



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>



Research Article

Population Proportion Estimator of Respondent Driven Sampling for Non-dichotomous Variables, Data Smoothing Approach

Arezoo Bagheri and Mahsa Saadati

National Population Studies and Comprehensive Management Institute, Tehran, Iran

Abstract

Background: Sampling and estimating of hidden population sizes, such as injection drug users are important issues for health policy makers, because of exposing these populations to high risks diseases, such as HIV/AIDS. **Materials and Methods:** Respondent driven sampling is a successful method in terms of resulting in representative sample of hidden populations and finding unbiased estimates comparing to the other existing conventional methods. **Results:** The main purpose of this study is to define population proportion estimation of this sampling method for dichotomous and non-dichotomous variables. For non-dichotomous variables, reciprocal approach results in over-determination equations which can be solved by either least squares or data smoothing approaches, though the late one is much more effective. A hypothetical data has been employed to find the estimation of dichotomous and non-dichotomous variables for respondent driven sampling method. **Conclusion:** The novelty of data smoothing procedure to find respondent driven sampling estimates has been proved by this hypothetical data. Respondent driven sampling method could result in unbiased estimates of population proportions and it has been recommended to be applied for studying hidden population proportions.

Key words: Population proportion estimator, dichotomous variables, non-dichotomous variables, respondent driven sampling method, data smoothing approach

Received: October 18, 2016

Accepted: January 24, 2017

Published: March 15, 2017

Citation: Arezoo Bagheri and Mahsa Saadati, 2017. Population proportion estimator of respondent driven sampling for non-dichotomous variables, data smoothing approach. *J. Applied Sci.*, 17: 196-203.

Corresponding Author: Mahsa Saadati, National Population Studies and Comprehensive Management Institute, No. 5, Second Street, Pakestan Street, Shahid Beheshti Street, Tehran, Iran

Copyright: © 2017 Arezoo Bagheri and Mahsa Saadati. This is an open access article distributed under the terms of the creative commons attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Collecting accurate information about the behavior and composition of hidden social groups specially those exposing high risk and transmitted diseases such as HIV/AIDS is one of the vital issues of the most policy makers all around the world. Some well-known examples of such populations are Injection Drug Users (IDU)¹⁻³, sex workers⁴ and men who have sex with men (MSM)⁵.

Conventional sampling methods can not achieve a good prospective of these populations who influence the public health of society. These methods require a known probability of selection which means to have a sampling frame from all members of these populations. This information mostly does not exist⁶. One of the efficient methods of collecting information about hidden populations is institutional sampling method. Sampling IDU population by this method will be restricted for example to those attending to a drug rehabilitation program. In this way, any inferences from the resulting sample are not statistically valid⁷.

Another two most common approaches of sampling these populations are targeted and time-space sampling methods. Both of these methods treat the hidden population members discretely and they do not use population's networks relationships to estimate accurately. Targeted sampling or street outreach mostly can not reach a large number of non-institutional members of the hidden population⁸. The sample is not selected randomly and the probability selection of sample is unknown. Time-space sampling method introduced to select samples with known probability in the identified venues and ease inferences in this way⁹. Safety and cost concerns of some venues cause them not to be accessible for the researchers and the sample in this way is not representative of target populations¹⁰. Coverage problem of this method which makes unknown bias to the hidden population estimates by Stueve *et al.*¹⁰.

Another common approach is chain-referral sampling which can successfully penetrate in hidden populations and recruit their members but the estimations from these populations are statistically invalid. These methods has been called as non-probability methods due to have unknown probability of sample selection¹¹. One of the most applicable chain-referral sampling methods is snowball sampling, which is first introduced by Goodman¹².

A new sampling method, called Respondent Driven Sampling (RDS) introduced by Heckathorn^{13,14}. It is an altered method of snowball sampling. Since, last two decades of introducing RDS method, lots of researchers in different field of science, such as health and demography are studying and applying this sampling method¹⁵⁻²¹.

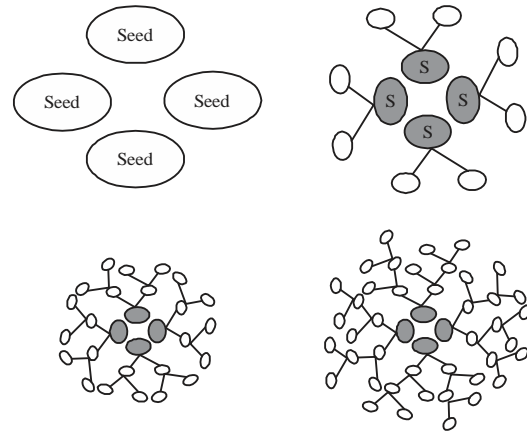


Fig. 1: Recruitment process in RDS method

The RDS sample is gathered in the same way as any chain-referral sampling method. Figure 1 presents the sampling process of RDS method. In this study, sampling starts from initial members, which are called seeds. Seeds are non-random members of the sample that are selected by researchers from their interested hidden population. They are those members who have bigger social networks among the others. These new members (recruiters) start recruiting from their social networks and introduce new members to the sample (recruits). These new recruiters recruit new members and ties are formed. This procedure continues till the desired sample size is achieved.

The RDS is an indirect method of estimating from sample about the population. The sample is used to make estimations about social network connections and then this information could be applied to make asymptotically unbiased population proportion estimates⁶. The bias of these estimates is on the order of sample size inverse, so, it is negligible for meaningful sample sizes²².

To find population estimators of RDS samples, there are different approaches for dichotomous (for example, HIV+, HIV-) and non-dichotomous variables (for example, quartiles of age, 20, 30 and 30-40 and etc). The main aim of this study is to review estimator for dichotomous variables and define data smoothing estimators for non-dichotomous variables.

MATERIALS AND METHODS

The RDS recruitment network reflects pre-existing social relationships that link recruiters and recruits. These relationships are reciprocal. In this way, T_{ab} the number of ties from A-B equals those from B-A that means as shown in Eq. 1:

$$T_{ab} = T_{ba} \tag{1}$$

The number of such crosscutting ties depends on three following factors in Eq. 2:

$$T_{ab} = N_a D_a S_{ab} \quad (2)$$

The same results can also be concluded for T_{ba} . Setting these two equations in Eq. 1 and dividing 2 parts of the equation to N , population size can get following Eq. 3:

$$P_a D_a S_{ab} = P_b D_b S_{ba} \quad (3)$$

By substituting $(1-P_a)$ for P_b , the Eq. 3 will be an estimate of population size of group A, P_a , based on the reciprocity model given Eq. 4:

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

The same results can also be concluded for P_b . Equation 4 provides an estimation of the proportional size of the hidden population according to the transition probabilities based on recruitment patterns and self-reported personal network size^{14,23}.

To find unbiased estimators of the population proportion size, the transition probabilities and personal network size should be estimated. A recruitment matrix, R , where R_{ab} is the number of recruitments by the members of group A of members of group B could be considered as in Eq. 5:

$$R = \begin{bmatrix} R_{aa} & R_{ab} \\ R_{ba} & R_{bb} \end{bmatrix} \quad (5)$$

According to this matrix, the unbiased estimators of S_{ab} and S_{ba} can be computed as in Eq. 6:

$$\widehat{S}_{ab} = \frac{R_{ab}}{R_{aa} + R_{ab}} = \frac{R_{ab}}{RB_a} \quad (6)$$

$$\widehat{S}_{ba} = \frac{R_{ba}}{R_{bb} + R_{ba}} = \frac{R_{ba}}{RB_b}$$

where, $RB_a = R_{aa} + R_{ab}$ and $RB_b = R_{bb} + R_{ba}$. The second element of RDS population size estimator is the mean degree of each group which could be estimated as Salganik and Heckathorn⁶ given in Eq. 7:

$$\widehat{D}_a = \frac{n_a}{\sum_{i=1}^{n_a} \frac{1}{D_i}} \quad (7)$$

The RDS is based on the theories of Markov chain model so the recruitment process in RDS has the following characteristics.

A memory less process: The recruitment pattern of each recruiter depends only on its own recruiter. It has been termed as a first Markov process characteristic by Heckathorn¹³.

An ergodic process: A recruiter with one set of characteristics recruiting another subject with the same or different characteristics. After one or more recruitment waves, a recruit can have the same characteristics as the earlier recruiter.

To reduce the bias of not selecting RDS seeds randomly, the recruitment should be continued until equilibrium is reached. To compute equilibrium analytically, the law of large numbers for Markov chains can be applied. The equilibrium for a system of two groups, A and B could be defined as in Eq. 8:

$$E_a + E_b = 1 \quad (8)$$

$$E_a = S_{aa} E_a + S_{ba} E_b$$

Solving this Eq. 8 system results in Eq. 9:

$$\widehat{E}_a = \frac{S_{ba}}{S_{ab} + S_{ba}}, \widehat{E}_a = 1 - E_b \quad (9)$$

Population proportion estimator for non-dichotomous variables:

Reciprocity model for a system with N groups can be shown by a system of equations, where the 1st equation express equality of summing the population proportion size of groups to one. Each of the other $\frac{N(N-1)}{2}$ equations shows reciprocity principals for each of the other two groups that can be shown as following in Eq. 10 and 11 for a 4 group system of A, B, C and D:

$$1 = P_a + P_b + P_c + P_d \quad (10)$$

$$\begin{aligned} P_a D_a S_{ab} &= P_b D_b S_{ba} \\ P_a D_a S_{ac} &= P_c D_c S_{ca} \\ P_a D_a S_{ad} &= P_d D_d S_{da} \\ P_b D_b S_{bc} &= P_c D_c S_{cb} \\ P_b D_b S_{bd} &= P_d D_d S_{db} \\ P_c D_c S_{cd} &= P_d D_d S_{dc} \end{aligned} \quad (11)$$

This system of equations are over-determined because the number of unknowns and equations are not equal. This problem could happen for non-dichotomous variables (any three or more categories variables). There are two different approaches for dealing with this problem. Linear least square

is a standard solving process for these systems which has the same logic as linear regression²⁴. Another approach that is explained in the study is an alternative method for calculating the population proportion estimates that is drawn from the same logic as reciprocity models.

Data smoothing approach: Data smoothing approach as a solution for overcoming the problem of over-determination in finding population proportion estimates for non-dichotomous variables has been introduced by Heckathorn¹⁴. The main idea in this study which comes from reciprocity models is about equality of the number of recruits, RO and the number of recruitments, RB for each group. By considering random recruitments from personal networks, cross-group recruitments will be equal for each pair of groups. The best estimate for the number of cross-recruitments between each pair of groups is the mean of recruitments in each direction. In this way, the problem of over-determination is solved by reducing the number of terms from which population estimates are calculated. This approach results in more efficient estimates comparing to linear-least squares approach²⁵. The point estimations are not affected for dichotomous variables by this approach, though for estimating variance it could be effective²⁵. This study can be done in the following steps.

Ranking or demographic adjustment step: That could be done by transforming the recruitment matrix using two conditions of not changing recruitment pattern and equaling the row and column sums for recruitment matrix (for any group, RO=RB). Each element in the transformed recruitment matrix (R_{ab}) is the product of three terms of the selection proportion (\widehat{S}_{ab}), the equilibrium for the recruiter's group (\widehat{E}_a) and the total number recruitments (RB). For a system with N group, the adjusted recruitment matrix R^* is given in Eq. 12:

$$R^* = \begin{bmatrix} \widehat{S}_{aa}\widehat{E}_aRB & \widehat{S}_{ab}\widehat{E}_aRB & \dots & \widehat{S}_{aN}\widehat{E}_aRB \\ \widehat{S}_{ba}\widehat{E}_bRB & \widehat{S}_{bb}\widehat{E}_bRB & \dots & \widehat{S}_{bN}\widehat{E}_bRB \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{S}_{Na}\widehat{E}_N RB & \widehat{S}_{Nb}\widehat{E}_N RB & \dots & \widehat{S}_{NN}\widehat{E}_N RB \end{bmatrix} \quad (12)$$

$$= \begin{bmatrix} R_{aa}^* & R_{ab}^* & \dots & R_{aN}^* \\ R_{ba}^* & R_{bb}^* & \dots & R_{bN}^* \\ \vdots & \vdots & \ddots & \vdots \\ R_{Na}^* & R_{Nb}^* & \dots & R_{NN}^* \end{bmatrix}$$

So for any groups, such as A, $RO_a^* = RB_a^*$ and $\frac{RO_a^*}{RO} = \frac{RB_a^*}{RB} = \widehat{E}_a$.

Smoothing step: The best estimate for demographically adjusted recruitment counts are smoothed counts which are calculated by finding the mean of these counts across groups. In this way, the smoothed matrix, R^{**} is found as in Eq. 13:

$$R^{**} = \begin{bmatrix} \frac{\widehat{S}_{aa}\widehat{E}_aRB + \widehat{S}_{ba}\widehat{E}_bRB}{2} & \dots & \frac{\widehat{S}_{aN}\widehat{E}_aRB + \widehat{S}_{Na}\widehat{E}_N RB}{2} \\ \frac{\widehat{S}_{ab}\widehat{E}_aRB + \widehat{S}_{ba}\widehat{E}_bRB}{2} & \dots & \frac{\widehat{S}_{bN}\widehat{E}_bRB + \widehat{S}_{Nb}\widehat{E}_N RB}{2} \\ \vdots & \ddots & \vdots \\ \frac{\widehat{S}_{aN}\widehat{E}_aRB + \widehat{S}_{Na}\widehat{E}_N RB}{2} & \dots & \frac{\widehat{S}_{NN}\widehat{E}_N RB}{2} \end{bmatrix} = \begin{bmatrix} R_{aa}^{**} & R_{ab}^{**} & \dots & R_{aN}^{**} \\ R_{ba}^{**} & R_{bb}^{**} & \dots & R_{bN}^{**} \\ \vdots & \vdots & \ddots & \vdots \\ R_{Na}^{**} & R_{Nb}^{**} & \dots & R_{NN}^{**} \end{bmatrix} \quad (13)$$

The R^{**} becomes the basis for all the other calculations and the smoothed population proportion estimate could be computed according to this matrix.

RESULTS

In this study, by considering a hypothetical population which is made of two groups, Injection Drug User (IDU) with positive HIV (HIV⁺), group A and Injection Drug User (IDU) with negative HIV (HIV⁻), group B, the population proportion estimates are calculated. Table 1 presents this population consisting of 30 cases. According to Table 1, ID is respondent identification, IDR is respondent identification of each respondent's recruiter, D is respondent's self-reported degree and V is a dichotomous variable (IDU with HIV⁺ and IDU with HIV⁻) and U is a non-dichotomous variable (educational levels, (1) <Diploma, (2) Diploma and bachelor and (3) >Bachelor).

Respondent driven sampling estimates for dichotomous variables: The population proportion estimates for these two groups, IDU with HIV⁺ and IDU with HIV⁻ are computed by RDS estimators. According to the Table 1, the first respondent is seed, so it doesn't have any recruiter. Note that the degree data for cases of 2, 16 and 26 are missing.

Table 1: Characteristics of hypothetical population

ID	IDR	D	V	U	ID	IDR	D	V	U	ID	IDR	D	V	U
1	NA	10	A	1	11	5	6	B	2	21	12	8	B	3
2	1	NA	A	1	12	6	8	B	2	22	14	9	A	3
3	1	10	B	2	13	7	7	A	1	23	14	5	B	2
4	1	8	A	1	14	7	8	A	3	24	15	4	A	1
5	2	5	B	3	15	8	10	A	3	25	15	5	A	2
6	3	3	B	2	16	9	NA	A	1	26	15	NA	B	2
7	3	8	A	1	17	9	3	B	2	27	16	3	B	1
8	3	6	B	1	18	10	2	B	1	28	17	6	A	1
9	4	8	A	2	19	11	1	B	2	29	18	7	A	3
10	4	4	A	3	20	12	9	A	3	30	19	4	A	3

Table 2: Recruitment of matrix of IDU with HIV+ and IDU with HIV- groups

Group of recruiters	Group of recruits (selection proportion)		
	Group A	Group B	RB
Group A	10	7	17
	-0.59	-0.41	
Group B	6	6	12
	-0.5	-0.5	
RO	16	13	29
Recruitment proportion	0.55	0.45	1

Table 2 presents recruitment matrix of data in Table 1. As it is presented in Table 1, the first respondent (seed) is a member of group IDU with HIV+, recruited respondents 2, 3 and 4, respondents 2 and 4 are members of group IDU, respondent 3 is a member of group IDU with HIV-. So, transition probabilities could be computed according to this recruitment pattern. Table 2 also consists of recruitment proportions which are computed from the recruitment counts.

The number of each group can be found from extracting seeds and members with missing data of degree, so $n_a = 17 - 1 - 2 = 14$ and $n_b = 13 - 1 = 12$. The estimated degrees for these two groups according to the Eq. 7 are:

$$\widehat{D}_a = \frac{n_a}{\sum_{i=1}^{n_a} \frac{1}{D_i}} = \frac{14}{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} + \frac{1}{7} + \frac{1}{8} + \frac{1}{10} + \frac{1}{9} + \frac{1}{9} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{4}} = 6.29$$

$$\widehat{D}_b = \frac{n_b}{\sum_{i=1}^{n_b} \frac{1}{D_i}} = \frac{12}{\frac{1}{10} + \frac{1}{5} + \frac{1}{3} + \frac{1}{6} + \frac{1}{6} + \frac{1}{8} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1} + \frac{1}{8} + \frac{1}{5} + \frac{1}{3}} = 3.35$$

The population estimates for these two groups also are shown in Eq. 14:

$$\widehat{P}_a = \frac{\widehat{S}_{ba} \widehat{D}_b}{\widehat{S}_{ba} \widehat{D}_b + \widehat{S}_{ab} \widehat{D}_a} = \frac{(0.5)(3.35)}{(0.5)(3.35) + (0.41)(6.29)} = 0.39 \quad (14)$$

$$\widehat{P}_b = 1 - 0.39 = 0.61$$

Table 3: Recruitment matrix of three educational levels

Educational levels of recruiters	Educational levels of recruits			RB
	<Diploma	Diploma and bachelor	>Bachelor	
<Diploma	4	2	5	11
Diploma and bachelor	4	4	3	11
>Bachelor	2	4	1	7
RO	10	10	9	29

Table 4: Selection probabilities of three educational levels

Educational levels of recruiters	Educational levels of recruits		
	<Diploma	Diploma and bachelor	>Bachelor
<Diploma	0.364	0.182	0.454
Diploma and bachelor	0.364	0.364	0.273
>Bachelor	0.286	0.571	0.143

Table 5: Equilibrium of three educational levels

<Diploma	Diploma and bachelor	>Bachelor
0.342	0.363	0.295

$$\widehat{E}_a = \frac{\widehat{S}_{ab}}{\widehat{S}_{ba} + \widehat{S}_{ab}} = \frac{0.41}{0.5 + 0.41} = 0.45 \quad (15)$$

$$\widehat{E}_b = 1 - 0.45 = 0.55$$

The equilibriums for these two groups are calculated from transition probabilities as given in Eq. 15:

Respondent driven sampling estimates for non-dichotomous variables:

In this study, data smoothing procedure is examined in driving population proportion estimates for variables with three or more categories. Table 3 demonstrates recruitment matrix by educational levels. Table 4 also presents selection probabilities for this variable.

According to the selection probabilities in Table 4, solving following equations results in equilibrium of these three levels of education which has been shown in Table 5:

$$1 = E_a + E_b + E_c$$

$$E_a = 0.364 E_a + 0.364 E_b + 0.286 E_c$$

$$E_b = 0.182 E_a + 0.364 E_b + 0.571 E_c$$

Table 6: Demographically adjusted matrix of three educational levels

Educational levels of recruiters	Educational levels of recruits			RB
	<Diploma	Diploma and bachelor	>Bachelor	
<Diploma	3.61	1.81	4.5	9.92
Diploma and bachelor	3.83	3.83	2.87	10.53
>Bachelor	2.45	4.88	1.22	8.55
RO	9.89	10.52	8.59	29

Table 7: Data smoothed recruitment matrix of three educational levels

Educational levels of recruiters	Educational levels of recruits			RB
	<Diploma	Diploma and bachelor	>Bachelor	
<Diploma	3.61	2.82	3.48	9.91
Diploma and bachelor	2.82	3.83	3.88	10.53
>Bachelor	3.48	3.88	1.22	8.58
RO	9.91	10.53	8.58	29

Table 8: Data smoothed selection probabilities of three educational levels

Educational levels of recruiters	Educational levels of recruits		
	<Diploma	Diploma and bachelor	>Bachelor
<Diploma	0.364	0.285	0.351
Diploma and bachelor	0.268	0.364	0.368
>Bachelor	0.406	0.452	0.142

To find the demographically adjusted matrix that is presented in Table 6 similar to the demographically adjusted matrix in Eq. 7, each cell has been multiplied by the selection probabilities in Table 4, equilibrium in Table 5 and the total recruitment (RB).

Table 7 presents smoothed recruitment matrix according to the Eq. 13 that is found by averaging the cross-recruitment counts. Then this matrix could be employed to recalculate all the other terms such as data smoothed selection probabilities in Table 8 and estimated degrees:

$$\widehat{D}_a = \frac{n_a}{\sum_{i=1}^{n_a} \frac{1}{D_i}} = \frac{11-1-2}{\frac{1}{8} + \frac{1}{8} + \frac{1}{6} + \frac{1}{7} + \frac{1}{2} + \frac{1}{4} + \frac{1}{3} + \frac{1}{6}} = 4.42$$

$$\widehat{D}_b = \frac{n_b}{\sum_{i=1}^{n_b} \frac{1}{D_i}} = \frac{10-1}{\frac{1}{10} + \frac{1}{3} + \frac{1}{8} + \frac{1}{6} + \frac{1}{8} + \frac{1}{3} + \frac{1}{1} + \frac{1}{5} + \frac{1}{5}} = 3.48$$

$$\widehat{D}_c = \frac{n_c}{\sum_{i=1}^{n_c} \frac{1}{D_i}} = \frac{9}{\frac{1}{5} + \frac{1}{4} + \frac{1}{8} + \frac{1}{10} + \frac{1}{9} + \frac{1}{8} + \frac{1}{9} + \frac{1}{7} + \frac{1}{4}} = 6.36$$

According to the results of estimated degrees and smoothed selection probabilities, following equations could be concluded:

$$1 = \widehat{P}_a + \widehat{P}_b + \widehat{P}_c$$

$$\widehat{P}_a(4.42)(0.285) = \widehat{P}_b(3.48)(0.268)$$

$$\widehat{P}_a(4.42)(0.351) = \widehat{P}_c(6.36)(0.406)$$

So, the population proportion estimates for three educational levels are equal to $\widehat{P}_a = 0.34$, $\widehat{P}_b = 0.46$ and $\widehat{P}_c = 0.20$.

DISCUSSION

Most of present studies about hidden populations result in samples which have been collected by lots of effort and could not be generalized to the interested populations^{6,15-21,26,27}. So, in this way, some descriptive statistics could be drawn from these samples and no statistical inferences could be possible to conclude^{13,14}. It also leads researchers to some misleading conclusions⁶. The RDS method which has been introduced by Heckathorn^{13,14} and reviewed in this study could solve this problem. However, if the expected procedure in RDS couldn't be followed it will not result in more efficient estimators comparing to chain referral sampling methods²⁸. The RDS results in asymptotically unbiased population proportion estimates when theoretical and analytical assumptions of this sampling method could achieve⁶. In the condition that some of non-sampling errors, such as non-random selection of seeds, not-achieving to equilibrium in most important interested variables in the study and countering homophile which is recruiting non-randomly from social networks exist, RDS will not conclude in unbiased estimators^{6,13,14}. Moreover, no statistical inferences could be drawn from RDS samples.

For finding estimates of non-dichotomous variables in RDS method, reciprocal approach introduced by Salganik and Heckathorn⁶ and Heckathorn¹⁴ results in over-determination equations which can be solved by either least squares or data smoothing approaches. Data smoothing approach has been reviewed and calculated for a hypothetical data in this study. This method has been claimed that is much more effective in calculating unbiased estimators comparing to least squares^{6,13,14}.

CONCLUSION

The RDS sample could result in asymptotically unbiased estimates of population proportions, by collecting the information about social networks. By applying the theories behind reciprocal models, RDS results in population proportion estimates for dichotomous and non-dichotomous variables. The main purpose of this study is to introduce data smoothing that is an effective method to overcome over-determination problem in non-dichotomous variables case has been reviewed. An hypothetical data has been

applied to show the novelty of this method for finding population proportion estimates for non-dichotomous variables.

ACKNOWLEDGMENTS

The authors would like to acknowledge the following article extracted from a survey under the title of "Respondent Driven Sampling Statistical Inferences to Estimate Demographical Parameters" which is supported by National Population Studies and Comprehensive Management Institute in 2015 by the registered number of 20/18616.

REFERENCES

1. Mumtaz, G.R., H.A. Weiss, S.L. Thomas, S. Riome and H. Setayesh *et al.*, 2014. HIV among people who inject drugs in the Middle East and North Africa: Systematic review and data synthesis. *PLoS Med.*, Vol. 11. 10.1371/journal.pmed.1001663.
2. Young, A.M., R.J. DiClemente, D.S. Halgin, C.E. Sterk and J.R. Havens, 2014. Drug users' willingness to encourage social, sexual and drug network members to receive an HIV vaccine: A social network analysis. *AIDS Behav.*, 18: 1753-1763.
3. Stormer, A., W. Tun, L. Guli, A. Harxhi and Z. Bodanovskaia *et al.*, 2006. An analysis of respondent driven sampling with Injection Drug Users (IDU) in Albania and the Russian Federation. *J. Urban Health*, 83: 73-82.
4. Liu, H., H. Liu, Y. Cai, A.G. Rhodes and F. Hong, 2009. Money boys, HIV risks and the associations between norms and safer sex: A respondent-driven sampling study in Shenzhen, China. *AIDS Behav.*, 13: 652-662.
5. Chopra, M., L. Townsend, L. Johnston, C. Mathews, M. Tomlinson, H. O'Bra and C. Kendall, 2009. Estimating HIV prevalence and risk behaviors among high-risk heterosexual men with multiple sex partners: Use of respondent-driven sampling. *J. Acquired Immune Deficiency Syndromes*, 51: 72-77.
6. Salganik, M.J. and D.D. Heckathorn, 2004. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol. Methodol.*, 34: 193-240.
7. Watters, J.K. and Y.T. Cheng, 1987. HIV-1 (Human Immunodeficiency Virus-type one) infection and risk among intravenous drug users in San Francisco: Preliminary results and implications. *Contemp. Drug Problems*, 14: 397-410.
8. Watters, J.K. and P. Biernacki, 1989. Targeted sampling: options for the study of hidden populations. *Social Problems*, 36: 416-430.
9. Muhib, F.B., L.S. Lin, A. Stueve, R.L. Miller and W.L. Ford *et al.*, 2001. A venue-based method for sampling hard-to-reach populations. *Public Health Rep.*, 116: 216-222.
10. Stueve, A., L.N. O'Donnell, R. Duran, A. San Doval and J. Blome, 2001. Time-space sampling in minority communities: Results with young Latino men who have sex with men. *Am. J. Public Health*, 91: 922-926.
11. Kalton, G., 1983. *Introduction to Survey Sampling-Vol. 35 (Quantitative Applications in the Social Sciences)*. 1st Edn., SAGE Publications, California, ISBN-13: 978-0803921269, Pages: 96.
12. Goodman, L.A., 1961. Snowball sampling. *Ann. Math. Stat.*, 32: 148-170.
13. Heckathorn, D.D., 1997. Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44: 174-199.
14. Heckathorn, D.D., 2002. Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*, 49: 11-34.
15. Bagheri, A. and M. Saadati, 2014. A survey on link-tracing sampling methods (social networks) and its application on demography. National Population Studies and Comprehensive Management Institute, Iran.
16. Bagheri, A., 2016. Respondent driven sampling: A new approach to sampling rare and hidden diseases. *J. Health Syst. Res.*, 11: 753-761.
17. Saadati, M. and A. Bagheri, 2016. Respondent driven sampling method compared with other sampling methods of hidden populations. *Iran. J. Epidemiol.*, 12: 63-74.
18. Bagheri, A. and M. Saadati, 2015. Exploring the effectiveness of chain referral methods in sampling hidden populations. *Indian J. Sci. Technol.*, 8: 1-8.
19. Saadati, M. and A. Bagheri, 2015. Adaptive versus conventional sampling in demographic studies. *Proceedings of the 3rd Conference of Asian Population Association*, July 27-30, 2015, Kuala Lumpur, Malaysia.
20. Saadati, M. and A. Bagheri, 2015. Sampling migrants by respondent driven sampling method. *Proceedings of the the International Conference on Migration: Pattern, Consequences and Policies*, October 27-28, 2015, Tehran, Iran.
21. Bagheri, A. and M. Saadati, 2015. Sampling hard to reach populations: Comparing two chain referral approaches. *Proceedings of the 4th International Conference on Computer Science and Computational Mathematics*, May 7-8, 2015, Langkawi, Malaysia.
22. Erdos, P. and A. Renyi, 1959. On random graphs. *I. Publicationes Mathematicae*, 6: 290-297.
23. Heckathorn, D.D., 2007. Extensions of respondent-driven sampling: Analyzing continuous variables and controlling for differential recruitment. *Sociol. Methodol.*, 37: 151-207.

24. Farebrother, R.W., 1988. Linear Least Squares Computations. CRC Press, USA., ISBN: 9780824776619, Pages: 320.
25. Volz, E. and D.D. Heckathorn, 2008. Probability based estimation theory for respondent driven sampling. *J. Official Stat.*, 24: 79-97.
26. Saadati, M. and A. Bagheri, 2016. Homophily and unbiased estimators of respondent driven sampling method. Proceedings of the 13th Iranian Statistical Conference, August 23-25, 2016, Shahid Bahonar Kerman University, Iran.
27. Saadati M. And A. Bagheri, 2017. Unbiased Estimator of Population Proportion for Hidden Populations Exposing High Risk Diseases, *J. Health Syst. Res.*, Vol 12, No. 4, (In press).
28. Verma, V., 2013. Sampling elusive populations: Applications of child labour. ILO., Geneva. http://www.ilo.org/global/statistics-and-databases/WCMS_314425/lang-en/index.htm.