon und Computer*. Weinheim: Beltz Psychologie Verlags Union.
[14] Glasser, G. J. and Metzger, G. D. (1972) ‘Random Digit Dialing as a Method of Telephone Sampling’, *Journal of Marketing Research*, 9, 59-64.
[15] Groves, R. M. and Lyberg, L. E. (1988) ‘An Overview of Non Response Issues in Telephone Surveys’, in: Groves, R. M., Biemer, P. P., Lyberg, L. E., Massey, J. T., Nicholis, W. L. and Waksberg, J. (eds.). *Telephone Survey Methodology*. New York: Wiley.
[16] Heckathorn, D. (1997) ‘Respondent-driven sampling: A New Approach to the Study of Hidden Populations’, *Social Problems*, 44, 174-199.
[17] Hendricks, V. and Blanken, P. (1992) ‘Snowball Sampling: Theoretical and Practical Considerations, in: Hendricks, V. and Blanken, P. (eds.). *Snowball Sampling: A Pilot Study on Cocaine Use*. Rotterdam: IVO.
[18] Hoops, C., Glantz, A. and Michael, T. (2013)

'Innovations in Rolling-Cross-Section-Design', *Journal of Political Science*, 23(2) 215-243.

[19] Hoops, C. and Schnapp, K.-U. (2013) '*Sampling of Hard-to-Reach-Populations*', Paper presented at European Survey Research Association Conference, Ljubljana, July 15-19.

[20] Hoops, C., Schnapp, K.-U. and Schaefer-Rolffs, A. (2013) *Zur Ziehung von Stichproben in schwer erreichbaren Zielgruppen mit Gravitations-analytischen Methoden*. Hamburg: University of Hamburg (WiSo-HH Working Paper Series; WP-2013-4).

[21] Huff, D. L. (1964) 'Defining and Estimating a Trading Area', *Journal of Marketing*, 28(3) 34-38.

[22] Illenberger, J., Flötteröd, G. and Nagel, K. (2008) *An Approach to Correct Biases Induced by Snowball Sampling*. Berlin: Technical University Berlin (VSP Working Paper 8-16).

[23] Jäckel, E. (1959) *Die Schleswig-Frage seit 1945: Dokumente zur Rechtsstellung der Minderheiten beiderseits der deutsch-dänischen Grenze*. Frankfurt am Main: Metzner.

[24] Johnston, L. and Sabin, K. (2010) 'Sampling Hard to Reach Populations with Respondent Driven Sampling', *Methodological Innovations Online*, 5(2) 38-48.

[25] Kendall, C., Kerr, L. R., Gondim, R. C., Werneck, G. L., Macena, R. H., Pontes, M. K., Johnston, L. G., Sabin, K. and McFarland, W. (2008) 'An Empirical Comparison of Respondent-Driven Sampling, Time-location Sampling, and Snowball Sampling for Beha-vioral Surveillance in Men Who Have Sex with Men, Fortaleza, Brazil', *AIDS and Behavior*, 12(1) 97-104.

[26] Lopes, C. S., Rodrigues, L. C. and Sichieri, R. (1996) 'The Lack of Selecton Bias in a Snowball Sampled Case-Control Sudy on Drug Abuse', *International Journal of Epidemiology*, 25(6) 1267-1270.

[27] Lu, Y.-E., Roberts, S., Lio, P., Dunbar, R. and Crowcroft, J. (2009) 'Size Matters: Variation in Personal Network Size, Personality and Effect on Information Transmission', *Computational Science and Engineering*, 4, 188-193.

[28] Magnani, R., Sabin, K., Saidel, T. and Heckathorn, D. (2005) 'Review of Sampling Hard-to-Reach and Hidden Populations for HIV Surveillance', *AIDS*, 19(2) 67-72.

[29] Marchant, T., Nathan, R., Jones, C., Mponda, H., Bruce, J., Sedekia, Y., Schellenberg, J., Mshinda, H. and Hanson, K. (2008) 'Individual, Facility and Policy Level Influences on National Coverage Estimates for Intermittent Preventive Treatment of Malaria in Pregnancy in Tanzania', *Malaria Journal*, 7, 260.

[30] Marpsat, M. and Razafindratsima, N. (2010) 'Survey Methods for Hard-to-Reach Populations: Introduction to the Special Issue', *Methodological Innovations Online*, 5(2) 3-16.

[31] Massey, J. T. and O'Connor, D. (1997) '*Response Rate in Random Digit Dialing (RDD) Telephone Survey*' Paper presented at American Statistical Association Conference, Anaheim, August 10-14.

[32] McKenzie, D. J. and Mistiaen, J. (2009) 'Surveying Migrant Households: A Comparison of Census-Based, Snowball, and Intercept Point Surveys', *Journal of the Royal Statistical Society* 172(2) 339-360.

[33] Miller, K. W., & Bilder, L. B., Stillman, F. A. and Becker, D. M. (1997) 'The Feasibility of a Street-Intercept Survey Method in an African-American Community', *American Journal of Public Health* 87(4) 655-658.

[34] Nakanishi, M. and Cooper, L. G. (1974) 'Parameter Estimation for a Multiplicative Competitive Interaction Model - Least Squares Approach', *Journal of Marketing Research* 11(3) 303-311.

[35] Salentin, K. (2007) 'Die Aussiedler-Stichproben-ziehung', *Methoden - Daten - Analysen* 1(1) 25-44.

[36] Salganik, M. and Heckathorn, D. (2004) 'Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling', *Sociological Methodo-logy*, 34, 193-239.

[37] Sangngam, P. and Suwattee, P. (2010) 'Stratified Inverse Sampling', *Journal of Thai Statistical Association*, 10(1) 69-86.

[38] Shaghaghi, A., Bhopal, R. S. and Sheikh, A. (2011) 'Approaches to Recruiting "Hard-To-Reach" Populations into Research: A Review of the Literature', *Health Promotion Perspectives*, 1(2) 86-94.

[39] Turner, A. G., Angeles, G., Tsui, A. O., Wilkinson, M. and Magnani, R. (2001) *Sampling Manual for Facility Surveys: For Population, Maternal Health, Child Health and STD Programs in Developing Countries*. Chapel Hill: University of North Carolina (Measure Evaluation Manuel Series; Working Paper 3).

Author Profiles

Christian Hoops is a German statistician who works for Ipsos Public Affairs. In 2011 he received the EHI science prize for innovative methodology development, after which he moved to the Technical University of Dortmund as a lecture of quantitative methods and scientific work. His main research priority focuses on sampling, survey design and electoral research.

Rahul Pathare holds different certificates of industrial electronics, engineering and technology. He has 3 years of work experience in academics as an assistant professor at the Narsee Monjee Institute of Management Studies at the University of Mumbai. His area of focus has been business, statistics, scenario planning & technology.