Approaches to Recruiting 'Hard-To-Reach' Populations into Re-search: A Review of the Literature

Health Promotion Perspectives, Vol. 1, No. 2, 2011; P: 86-94

ORIGINAL ARTICLE **Open Access**

Approaches to Recruiting 'Hard-To-Reach' Populations into Research: A Review of the Literature

Abdolreza Shaghaghi[1, 2], Raj S Bhopal[3], Aziz Sheikh[4]

[1]Department of Medical Education Research Centre, R & D Campus, Tabriz University of Medical Sciences, Tabriz, Iran
[2]Department of Health Education & Promotion, School of Health & Nutrition, Tabriz University of Medical Sciences, Tabriz, Iran
[3]Bruce and John Usher Professor of Public Health, Edinburgh Ethnicity and Health Research Group, Centre for Population Health Sciences, University of Edinburgh, UK
4 Department of Primary Care Research & Development, University of Edinburgh, UK

(Received: 14 Sep 2011/ Accepted: 06 Oct 2011)

ABSTRACT

Background: 'Hard-to-reach' is a term used to describe those sub-groups of the population that may be difficult to reach or involve in research or public health programmes. Application of a single term to call these sub-sections of populations implies a homogeneity within distinct groups, which does not necessarily exist. Different sampling techniques were introduced so far to recruit hard-to-reach populations. In this article, we have reviewed a range of approaches that have been used to widen participation in studies.

Methods: We performed a Pubmed and Google search for relevant English language articles using the keywords and phrases: (hard-to-reach AND population* OR sampl*), (hidden AND population* OR sample*) and ("hard to reach" AND population* OR sample*) and a consultation of the retrieved articles' bibliographies to extract empirical evidence from publications that discussed or examined the use of sampling techniques to recruit hidden or hard-to-reach populations in health studies.

Results: Reviewing the literature has identified a range of techniques to recruit hard-to-reach populations, including snowball sampling, respondent-driven sampling (RDS), indigenous field worker sampling (IFWS), facility-based sampling (FBS), targeted sampling (TS), time-location (space) sampling (TLS), conventional cluster sampling (CCS) and capture re-capture sampling (CR).

Conclusion: The degree of compliance with a study by a certain 'hard-to-reach' group depends on the characteristics of that group, recruitment technique used and the subject of interest. Irrespective of potential advantages or limitations of the recruitment techniques reviewed, their successful use depends mainly upon our knowledge about specific characteristics of the target populations. Thus in line with attempts to expand the current boundaries of our knowledge about recruitment techniques in health studies and their applications in varying situations, we should also focus on possibly all contributing factors which may have an impact on participation rate within a defined population group.

Keywords: Hard-to-reach populations; hidden populations; time-location sampling; time-space sampling; respondent driven sampling; capture-recapture.

86 ***Corresponding Authors. Abdolreza Shaghaghi**. Fax: 0098 411 3340634, Tel: 0098 411 3357580
E-mail: shaghaghir@tbzmed.ac.ir

Introduction

'Hard-to-reach' is a term used to describe those sub-groups of the population that are difficult to reach or involve in research or public health programmes due to their physical and geographical location (e.g. in mountains, forests or deserts) or their social and economic situation [1]. An alternative term 'hidden population' is sometimes used in the literature especially to refer to those who do not wish to be found or contacted (e.g. illegal drug users or migrants and homeless people) [2]. Application of a single term to call these subsections of populations implies a homogeneity within the distinct groups which does not necessarily exist [3]. Studies on hidden populations raise a number of issues which are usually less important when doing research involving known populations. Hard-to-reach populations are generally floating populations and socially invisible thus gaining access to them poses major barriers for their recruitment [4]. Hard-to-reach populations may also actively try to conceal their group identity [5] due to fear of confrontation with legal authorities (e.g. drug users) or simply because of social pressure they feel from other members of the broader community. Sensitivity of the variables under study also adds to the potential difficulties a researcher may face when working with hard-to-reach groups. Even after reaching and recruiting an individual from a hidden population actual or perceived threat from legal authorities when doing research on stigmatized or illegal behaviors can increase probability of concealing a particular behavior or characteristic. People may not agree to cooperate in a study if they feel their anonymity may be violated by their participation. Hard-to-reach populations may thus be characterized by a group of disadvantage attributes such as illiteracy or being uncooperative, but this is not invariably the case and this characterization does not necessarily extend to all hard-to-reach populations. In addition, even with distinguishing hard-to-reach populations by these underprivileged features we should try to avoid the stigma associated with this term

It is suggested that faith-based communities (when there are clashes between faiths) and newly arrived residents are among the hard-to-reach populations. Over-researched people also have been suggested to be generally reluctant to participate in research. People who feel they are disconnected from the mainstream political process also were added [6] to the list of hard-to-reach groups. Migrants are also among world's most hard-to-reach people due to their scattering on the host communities, living in temporary camps, cultural separateness or simply because of difficulties an outsider may experience to access the social network of a special migrant group (Box 1) [1].

Box 1: People who are sometimes categorized as being hard-to-reach or hidden
Those being under social pressure of the broader community
Those living in faith based communities
Those who fear of confrontation with legal authorities
Illiterates
Those who have no interest to be found or contacted
Migrants
Newly arrived residents
Over-researched people
Those living in remote physical and geographical location
Those living in vulnerable social and economic situation

The degree of consent to participate in a study by a certain hard-to-reach group depends on the characteristics of that group, recruitment method used and the subject of interest. A group may be hard-to-reach to some extent and some locations and not in all circumstances [6]. Even when studies are explicitly designed to reach socially excluded groups researchers generally face challenges in recruiting enough number of study participants in practice [7]. Cultural, economic and social factors or lack of a sampling frame can raise barriers to access a special subgroup of a population. Different sampling techniques were introduced so far to recruit hard-to-reach populations.

In this article, we have reviewed a range of approaches that have been used to widen participation in studies, and summarized their relative advantages and disadvantages.

Materials and Methods

We performed a Pubmed and Google search for relevant English language articles using the keywords and phrases: (hard-to-reach AND population* OR sample*), (hidden AND population* OR sample*) and ("hard to reach" AND population* OR sample*) and a consultation of the retrieved articles' bibliographies to extract empirical evidence from publications that discussed or examined the use of sampling techniques to recruit hidden or hard-to-reach populations in health studies. The search was carried out from March 2009 to June 2009 and updated in August 2011. The first author scanned titles and abstracts to select studies and papers for consideration. Further assessment of the full text was also done by first author and the those articles that discussed sampling techniques, their procedures and factors that influence success and failure of the techniques in hard-to-reach populations were included. Two other authors also scanned the retrieved articles for consideration and initial disagreements on the scientific evidence selection were resolved by consensus.

Current status of knowledge

Snowball sampling

Snowball sampling is a non-probability method used when the desired sample characteristic is rare or when the studied population is broader and more heterogeneous than that can be easily accessible through other more reliable sampling methods [8].

It is not always possible to undertake a probability method of sampling when, for example, there is not a complete or easily accessible sampling frame, which is common for certain groups of population including migrants [9]. In these circumstances, it can be very difficult or expensive to recruit study subjects. In the snowball sampling method reliance is based on referrals from initial known subjects to recruit new additional subjects. This method is often used when the population under investigation is hard-to-reach due to their special characteristics or sensitivity of the study subject [10]. The known cases may be contacted to acquire needed data. If these known cases mutually agree to participate in the study they will be asked to nominate and facilitate introductions to other people whom they know according to the interpersonal relations and connections between people. Accordingly, the introduced nominee will be contacted and invited to participate in the study and with consent he or she is asked to introduce other people who also fulfill the study inclusion criteria [8, 10].

This strategy is regarded as a potential solution to the problem of sampling concealed populations [2]. The basic assumption in snowball sampling is that a link exists between the initial known subjects and others in the same target population. If this assumption is accurate it will allow a chain of acquaintance to be created

originating from primary contacts [2]. The main value of snowball sampling is its usefulness where some degree of trust is needed to initiate study subjects' recruiting process. Limited validity of data resulting from selection bias is the most important consideration for snowball sampling. Therefore findings from data gathered through snowball sampling are, it is suggested, not easily generalisable to the target population. It is also discussed that in snowball sampling the emphasis is on the inter-relationships, which isolates those who are not connected to any social network and thus are under-represented in the sample. The problem of selection bias is recommended to be solved relatively through selection of large sample and also by replication of results [2].

Respondent-driven Sampling (RDS)

As Heckathorn et al. [11] stated the main criticism about chain-referral or snowball sampling is bias toward recruiting more cooperative subjects and masking which is protecting close friends or relatives by not referring them when specially there is a strong privacy concern associated with the subject of the study. It is also suggested that those with extended personal networks to be over-sampled and isolated people to be excluded in the study.

Developed by Heckathorn [12] in 1997, RDS is a form of chain-referral sampling that was designed to eliminate the above mentioned sources of bias that are not inherent in the method. In order to facilitate recruitment procedure three mechanisms were suggested to be employed in RDS. These are: use of recruitment incentives (e.g. payment for participation and also for recruiting peers), limiting the number of recruits permitted per participant and not violating participants' confidentiality by letting them decide whether to become known to researcher or not (respondents should be recruited by their peers rather than by researchers) [11]. RDS combines snowball sampling with a mathematical model that weights the sample to compensate for the fact that the sample was collected non-randomly [12]. But still many open questions remain with RDS including bias, which can emerge from variable recruitment success rate by different types of people in an individual study [13].

Indigenous field worker sampling (IFWS)

In this sampling method instead of using formal trained investigators, they are selected from local community. Then they undergo special training relevant with objectives of the study including interview skills and fieldwork protocol. The selected people should have privileged access to the study target population [14, 15]. It is believed that use of this technique can reduce masking, volunteer bias and under-reporting of socially undesirable behaviors [16, 17].

The indigenous fieldworkers track down individuals known to them within the target area and recruit them into the study. Data collection takes place in the community setting separate from the rest of research team. An incentive is given to participants and they are asked to introduce their peers to the interviewer. To ensure wide coverage of the target population use of multiple sites and recruitment networks is recommended. Data collectors' safety and steady progress of recruitment process are main reported concerns [14].

Facility-based sampling (FBS)

Facility-based sampling refers to recruiting members of target population from a variety of facilities including correctional and drug treatment centers, sexually transmitted diseases clinics or general health centers and hospitals in certain suburban areas [18]. Each of these facilities can be used to recruit individuals from hidden population, but similar biases may occur due to under-sampling of those who are reluctant to seek and obtain services

especially when their behaviors are stigmatized. Other limitations of this sampling method include that in many part of the world, particularly in less developed countries, dedicated services to high risk groups are not common and even where provided; equal access to them by deprived subgroups of population is not guaranteed [18].

Targeted sampling (TS)

The targeted or purposive sampling method has been developed to overcome the limitation of snowball sampling when we would like to include specific pre-defined subgroups of population in our sample [18, 19]. This sampling method generally includes an initial assessment aimed at identifying the various subgroups that might exist in the population of interest. The identified subgroups are then regarded as sampling strata, which should have a pre-defined quota in the final sample. The magnitude of success in this sampling method depends mainly on thoroughness of the initial assessment and to some extent the time and resources available for its undertaking [18]. Application of this method can reach readily accessible subgroups very quickly [20], but on the other hand reaching isolated people will be very time consuming and therefore expensive.

Time-location (space) sampling (TLS)

Some members of hidden populations e.g. migrant workers tend to gather at certain types of location within the community and therefore time-location sampling is used to recruit these groups of hard-to-reach populations at locations where they may be found [18, 21]. Generally TLS begins with a formative phase of mapping different venues and establishments where individuals from hidden groups are known to congregate. The mapping generates a sampling frame of venues and time periods through which recruitment of individuals from a specific group of a hidden population will be possible [22, 23]. At a later stage the sampling frame is divided into venue-day-time increments that form the unit of random sampling [22]. These steps are necessary to ensure inclusion of individuals with varying venue and time attendance patterns [23].

It is suggested that bias due to masking and chain-referral selection of study participants is eliminated in TLS [21] however, unless a high percent of venues where members of hidden populations gather are identified and a very high percent of members from target population visit such locations, TLS also can suffer from potentially considerable bias. Isolated people for instance who do not visit such locations will be under-represented in the sample [18].

Conventional cluster sampling (CCS)

Conventional cluster sampling can be applied in limited circumstances to recruit hard-to-reach population. Cluster sampling is reasonable when there is no list of people to be selected, but a good list of locations where individuals from hard-to-reach group are gathered. The primary presumption for use of this sampling method is that the distribution of the variable of interest is similar between locations (clusters) [18, 24]. Clusters then are randomly selected for data gathering and thus recruitment costs will be minimized since the number of locations from which recruitment take place reduces. Clusters can be perinatal clinics, drug treatment centers, restaurants or health centers. Other requirement to be met in using this method is the possibility of ready access to all individuals from the population of interest in the clusters; otherwise cluster sampling will be an infeasible option to reach hidden populations.

Capture re-captures sampling method (CR)

This method which also known as contact re-contact method originally comes from application of a series of procedures to estimate the numbers of wild animals in

nature [25] but recently it was also indicated to be suitable to gauge population size, coverage of census or sampling success rate in epidemiological studies [26-30]. In employing this method, the primary assumption is that the population under study is stable during the observation period. To estimate the coverage or sampling success rate (R_{c-s}) at least two independent observers will try to contact study target group.

The number of people recruited by the first (n_1) and the second observer (n_2) and also those who have been recruited by both observers (k) are needed to calculate R_{c-s} (Figure 1).

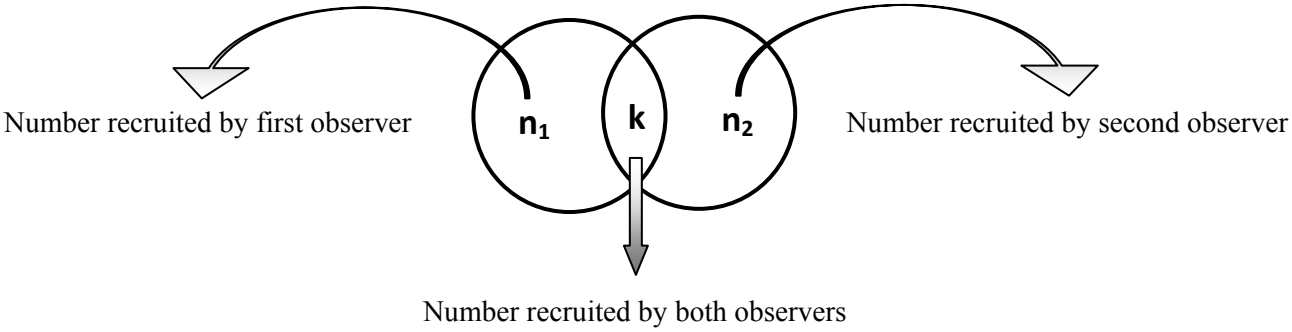

Figure 1: Capture re-capture sampling procedures

The estimated population size (N_E) will be computed using the formula below: [28]

$$N_E = \frac{(n1+1)\,(n2+1)}{(k+1)} - 1$$

To estimate 95% confidence interval (CI) for N_E the following formula may be utilized: [30]

$$95\%\ CI\ N_E = N_E \pm \sqrt{\frac{n1 * n2\,(n1-k)\,(n2-k)}{k^3}}$$

The coverage or sampling success rate (R_{c-s}) can be estimated by applying the following formula:

$$R_{c-s} = \frac{n_{1+}\, n_2 - k}{N_E} * 100$$

Two other assumptions that should be fulfilled to be able to use CR method is the same probability of being captured in both observations and also possibility of matching identified cases in two groups [28]. These calculations in two lists CR are simple but in the multiple lists CR a log-linear model should be fitted to the data [26, 28]. To prevent overestimation of N_E the total number of n_{1+} n_2 should be greater than N_E and k greater than 7 in two lists CR. Failure to meet all discussed prerequisites in using CR may lead to misleading results.

Conclusions

Recruitment of hard-to-reach populations: a holistic approach beyond techniques

To sum up, irrespective of potential advantages or limitations of the discussed techniques, their successful use depends mainly upon our knowledge about specific

characteristics of a target subgroup within a larger population. Without having such essential information it is very difficult to clarify which method will work best to recruit different hard-to-reach populations in varying settings and circumstances. Moreover our current knowledge about the recruitment techniques is based on their application in a wide range of topics and mostly within socially, culturally or behaviorally homogenous population subgroups. People belonging to a specific ethnic minority may be classified socially and culturally in a wider spectrum and are not necessarily homogenous. Thus in line with attempts to expand the current boundaries of our knowledge about recruitment techniques and their applications in varying situations, we should also focus on possibly all contributing factors which may have an impact on participation rate within a defined population group. Meticulous utilization of recruitment techniques could have great implications for health resource allocation towards hidden populations.

Acknowledgement

This paper is a part of the literature review section of PhD thesis A Shaghaghi has prepared under co-supervision of Professor R S Bhopal, Professor A Sheikh and Dr F Namdaran in the University of Edinburgh. A Shaghaghi prepared early draft of the article; RS Bhopal and A Sheikh provided critical comments on several drafts of the manuscript. The authors would like to thank Dr F Namdaran for his comments that helped to improve the early drafts of the paper. All authors approved the final draft.

Ethical approval

Not required.

Competing interests

None declared

References

[1] Managing Community Health Services. Bringing services to hard to reach populations. [Internet]. [2008?] [cited 2008 Jun 6]; Available from: URL:http://erc.msh.org/mainpage.cfm?file=2.2.3.htm&module=chs&language=English

[2] Atkinson R, Flint J. Accessing Hidden and Hard-to-Reach Populations: Snowball research strategies. [Internet]. [2001?] [cited 2008 Jun 6]; Available from: URL:http://sru.soc.surrey.ac.uk/SRU33.pdf

[3] Brackertz N. Who is hard to reach and why? [Internet]. [2007?] [Cited 2008 Jun 6]; Available from: URL:http://www.sisr.net/publications/0701brackertz.pdf

[4] Faugier J. Sampling hard to reach populations. J Adv Nurs 1997 Oct;26(4):790-7.

[5] Duncan DF, White JB, Nicholson T. Using Internet-based surveys to reach hidden populations: case of nonabusive illicit drug users. Am J Health Behav 2003 May-Jun;27(3):208-18.

[6] Brackertz N, Zwart I, Meredyth D, Ralston L. Community consultation and the 'hard to reach':concepts and practice in Victorian local government. [Internet]. [2005 Dec] [cited 2011 Sep 10]; Available from: URL:http://www.sisr.net/Flagships/democracy/docs/HardtoReach_append.pdf

[7] Emmel N, Hughes K, Greenhalgh J, Sales A. Developing methodological strategies to recruit and research socially excluded groups. [Internet]. [2007 Feb] [cited 2008 Jun 6]; Available from: URL:http://www.ccsr.ac.uk/methods/publications/documents/emmel.pdf.

[8] Council of Europe. Co-operation group to combat drug abuse and illicit trafficking in drugs (Pompidou Group). Handbook on snowball sampling. Strasbourg: Council of Europe; 1997.

[9] Gillies A. Using Research in Nursing: A Workbook for Practitioners. UK (Oxon): Radcliffe Publishing; 2002.

[10] Browne K. Snowball sampling: using social networks to research non-heterosexual women. Int J Soc Res Meth 2005;8(1):47-60.

[11] Heckathorn DD, Semaan S, Broadhead RS, Hughes JJ. Extensions of respondent-driven sampling: a new approach to the study of injection drug users aged 18–25. AIDS Behav 2002 March;6(1):55-67.

[12] Respondent Driven Sampling [Internet]. 2006 Apr 21 [cited 2008 Jun 17]; Available from: URL:http://www.respondentdriven-sampling.org/

[13] Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. Sociol Methodol 2004;34(1):193-239.

[14] Platt L, Wall M, Rhodes T, Judd A, Hickman M, Johnston LG, et al. Methods to recruit hard-to-reach groups: comparing two chain referral sampling methods of recruiting injecting drug users across nine studies in Russia and Estonia. Bull N YAcad Med 2006;83(7):i39-i53.

[15] Rhodes T, Platt L, Maximova S, Koshkina E, Latishevskaya N, Hickman M, et al. Prevalence of HIV, hepatitis C and syphilis among injecting drug users in Russia: a multi-city study. Addiction 2006 Feb;101(2):252-266.

[16] Power R, Harkinson S. Accessing hidden populations: the use of indigenous interviewers. In: Aggleton P, Davies P, Hart G, editors. Aids: facing the second decade. London: Falmer Press; 1993. p. 109–117.

[17] Power R. Some methodological and practical implications of employing drug users as indigenous fieldworkers. In: Boulton M, editor. Challenge and innovation: methodological advances in social research on HIV/AIDS. London: Taylor and Francis; 1994. p. 97–110.

[18] Magnani R, Sabin K, Saidel T, Heckathorn D. Review of sampling hard-to-reach and hidden populations for HIV surveillance. AIDS. 2005 May;19(Suppl 2):S67-72.

[19] Nonprobability Sampling [Internet]. 2006 Oct 20 [cited 2008 Jun 18]; Available from: URL:http://www.socialresearch-methods.net/kb/sampnon.php

[20] Watters JK, Biernacki P. Targeted Sampling: Options for the study of hidden populations. Soc Problems 1989 Oct;36(4):416-30.

[21] Karon JM. ASA Section on Survey Research Methods: The analysis of time-location sampling study data [Internet]. [2004?] [cited 2008 Jun 19]; Available from: URL:http://www.amstat.org/Sections/Srms/Proceedings/y2005/Files/JSM2005-000306.pdf

[22] Kendall C , Kerr LRFS, Gondim RC, Werneck GL, Macena RHM, Pontes MK et al. An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. AIDS Behav [Internet]. 2008 Jul [cited 2008 Sep 25]; 12(Suppl 1):97-104. Available from: URL:http://www.springerlink.com/content/v2u5r3r153137w75/fulltext.pdf

[23] Ferreira LOC, de Oliveira ES, Raymond HF, Chen SY, McFarland W. Use of time-location sampling for systematic behavioral surveillance of truck drivers in Brazil AIDS Behav [Internet]. 2008 Jul [cited 2008 Sep 25]; 12(Suppl 1):32-8. Available from: URL:http://www.springerlink.com/content/t863027t5038411n/

[24] Shepherd BE, Rossini AJ, Soto RJ, De Rivera IL, Mullins JI. Sampling designs for HIV molecular epidemiology with application to Honduras. AIDS Res Hum Retroviruses 2005 Nov 1;21(11):907-14.

[25] Marpsat M, Razafindratsima N. Survey methods for hard-to-reach populations: introduction to the special issue. Methodological Innovations Online (2010) 5(2) 3-16.

[26] Ismail AA, Beeching NJ, Gill GV, Bellis MA. How many data sources are needed to determine diabetes prevalence by capture-recapture? Int J Epidemiol 2000;29:536-41.

[27] Chao A, Tsay PK, Lin SH, Shau WY, Chao DY. The applications of capture-recapture models to epidemiological data. Stat Med 2001 Oct 30;20(20):3123-57.

[28] Gill GV, Ismail AA, Beeching NJ. The use of capture-recapture techniques in determining the prevalence of type 2 diabetes. QJM 2001 Jul;94(7):341-6.

[29] Gill GV, Ismail AA, Beeching NJ, Macfarlane SBJ, Bellis MA. Hidden diabetes in the UK: use of capture–recapture methods to estimate total prevalence of diabetes mellitus in an urban population. J R Soc Med 2003;96:328-332.

[30] Geibel S, van der Elst EM, King'ola N, Luchters S, Davies A, Getambu EM, Peshu N, Graham SM, McClelland RS, Sanders EJ. Are you on the market?': a capture–recapture enumeration of men who sell sex to men in and around Mombasa, Kenya. AIDS 2007, 21:1349–1354.