Management and Implementation of Sampling from Injecting Drug Users Exposed to High Risk Diseases

Arezo Bagheri, Mahsa Saadati

Arezo Bagheri
Assistant Professor of National Population Studies & Comprehensive Management Institute, Tehran, Iran

Mahsa Saadati
Assistant Professor of National Population Studies & Comprehensive Management Institute, Tehran, Iran
E-mail: mahsa.saadati@gmail.com

Abstract

Objective: The study of hidden populations, such as Injecting Drug Users (IDU) is very crucial due to their exposure to high risk diseases and their role in public health. Conventional statistical methods for sampling from these populations are not efficient because of the restrictions in attendance among these populations. Despite the introduction of respondent-driven sampling as a successful method for sampling from hidden populations, a limited number of studies have used this method due to the lack of researchers’ knowledge. Method: The main purpose of the current research was to study the influential factors in managing and implementing respondent-driven sampling method. Results: Researchers should consider key points in designing coupons, defining the degree and payment method of incentives, and preventing the irregular growth of the sample size for monitoring this sampling method. Conclusion: Without taking into consideration the requirements of sampling implementation and management, one cannot expect to achieve representative samples from these populations via this sampling method.

Keywords: hidden populations, injection drug users (IDU), high risk diseases, respondent driven sampling method
Introduction

Hidden populations or hard-to-reach populations refer to the sub-populations that usually constitute 1 to 10% of the total population of the society. Some of these populations, such as injecting drug users have received the attention of health researchers since this group of population is exposed to high-risk diseases, such as AIDS and threatens the health of the community. Sampling from these populations has particular problems since standard statistical sampling methods require a list of members of the population to be sampled. The use of family planning framework is inefficient when the desired population is small relative to the target population, is geographically dispersed, contains illegal behaviors (such as addiction) or consists of the networks that can be penetrated with difficulty (Watters & Biernacki, 1989). For example, researchers cannot sample injecting drug addicts through family plans because these societies often hide their habits from those who live with them.

In some studies on hidden populations, probabilistic sampling methods have been used, but these methods suffer from the issue of no absolute coverage of the target population (Ramirez-Valles et al., 2005). On the other hand, these methods do not include the addicts who come to these institutions. Institutional sampling is one of these sampling methods; for example, the addicts who are in jail or who refer to addiction treatment centers for drug abuse counseling have not been randomly selected.

Other studies in which non-probability sampling methods have been used for sampling from these populations provide a more comprehensive coverage of the target population compared to probability methods; however, the results of these methods are not statistically valid. Chain referral methods are among these methods. For example, in Snowball sampling method (Goodman, 1961), which is one of these sampling methods, a more complete coverage of the target community is provided through access to people via their social networks; however, it still faces inaccurate statistical estimates. In order to eliminate the interpretation problems of the results in chain referral methods, researchers attempted to modify and change the snowball sampling method and to convert it into probability sampling (Frank & Snijders, 1994). These methods are, in fact, a subset of the probability sampling methods that are referred to as adaptive designs or link tracing designs (Thompson & Frank, 2000). The respondent-driven sampling method is one of the variety of adaptive designs that was used by Heckathorn in 1997 for the first time in sampling from American drug addicts and is an efficient method to sample the hidden populations. This sampling method has been used in studies assessing the hidden populations, such as injecting drug users (Mumtaz et al., 2014; Yaung, Di Clemente, Halgin, Sterk & Havens, 2014; Stromer et al., 2006), sex workers (Liu, Liu, Cai, Rhodes & Hong, 2009), and Men who have Sex with Men (MSM) (Chopra et al., 2009).

In Iran, a pilot study was conducted to administer respondent-driven sampling method in Tehran in 2005, which demonstrated the effectiveness of
this method for studying the population of injecting drug users (Arzani et al., 2007). Injecting drug users aged 18 years and older in Tehran were also studied in a research on access to harm reduction programs (syringe replacement and methadone treatment programs) (Rahnema et al., 2014). In addition, a study was also conducted to estimate the prevalence of HIV infection among injecting drug users in Tehran in 2015 (Malekinejad et al., 2015). Bagheri & Saadati also introduced respondent-driven sampling method, compared it with other common sampling methods, and examined the methods of its estimation in 2014 and 2015.

Despite the introduction of this sampling method by Heckathorn nearly two decades ago, the complexities of the implementation and management of this sampling method have challenged its employment for researchers (Bagheri & Saadati, 2015; Saadati and Bagheri, a and b, 2016; Bagheri & Saadati a and b, 2016; Bagheri & Saadati, 2017). Therefore, the current study aims at investigating the implementation procedure of the respondent-driven sampling method and the considerations that lead to the efficiency in the management and implementation of this method, especially for studies in the field of drug addiction.

Method

Population, sample, and sampling method

One of the most important steps in designing the respondent-driven sampling method is the formative assessment that studies the diversity of social networks in the target population about demographic characteristics, the assessment of the target population's willingness to participate in the sample design, finding seeds for the start of recruitment chains, training of respondents (interviewers) in order to interview their own peers (subscribed) and decide on applicable issues, such as the use of incentives, the location where the questionnaire and interviews are to be conducted.

The next step is to select the seeds, which are, in fact, individuals with a degree (network size). Sampling from seeds begins and researchers randomly select them from the statistical population. The distinctive features of the seeds in injecting drug users are the number of years of drug injection, the preference of the type of consumed drugs, the use or non-use of the used syringes, and attempt or non-attempt for sex working.

Another important element in the implementation and management of this sampling method is the coupon. Coupon is, in fact, the quota of each person through which s/he can make his/her peers join the sample design. It has been mentioned that three coupons are usually assigned to each member in this method (Abdul-Quader, Heckathorn, Sabin & Saidel, 2006). Coupons provide information on the design, working time and location of the census bureau, the involvement of the recruiter and the recruited individuals through the allocation
of unique coupon identification numbers, assistance in the recruitment process, and the allocation management of incentives.

Another component that influences the way in which this sampling method is implemented and managed is the incentives that are assigned to the recruiters. Respondent-driven sampling design as a reward for the assignment of respect and value to the time and the effort that participants make towards recruitment and participation in the sampling (Verma, 2013). There are two types of these incentives. Primary incentives are the ones that are awarded to the recruiter for his/her participation in the sampling. On the other hand, secondary incentives are the ones the recruiter receives after the recruitment of his/her peers in the sampling.

In the primary stages of the respondent-driven sampling method, formative assessments are carried out and, then, this sampling method commences. Researchers non-randomly choose the seeds from the statistical population for membership in the sample. The seeds that have completed the sampling process as recruiters receive some coupons to recruit their peers or recruited members. The first wave of sampling is produced by the recruitment of seeds. In the following waves, each of the recruited members becomes the recruiter of the new wave. The first wave recruitment produces the second wave, and the sampling continues to reach the appropriate sample size. Primary incentives as well as secondary incentives, if necessary, are awarded to all sample members. Figure (1) shows how different networks have been produced from different seeds. The seeds and their corresponding waves are called chains.

![Figure 1: Formation of different chains from different seeds](image-url)
The estimation of the proportion of hidden populations, especially the populations exposed to high-risk diseases is highly regarded by the policy-makers who are in contact with these populations. To calculate this estimate, the hypothetical population consists of two groups; for example, IDUs are divided into two groups, i.e. HIV positive (group A) and HIV negative (group B). When the whole information about the social network exists in variable $X$, $x_{ij} = 1$ if there is a direct relationship between individuals $i$ and $j$; otherwise, $x_{ij} = 0$. In this sampling method, relationships are reciprocal in such a way that if $x_{ij} = 1$, then $x_{ji} = 1$. The degree of the $i$th individual is defined as $D_i = \sum x_{ij}$, and $T_a$ represents the total number of relations that arise from individuals in group A. If $N_a$ is the number and $D_a$ is the sum of the degrees of members of group A, then, $T_a$ is defined as:

$$T_a = \sum_{i \in A} D_i = N_aD_a \quad (1)$$

If the social network $X$ is assumed, the probability of a random relationship that connects a member of group A to a member of group B is defined as follows:

$$S_{ab} = \frac{T_{ab}}{T_a} \quad (2)$$

where $T_{ab}$ is the number of nodes that are members of group A and B.

The number of connections from group A to group B is equal to the number of connections from group B to group A. This number of connections can be calculated from the number of connections that have been derived from group A ($T_a$ in the probability that one of these connections with an individual from group B is $S_{ab}$ ($T_aS_{ab} = T_{ab}$)) and from group B ($T_b$ in the probability that one of these connections with an individual from group A is $S_{ba}$ ($T_bS_{ba} = T_{ab}$)). Assuming the number of connections to be equal with each other and using $T_a$ and $T_b$ definitions in Equation (1), the following equation is obtained:

$$N_aD_aS_{ab} = N_bD_bS_{ba} \quad (3)$$

If both sides of equation (3) are divided by N, i.e. the total population, then $P_a$ and $P_b$ in $P_aD_aS_{ab} = P_bD_bS_{ba}$ can be computed. In this way, by considering $P_a + P_b = 1$ and using the system resolution with two equations and two unknowns, the values of population ratios can be calculated from the following equations using information about the structure of the networks that connect the groups in the population to each other:

$$P_a = \frac{S_{ba}D_b}{S_{ba}D_b + S_{ab}D_a} \quad (4)$$

The estimate of $P_b$ can also be calculated similarly.

**Results**

This section examines the key points in the management and implementation of respondent-driven sampling method, which include the design, coupon
identification and management techniques, incentive management, sample
growth control, and recruitment termination.

A) Design and methods for the identification and management of coupons:

Coupons play an important role in the implementation process of the respondent-
driven sampling method whose design method will be discussed below.

* Essential and basic characteristics: Each coupon must contain information,
including a unique identification number in the design, location, working hours
and days of the Census Directorate, phone number, and expiration date. The
factors effective in the design of coupons include the spoken language, the level
of education, and age of the participants in the design, the number of Census
Directorate, the various populations that are being sampled at the same time, the
number and type of various logos belonging to the supporting organizations of
the design that is printed on the coupon, and the variety of the available papers
for printing the coupons. In order not to have a coupon shortage at the
Directorates, it is better to print a number of coupons three times as large as the
sample size. If more than one respondent-driven sampling design is being
performed simultaneously, it would be better to consider different colors for the
coupons of these designs. The size of coupons should not be too small to be lost
nor be too large to be maintained with difficulty.

* Expiration Date: It refers to the time framework during which a participant
is expected to deliver a coupon to his/her peer for recruitment and the peer is to
deliver it to the Census Directorate. There are many reasons to put an expiration
date on coupons; for example, the increase in the recruiter and new member’s
enthusiasm for the delivery of coupons, the management of coupons, and the
determination of the probabilistic number of participants involved in the design,
and the possibility of termination of the design in case of the absence of valid
coupons in the statistical population (World Health Organization, 2013).

* Activating time period: It is inserted on the coupon and it begins from the
moment the participant leaves the Directorate until the moment the coupons a
re delivered to his/her peers (1 to 3 days). The main objective of this period is to
reduce the recruiting speed and provide an incentive for the random recruitment
of peers.

Coupon Identification Methods: All coupons have identification numbers for
the recognition of both the recruiter and recruited person. Coupon identification
method depends on the implementation method of the coupon in two ways as
follows. Systematic identification method: This method is useful when there is
no computer in the Census Directorate to manage the recruitment of participants
in the design (Verma, 2013). Depending on the number of seeds in the design,
this method begins by assigning a unique number to each of the seeds. For
example, in a 10-seed design, the first two digits of each coupon are the seed
number, which are from 1 to 10. In most of respondent-driven sampling designs,
up to three peers are eligible for recruitment. In this state, the number that
appears following the seed number is actually the three coupons that are assigned
to the seed for recruitment. For example, if the seed number is 5 and three coupons are assigned to it, then, as shown in Fig. 2, it will receive the coupons numbered 51, 52, and 53 (World Health Organization, 2013). If the recruited member numbered 53 is interviewed, s/he will be assigned coupons numbered 531, 532, and 533. This process proceeds based on the number of waves per design. As a result, the coupon numbered 533 represents the second wave of the seed numbered 5. Among the advantages of this approach, one can refer to the ease of the management process of coupons, addition of seeds, and the ease of determining which seed member and wave belong to the recruited person. The disadvantages of this method are the need for accurate numbering and the possibility of assigning very long numbers, which raises the possibility of an error.

![Figure 2: Coupon Numbering for Seed No. 5](image)

* Sequential Identification Method: This method uses 4-digit numbers for the numbering of coupons in the design, where these numbers range from 1000 to the last number required for all coupons. In this method, coupon identification numbers are inserted and printed on them beforehand. The advantages of this approach include easy numbering and follow-up of the recruitment process when using a computer, printing coupon identification numbers on coupons before the design implementation, and short coupon numbers (only four digits). Among the disadvantages of this method, one can refer to the lack of direct supervision over the recruitment process in the design, the possibility of error occurrence in the numbering of the coupons due to the irrelevance of coupon numbers with each other, and the inapplicability of this method in the absence of a computer system for numbering.

Coupon Management System: Coupon management is performed to follow up the recruiter and the recruited person, to ensure that the incentives are given to the participant correctly, to follow up the number of completed waves in the design, to follow up the seeds' chains that are in the growing process and the seeds to be added to the design, to assess and decide on when the number of coupons should be reduced, when the design should be terminated, and when the
coupon should be stopped, and when the recruitment patterns should be analyzed.

B) Management of Incentives: Incentives are awarded to members without judgment about how they are used (Semaan, Santibanez, Garfèin, Heckathorn & Des Jarlais, 2009). The incentives should not be so small in value that illegitimate populations are reluctant to participate in the design, nor be so large in value that the likelihood of the sale or the theft of coupons increases. In addition, the high values of incentives make people who are not eligible for inclusion into the design to pretend to be qualified for participation. Verma (2013) referred to an East African sampling design on injecting drug addicts wherein the value of high incentives in the design had led many non-injecting drug users to pretend injecting drugs in order to receive incentives, and those who had participated in the design sold their coupons for the purpose of profitability.

* Types of incentives: Material and non-material incentives include such examples as food coupons and gifts, including clothes and telephone cards to be awarded to participants in the design.

* The motives for people's participation: Although it seems that the receipt of incentives is a clear and obvious factor for the participation of people in the design, this is not the main reason for their participation. In most of the respondent-driven sampling questionnaires, there are some question samples to examine the causes of coupon acceptance, such as the receipt of incentives and results of high-risk disease tests, acceptance of the recruiter's suggestion, finding the design interesting and helpful, and the availability of enough time (WHO, 2013). According to the results obtained from the addicts' sampling design, more than 60% of the female sex workers and Men who have Sex with Men in the Dominican Republic in 2008 announced the receipt of HIV test results as their main reason for participation (Johnston, Malekinejad, Kendall, Luppa & Rutherford, 2008).

Amount of incentives: A formative assessment is required to determine the appropriate incentives to the statistical population. The amount of incentives depends upon the project budget, living standards in the host country, government policies, and the population under study.

Requests for incentives: In order request for the primary incentives of participants in the design, they should enjoy conditions such as having a recruiting coupon (other than the seeds), eligibility for participation in the design, completion of the interview stages. In order to receive secondary incentives, the person usually needs to refer to the Census Directorate once more. It is suggested that the recruiter's secondary incentives not be awarded at the same time as receiving the primary incentives; in addition, the recruiter's incentives and the recruited person's incentives are suggested to be awarded on independent days (Verma, 2013).
C) Sample growth control and recruitment termination: The estimation of the length of the time required to achieve the optimal sample size in the respondent-driven sampling design is not an easy task, and it depends upon the statistical population under study, the optimal sample size, the number of seeds with which the design begins, the amount of incentives, and the size of social networks.

* Sample growth control: On the one hand, the high growth of sample size has caused some problems in the Census Directorate due to an increase in the number of visits; on the other hand, the sudden cessation of recruitment on the verge of achieving the desired sample size results in the accumulation of a large number of undelivered coupons in the statistical population. The effect of the exponential increase in the sample size can be reduced through the control of the linear growth rate of the sample and by limiting the number of coupons (Johnston et al., 2008).

* The end of the respondent-driven sampling design: In the respondent-driven sampling design, the Census Directorate should stay open until the payment of secondary incentives and provision of consultation to the people coming to the Directorate for participation in the design after completing the design. Information on the "completion of the design after reaching the desired sample size" should be inserted on the coupon. It should be ensured that the coupons have an expiration date that indicates the duration of their validity. The expiration date of the coupon must be some time after the date of delivery. When the desirable sample size is obtained, the expiration date of the coupons should be shortened so that the recruiter does not go to the Directorate after the end of the design. The delivery of coupons should be stopped when the design achieves the desired sample size. The Census Directorate would better stay open and active several weeks after the achievement of the desired sample size in the design in order to explain the reasons for the termination of the recruitment to the clients, provide health counseling for the recruited people with valid coupons, and award the secondary incentives.

Discussion and Conclusion

The study of hidden populations, such as injecting drug users, is of interest to policy-makers since these populations are at exposed to high-risk diseases, such as HIV and Hepatitis C (HVC) because of the shared use of the injection tools and also due to unprotected sex (Friedman et al., 2017; Grinsztejn et al., 2017; Visavakumet al., 2016; Nielsen et al., 2016; Wenz et al., 2016). Knowledge of demographic characteristics and the suffering level of these populations from high-risk diseases is essential because it is possible to adopt health care programs and counseling services for the individuals suffering these diseases that need as well as for the healthy people who are exposed to these infections.

Some studies on hidden populations of injecting drug addicts have been carried out through the commonly used sampling methods where the generalization of the results of these methods to these populations is not
statistically valid (Fotiou et al., 2016; Lewis et al., 2016). The respondent-driven sampling method, which is one of the chain-referral sampling methods, is an efficient sampling method that produces asymptotically unbiased estimates of the rate of hidden populations, such as injecting drug users due to the possession of strong social networks (Bagheri & Saadati, 2014 & 2015).

Therefore, the main objective of this study is to familiarize researchers interested in studying injection drug users with this sampling method, and also to point out the practical aspects in order to increase the efficiency of using this sampling method in such studies. In the following, some results of this study in the management and implementation of this sampling method are referred to.

The most important stages of the management and implementation of the respondent-driven sampling method include the design, coupon identification and management techniques (each recruiter's quota for recruiting from among his/her peers), incentive management (the amount paid to each recruiter to compensate for the time and cost of participating in the design), and sample growth control and recruitment termination (the time and completion procedure of the sampling). The key points that should be considered when designing coupons are the legibility and intelligibility of the coupons even for uneducated students. In addition, when the recruiting process goes ahead in a slow and an unexpected fashion, the number of coupons delivered to the recruiters should not be reduced. The slowness of the recruitment process in the first months of the census may have been arisen from the non-recruiting activity of some of the seeds. Coupon numbering can also manage efficient seeds, infertile seeds (seeds with no recruitment), and the number of waves completed by each seed (World Health Organization, 2013).

In addition to the payment of incentives, Verma (2013) referred to some successful financial incentives, such as financial payment equivalent to a common meal, the budget for one week's grocery shopping, fund for one week's urban transit and transportation. In a study carried out in 2008, a list of incentives in terms of their quantity and type has been presented. This list is the result of the experiences from 107 respondent-driven sampling methods conducted outside of the United States (Malkinezhad et al., 2008). The payment of secondary incentives has been controversial in some previous studies. The payment of these incentives leads to the participation of people with low socioeconomic status, the presence of repetitive and non-eligible participants, and also the identification of the individuals in the population who compete with others for recruitment. In the review of previous respondent-driven sampling designs, secondary incentives were of less value than the primary ones (Malekinejad et al., 2008).

It is also not an easy task to predict the completion time of the design except in the designs with a previous executive record. It should be noted that even if the sample size has reached the desired level and the expiration date of the design has not expired, each valid coupon coming to the Directorate should be included
in the sample. In case of the non-sampling and receipt of incentives for the volunteers for participation in the design, other designs that will be carried out in the future will be affected (Verma, 2013).

Considering the applied considerations addressed in this study towards the management and implementation of successful respondent-driven sampling method, researchers are advised to use this sampling method in their studies on addiction to achieve generalizable results to the population of injecting drug users.

Reference


