Scientific Research

Estimating the Size of an Injecting Drug User Population

Yang Zhao

Department of Mathematics and Statistics, University of Regina, Saskatchewan, Canada.
Email: zhaoyang@uregina.ca

ABSTRACT

This article describes a sampling and estimation scheme for estimating the size of an injecting drug user (IDU) population by combining classical sampling and respondent-driven sampling procedures. It is designed to use the information from prevention programs, especially, Needle Exchange Programs (NEPs). The approach involves using respondent-driven sampling design to collect a sample of injecting drug users who appear at site of NEP in a certain period of time and to obtain retrospective self-report data on the number of friends among the IDUs and number of needles exchanged for each sampled injecting drug user. A methodology is developed to estimate the size of injecting drug users who have ever used the NEP during the fixed period of time, and which allows us to estimate the proportion of injecting drug users in using NEP. The size of the IDU population is estimated by dividing the total number of IDUs who using NEPs during the period of time by the estimated proportion of IDUs in the group. The technique holds promise for providing data needed to answer questions such as "What is the size of an IDU population in a city?" and "Is that size changing?" and better understand the dynamics of the IDU population. The methodology described here can also be used to estimate size of other hard-to-reach population by using information from prevention programs.

*Keywords***:** *Injecting Drug Users, Needle Exchange Programs, Respondent-Driven Sampling, Estimating the Size of an IDU Population*

1. Introduction

It is reported that HIV epidemics are primarily driven by injecting drug use in Eastern Europe and Central Asia [1]. In Central and South America, injecting drug use and unsafe sex have been the main route of HIV transmission [1,2]. The negative health consequences of injecting drug use are not limited to just HIV infection. Sharing injection equipment carries a high risk of transmission of other blood-borne infectious diseases such as hepatitis B and hepatitis C. Also, injection drug use contributes to the epidemic's spread far beyond the circle of those who inject. Injection drug users (IDUs), their partners, and their children account for at least 36% of all AIDS cases reported in the U.S. through 1999 [3]. In Canada, injection drug use is also a problematic activity. Although it is difficult to obtain accurate data on the prevalence and profile of IDUs in Canada, it is clear that there are large numbers of IDUs across the country [4-6]. Policy makers and researchers have realized that data on the size of IDUs and pattern of drug injection need to be systemati-

cally collected for a comprehensive understanding of the HIV epidemics among IDUs [7-9]. Understanding something about the dynamics of the injection drug users makes it possible not only to assess the likely impact of the spread of HIV/AIDS and other related diseases, but also alert policy makers to a worsening situation, or alternatively to provide evidence so that other initiatives may be working. However, it is difficult to get reliable estimates of the number of IDUs by systematic surveillance. Policy makers and researchers have met the problem of collecting accurate information about IDUs, since they are not easily captured in a general population based survey (a typical general population based survey is to survey individuals in a random sample of households in a district, province or a country, depending on the scale of the study). Injectors sometimes even hide their habit from those with whom they live, including parents, roommates, and sexual partners because injecting drug could be viewed as socially unacceptable. Therefore, the traditional sampling and estimation methods can not be used to collection information on this population, so

called as hard-to-reach population. To overcome the difficulties, a network sampling technique, respondent-driven sampling (RDS), has been developed and widely used to sample from the hard-to-reach population [10-14]. The methodology shown in this article is to estimate the size of an IDU population and to understand the HIV/AIDS epidemics among IDUs, based on information from needle exchange programs (NEPs) by systematic surveillance.

2. Estimating the Size of IDU Population Who Use NEPs

Needle or syringe exchange programs, which sterile needles and syringes are free or at a minimal cost for IDUs, are a convenient means of monitoring the prevalence of blood borne viral infections among large numbers of IDUs who are currently injecting drugs. There are well over two hundred NEPs in Canada, with more under development [15]. In addition, there are numerous pharmacies that provide needle exchange services [15]. For example, within the province of Ontario, 34 NEPs operate distributing over 3.2 million clean syringes annually to an approximate 41,100 people who inject drugs and it is estimated that 53 needles are distributed per injector per year [16].

The purpose in this Section is to estimate the size of IDU population who are using NEPs, based on information of the number of needles distributed by NEP centres. To obtain such information from NEP centres, a two stage-samples procedure is provided as the following: a sample of NEP centres is selected, then, in a survey week, all individuals in the selected NEP centers will be interviewed, which allows for the sampling of individuals at a limited number of centres. Because of time and cost constraints, considering characteristics of IDU population, a sampling strategy has to be developed to select NEP centers. They can be chosen completely at random from a list of the all centers of NEPs. However, to decrease the variability of parameter estimators based on data from the completed survey, sampling theory suggests that it is better to sample large centers (covering large-IDU-population) with higher probability than small centers (covering small IDU-population). So, the sampling design for NEP centers is quite general and frequently used, which is the Stratified Probability-Proportional-to-Size sampling. Based on the number of needles which has exchanged in each NEP centre in the last year, all centers can be divided across the country into L strata with K_h needle exchange centers in the h th stratum. In addition, needles are legally available for purchase through pharmacies [17], although the willingness of pharmacists to sell syringes to IDUs is variable. This kind of exchange sites can be one of strata. From the h th stratum,

k_h NEP centers are sampled with inclusion probabilities π_{h1}, $\pi_{h2}, \cdots$, π_{h,k_h}. The inclusion probabilities are determined by the number of needles that were exchanged in the last year of the survey week. The π_{hi} could be determined in the following way: assuming that sampling is without replacement, the inclusion probability of the hi th NEP center in the list of k_h NEP centers is

$\pi_{hi} = k_h Q_{hi} / \sum_{j=1}^{K_h} Q_{hj}$, where Q_{hj} is the total needles that

were exchanged at the i th centre in h th strata in the last year of the survey week. It is necessary to require $\pi_{hi} < 1$. This is the case for all i, when $k_h = 1$. It could be expected that there is no very large Q_{hj} within each stratum. So, the requirement can be satisfied. Otherwise, it could be done by setting $\pi_{hi} = 1$ for the following inequality being valid [18]

$$k_h Q_{hi} / \sum_{j=1}^{K_h} Q_{hj} > 1 .$$

It is clear that the size of the IDU population which a NEP center covers is correlated with the number of needles distributed. Therefore, Probability-Proportional-Size sampling offers the possibility of decreased variability of estimators of totals [19]. It is also offer some operational advantages in multistage sampling in that it can equalize the workload across geographic areas sampled at the first stage of sampling. All IDUs attending the selected needle exchange centres during the survey week will be asked to complete a brief questionnaire, a question, such as "how many needles did you use in the last three months?" could be in the questionnaire.

Let T_{hi} be the number of needles that are exchanged during the period of the three months in the i th sampled NEP centre of h th stratum. Suppose there are N_{hi} IDUs who were exchanged the needles in the last three months of the survey week in this centres and N_{hi_j} is the number of IDUs who has used j needles during the three month and all those needles come from this centre. The numbers of N_{hi} and N_{hi_j} include not only IDUs who attend the NEP centre but also those IDUs who use the needles from the centre through possible second hand exchange. A survey [20] conducted in American shows that 90% of NEPs actively encouraged secondary exchange which is defined as "providing needles that you know will be used by persons other than the exchanger". Therefore, the number of clients for a NEP centre does not equal to the number of IDUs who exchange needles from this centre.

Now we can get relation between the number of IDUs who use needles from the centre and the total number of needles that are exchanged in the centre for the period of the three months,

$$T_{hi} = \sum_{j=1}^{m} j N_{hi_j} = N_{hi} \sum_{j=1}^{m} j R_{hi_j}$$

the m is the maximum number of needles that an IDU could be used in the three month, the $R_{hi_j} = N_{hi_j} / N_{hi}$ which can be estimated by the sample obtained in the survey week. Notice that N_{hi} and N_{hi_j} are unknown. However, what we need to know is the proportion R_{hi_j} and it can be estimated by the sample proportion $r_{hi_j} = n_{hi_j} / n_{hi}$, where the n_{hi} is the number of IDUs who attend the needle exchanges in the exchange centre and complete the questionnaire in the survey week and among of them there are n_{hi_j} IDUs who used j needles in the three months. Therefore, the N_{hi} can be estimated by

$$N_{hi} = T_{hi} \Big/ \sum_{j=1}^{m} j r_{hi_j} .$$

In fact, the denominator $r_{hi} = \sum_{j=1}^{m} j r_{hi_j}$ is the mean of numbers of needles that exchanged for IDUs who used the NEP for their changes in the period of three months. The estimator of the size N of IDU population who used NEPs at least once during the three months in all jurisdictions of the country (or the area) is the weighted mean

$$N = \sum_{h=1}^{L} \sum_{i=1}^{k_h} \pi_{hi}^{-1} N_{hi} .$$

Suppose the survey week is a typical one for the IDUs who use the NEP centre. It is assumed that the sample of IDUs who are observed in the survey week is obtained by a simple random sampling without replacement in the IDUs who use the NEP centre in the period of the three months. From the surveys which were carried out for studying the risk behavior of IDUs in Australian [21] and in Canada [22], the IDUs observed from a survey week are representative for the IDUs who use the NEP centers in a period of three months. So, the above assumption seems reasonable. Algorithms for Stratified Probability Proportion-to-Size Sampling selection without replacement could be realized, based on many methods, such as Hanurav-Vijayan algorithm [23,24].

3. Estimating the Proportion by RDS

Under the assumption in the last Section, the IDU population is made up of two groups of people based on their status of participating NEPs. To estimate the size of IDU population, we have to know not only the size of the population who use NEPs but also the proportion of the population who participate the NEPs during the three months. The idea from respondent-driven sampling [14] will be used to estimate the proportion.

The sampled IDUs in the survey week in the hi th centre can be divided into two groups: one is for having

used NEPs at least one time in the period of the three months and the other one is for having never used NEPs in the period, the groups are denoted by Ω_{NEP}^{hi} and $\Gamma_{\overline{NEP}}^{hi}$, respectively. The size of the group $\Gamma_{\overline{NEP}}^{hi}$ usually is small. To increase the size and coverage, we can use each IDU in the group of $\Gamma_{\overline{NEP}}^{hi}$ as an initial seed to conduct a respondent-driven sample, that is, each seed will recruit certain number of IDUs and provide information on how many friends they have with NEPs users group and non-users group respectively. The new sample collected by members of the group $\Gamma_{\overline{NEP}}^{hi}$ will contain NEP users and non-NEP users.

Now, to estimate the population proportion, we need to know the networking structure, such as the average degree of friendships, and probability that a non-NEP user (or NEP user) have a friendship with a NEP user. The total number of friendships radiating from the IDUs in the group Ω_{NEP}^{hi} is denoted by Φ_{NEP}^{hi} which can be written as

$$\Phi_{NEP}^{hi} = n_{NEP}^{hi} \Psi_{NEP}^{hi} ,$$

where Ψ_{NEP}^{hi} is the average friendships of an IDU in the group and n_{NEP}^{hi} is the number of IDUs in the group. Let $\Pi_{NEP,\overline{NEP}}^{hi}$ be the number of friendships that all IDUs in group Ω_{NEP}^{hi} have with IDUs who don't use NEPs. Then the probability $P_{NEP,\overline{NEP}}$ that an IDU who uses NEPs has a friendship with an IDU who has never used the NEPs in the period of the three months can be estimated by

$$P_{NEP,\overline{NEP}} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{k_h} \Pi_{NEP,\overline{NEP}}^{hi}}{\sum_{h=1}^{L} \sum_{i=1}^{k_h} \Phi_{NEP}^{hi}} .$$

The average number of friendships of IDUs in the group Ω_{NEP} can be estimated by pooling the samples from each sampled centre together:

$$\Psi_{NEP} = \sum_{d=1} d \frac{\sum_{h=1}^{L} \sum_{i=1}^{k_h} f_{NEP}^{hi}(d)}{\sum_{h=1}^{L} \sum_{i=1}^{k_h} n_{NEP}^{hi}} ,$$

the $f_{NEP}^{hi}(d)$ is the number of IDUs in the group Ω_{NEP}^{hi} who have number d friendships among IDUs. Similarly, we can get $P_{\overline{NEP},NEP}$ and $\Psi_{\overline{NEP}}$.

Based on the status of an IDU using NEPs in the period of the three months, we say that an IDU in state S_{NEP} if he or she has ever used NEPs in this period, otherwise we say he or she is in state $S_{\overline{NEP}}$. Suppose we have respondent-driven sampling design to collect a sample, the initial IDU (a seed) is chosen in step 0. Another IDU could be chosen based on the degree of friendships of the initial seed. We say this IDU is chosen in step 1, and so on. Suppose the chance of recruiting another IDU with state S_{NEP} depends on this chosen IDU

only through his or her degree of friendships among IDUs. Suppose also that if the IDU chosen in step 0 is in state S_{NEP}, then an IDU in state $S_{\overline{NEP}}$ will be chosen with probability $P_{NEP,\overline{NEP}}$; and if the IDU chosen in step 0 is not in state S_{NEP}, then an IDU in state S_{NEP} is chosen with probability $P_{\overline{NEP},NEP}$ in step 1. Letting X_n denote the status of an IDU chosen in the n th step, then $\{X_n, n = 0,1,\cdots\}$ is a two-state Markov chain having a following transition probability matrix:

$$\begin{pmatrix} P_{NEP,NEP} & P_{NEP,\overline{NEP}} \\ P_{\overline{NEP},NEP} & P_{\overline{NEP},\overline{NEP}} \end{pmatrix}.$$

It is clear that it is an irreducible argotic Markov chain [25]. In fact, we assume that IDUs within the group of using NEPs have similar proportions of degree of friendships with group of IDUs not using NEPs (or using NEPs). Also, notice that $P_{NEP,\overline{NEP}} = 1 - P_{NEP,NEP}$ and $P_{\overline{NEP},\overline{NEP}} = 1 - P_{\overline{NEP},NEP}$. In practice, it is much easier for a sample is collected by 1 wave. So, we have to use the different method to approach the proportions π_{NEP} and $\pi_{\overline{NEP}}$ of a chain of IDUs in states S_{NEP} and $S_{\overline{NEP}}$ respectively. Based on the Markov theory, we have

$$\pi_{NEP} = \frac{P_{\overline{NEP},NEP}}{1 + P_{\overline{NEP},NEP} - P_{NEP,NEP}}$$

and

$$\pi_{\overline{NEP}} = \frac{1 - P_{NEP,NEP}}{1 + P_{\overline{NEP},NEP} - P_{NEP,NEP}}.$$

The π_{NEP} and $\pi_{\overline{NEP}}$ can be estimated by plug-in estimates $P_{NEP,NEP}$ and $P_{\overline{NEP},NEP}$. Now, based on the result given by [14], the proportion α of IDUs who have ever used NEPs in the period of the three months can be estimated by

$$\alpha = \frac{\pi_{NEP}\Psi_{\overline{NEP}}}{\pi_{\overline{NEP}}\Psi_{NEP} + \pi_{NEP}\Psi_{\overline{NEP}}}.$$

Now we have estimated the size N of IDU population who use NEPs and the proportion α of IDUs in using NEPs among all IDUs, then, this information can be used to estimate the size of IDU population as N/α.

4. Discussions and Conclusions

For many years researchers have tried to get accurate size of IDU population. We have shown that this goal could be fulfilled by the sample design combining prevention programs—NEPs. Switching to this prevention programs and using the information of number of needles exchanged in the NEP centres give us a fresh and novel approach to the estimation of the size of IDU population. Using NEPs allows us to design a sampling and estima-

tion scheme which could be both cheaper and more accurate. The prevention programs, such as NEPs, provide a comprehensive HIV and blood-born infections prevention model to prevent the further spread of the diseases among IDUs and they have been proved to be effective for intervention of risk behaviour among IDUs. It is possible that the network of NEPs could embrace a well distributed, age and sex representative population of the area where we are interested in. However, it may be difficult to choose all NEPs to participate the data collection system. For the purpose of estimating the size of IDUs using NEPs, the sample design we proposed follows the basic principle of sampling theory which is that each individual in the target population should have some nonzero chance of being sampled in the survey [26]. Notice that the target population in our first stage of estimation is all IDUs who have ever used the NEPs in the period of the three months. Therefore, each IDU using NEPs has some nonzero chance of being sampled in the survey of estimating the size of IDUs of participating NEPs. We use the idea of respondent-driven sampling to estimate the proportion of IDU population in using NEPs. However, in order to estimate the proportion, what we need to know is transition probability that a non-NEP user (or NEP user) has a friendship with a NEP user. Then the results from Markov chain and the article [14] were used for the estimates. We have provided estimators corresponding to the sample design.

Our approach has concentrated on estimating the size of IDU population. It may be necessary to combine this sampling and estimation strategy with study of risk behaviour of IDUs. Further study for variances of the estimators has to be carried out. We have presented a number of analytic results and these analytic arguments could be further supported with numerical simulation. Also, the possible bias exists when we estimate the transition probabilities, because the sample size of non-NEP users is usually small. However, it may be possible to predict the magnitudes and direction of this bias ahead of time.

5. Acknowledgements

REFERENCES

[1] Joint United Nations Programme on HIV/AIDS (UN-

AIDS), "Global Report: UNAIDS Report on the Global AIDS Epidemic 2010," Geneva, 2010.

[2] B. M. Mathers, L. Degenhardt, B. Phillips, L. Wiessing, M. Hickman, S. Strathdee, A. Wodak, *et al.*, "Global Epidemiology of Injecting Drug Use and HIV among People who Inject Drugs: A Systematic Review," *Lancet*, Vol. 372, No, 9651, 2008, pp. 1733-1745. doi:10.1016/S0140-6736(08)61311-2

[3] Centers for Diseases Control (CDC), "HIV Prevention Strategic Plan through 2005," Atlanta, 2001.

[4] C. P. Archibald, G. C. Jayaraman, C. Major, D. M. Patrick, S. M. Houston and D. Sutherland, "Estimating the Size of Hand-to-Reach Population: A Novel Method Using HIV Testing Data Compared to Other Methods," *AIDS*, Vol. 15, 2001, pp. s41-s48. doi:10.1097/00002030-200104003-00006

[5] J. Weekes and K. Cumberland, "Needle Exchange Programs (NEPs) FAQs," 2004. http://www.ccsa.ca/2004%20CCSA%20Documents/ccsa-010055-2004.pdf.Accessed on 11 May 2011

[6] Canadian Centre on Substance Abuse, "Injection Drug Users Overview," 2009, accessed on 11 May 2011. http://www.ccsa.ca/Eng/Topics/Populations/IDU

[7] K. L. Dehne, M. Adelekan, A. Chatterjee and G. Weiler, "The Need for a Global Understanding of Epidemiological Data to Inform Human Immunodeficiency Virus (HIV) Prevention among Injecting Drug Users," *Bulletin on Narcotics*, Vol. 44, No. 1-2, 2002, pp. 117-130.

[8] R. Muga, K. Langohr, J. Tor, A. Sanvisens, I. Serra and C. R. Rey-Joly, "Survival of HIV-Infected Injection Drug Users (IDUs) in the Highly Active Antiretroviral Therapy Era, Relative to Sex- and Age-Specific Survival of HIV-Uninfected IDUs," *Clinical Infectious Diseases,* Vol. 45, No. 3, 2007, pp. 370-376. doi:10.1086/519385

[9] Committee on the Prevention of HIV Infection among Injecting Drug Users in High-Risk Countries, "Preventing HIV Infection among Injecting Drug Users in High Risk Countries: An Assessment of the Evidence," The National Academies Press, Washington, D.C., 2006.

[10] D. D. Heckathorn, "Respondent-Driven Sampling: A new Approach to the Study of Hidden Populations," *Social Problems*, Vol. 44, No. 2, 1997, pp. 174-199. doi:10.1525/sp.1997.44.2.03x0221m

[11] S. Zhang, P. Yan and C. Archibald, "Estimating the Size of IDU Population Using Needle Exchange Programs," *Proceedings of Statistics Canada Symposium* 2004 *Innovative Methods for Surveying Difficult-to-Reach Populations*, Statistics Canada, Ottawa, 2004, pp. 1-10.

[12] D. D. Heckathorn, "Extensions of Respondent-Driven sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment," *Sociological Methodology*, Vol. 31, No. 1, 2007, pp. 151-207. doi:10.1111/j.1467-9531.2007.00188.x

[13] K. J. Gile and M. S. Handcock, "Respondent-Driven Sampling: An Assessment of Current Methodology," *Sociological Methodology*, Vol. 40, No. 1, 2010, pp. 285-327. doi:10.1111/j.1467-9531.2010.01223.x

[14] M. J. Salganik and D. D. Heckathorn, "Sampling and Estimation in hidden Populations Using Respondent-driven Sampling," *Sociological Methodology*, Vol. 34, No. 1, 2004, pp. 193-239. doi:10.1111/j.0081-1750.2004.00152.x

[15] Public Health Agency of Canada, "Harm Reduction and Injection Drug Use: An International Comparative Study of Contextual Factors Influencing the Development and Implementation of Relevant Policies and Program," 2003, accessed on 18 March 2006. http://www.phac-aspc.gc.ca/hepc/pubs/hridu-rmudi/index-eng.php

[16] C. Strike, L. Leonard, M. Millson, S. Anstice, N. Berkeley and E. Medd, "Ontario Needle Exchange Programs: Best Practice Recommendations," Ontario Needle Exchange Coordinating Committee, Toronto, 2006.

[17] C. L. Miller, M. Tyndall, P. Spittal, K. Li, A. Palepu and M. T. Schechter, "Risk-Taking Behaviors among Injecting Drug Users Who Obtain Syringes from Pharmacies, Fixed Sites, and Mobile Van Needle Exchanges," *Journal of Urban Health*, Vol. 79, No. 2, 2002, pp. 257-265. doi:10.1093/jurban/79.2.257

[18] C. Sarndal, B. Swensson and J. Wretman, "Model Assisted Survey Sampling," *Series in Statistics*, Springer, New York, 1992.

[19] W. G. Cochran, "Sampling Techniques," 3rd Edition, Wiley, New York, 1977.

[20] D. C. DesJarlais, C. McKnight, K. Eigo and P. Friedmann, "2000 United States Syringe Exchange Program Survey," The Baron Edmond de Rothschild Chemical Dependency Institute, New York, 2000.

[21] M. MacDonald, A. D. Wodak, R. Ali, N. Crofts, P. H. Cunningham, K. A. Dolan, M. Kelaher, W. M. Loxley, I. van Beek and J. M. Kaldor, "HIV Prevalence and Risk Behaviour in Needle Exchange Attenders: A National Study. The Collaboration of Australian Needle Exchanges," *The Medecal Journal of Australia*, Vol. 166, No. 5, 1997, pp. 237-240.

[22] M. Peggy, T. Myers, L. Calzavara, E. Wallace, C. Major, and N. Degani, "Regional Variation in HIV Prevalence and Risk Behaviours in Ontario Injection Drug Users," *Canadian Journal of Public Health*, Vol. 94, No. 6, 2003, pp. 431-435.

[23] K. Vijayan, "An Exact Sampling Scheme: Generalization of a Method of Hanurav," *Journal of the Royal Statistical Society*, Vol. 30, No. 6, 1968, pp. 499-513.

[24] K. W. R. Brewer, "A model of Systematic Sampling with Unequal Probabilities," *Australian Journal of Statistics*, Vol. 5, 1963, pp. 93-105. doi:10.1111/j.1467-842X.1963.tb00288.x

[25] S. M. Roos, "Probability Models," 6th Edition, Academic Press, London, 1997.

[26] E. L. Korn and B. I. Graubard, "Analysis of Health Surveys," *Wiley Series in Probability and Statistics Survey Methodology Section*, John Wiley & Sons, New York, 1999. doi:10.1002/9781118032619

[27] M. E. Thompson, "Theory of Sample Surveys," Chapman

& Hall, London, 1997.

[28] S. L. Lohr, "Sampling: Design and Analysis," Duxbury Press, New York, 1999.

[29] C. Sarndal, B. Swensson and J. Wretman, "Model Assisted Survey Sampling," *Series in Statistics*, Springer, New York, 1992.

Appendix

We discuss the variance of the estimator N in this appendix. First, we look at the estimator N_{hi}. Let Y_{hi} denote the number of needles exchanged by an injecting drug user in the centre for a period of the three months. The probability mass function of $Y_{hi} = j$ is denoted by $f_{hi}(y)$. The mean of Y_{hi} is $\overline{Y}_{hi} = \sum_j j f_{hi}(j)$. The variance of Y_{hi} conditioned on N_{hi} is

$$\frac{N_{hi}}{N_{hi}-1}\left[\sum_{j=1}^{N_{hi}} j^2 f_{hi}(j) - \left(\sum_{j=1}^{N_{hi}} j f_{hi}(j)\right)^2\right].$$

The $f_{hi}(j)$ can be substituted by r_{hi_j}; the $\dfrac{N_{hi}}{N_{hi}-1}$ can be replaced by 1 because N_{hi} is usually very large in our case. Notice that T_{hi} is known and $\overline{Y}_{hi}$ is estimated by r_{hi}, linearization methods [27,28] are used to approximate the variance of the estimator N_{hi}, then, we have

$$\mathrm{var}\left(N_{hi}\right) \approx \frac{1}{\overline{Y}_{hi}^2} N_{hi}^2 \, \mathrm{var}\left(\overline{Y}_{hi}\right),$$

where E denote the expectation. Finally, the variance N_{hi} can be estimated by

$$\frac{1}{\overline{Y}_{hi}^2} N_{hi}^2 \left(1-\frac{n_{hi}}{N_{hi}}\right)\frac{\left(\mathrm{var}\left(Y_{hi}\right)\right)}{n_{hi}},$$

where $\mathrm{var}\left(Y_{hi}\right)$ can be approximated by

$$\frac{N_{hi}}{N_{hi}-1}\left[\sum_{j=1}^{N_{hi}} j^2 r_{hi_j} - \left(\sum_{j=1}^{N_{hi}} j r_{hi_j}\right)^2\right].$$

Because the sampling is carried in each stratum independently and within a stratum the Probability Proportional-to-size sampling with fixed size without-replacement design is used, the variance of the estimator N for the size of IDU who use the NEP in the period of the three months can be written as the following [29]:

$$\mathrm{var}\left(N \mid N_{hi}, i = 1, \cdots, K_h\right)$$
$$= \frac{1}{2}\sum_{h=1}^{L}\sum_{i=1}^{K_h}\sum_{l=1}^{K_h}\left(\pi_{hi}\pi_{hl}-\pi_{hil}\right)\left(\frac{N_{hi}}{\pi_{hi}}-\frac{N_{hl}}{\pi_{hl}}\right)^2$$

where π_{hil} is the covariance between indicator variables

I_{hi} and I_{hl} ($I_{hi} = 1$ if the centre hi is included in the sample, 0 otherwise). In the case of fixed size without replacement design, the calculation of π_{hil} could be complicated. For example, if the sample size in hth stratum is 2, the π_{hil} has the following expression under the scheme [29]:

$$\pi_{hil} = \frac{2 Q_{hi} Q_{hl}\left(Q_h - Q_{hi} - Q_{hl}\right)}{Q_h\left(\sum_{j=1}^{K_h} c_{hj}\right)\left(Q_h - 2 Q_{hi}\right)\left(Q_h - 2 Q_{hl}\right)},$$

where

$$Q_h = \sum_{j=1}^{K_h} Q_{hj}$$

and

$$c_{hj} = Q_{hj}\left(Q_h - Q_{hj}\right) \Big/ \left(Q_h\left(Q_h - 2 Q_{hj}\right)\right).$$

For the first draw, the scheme gives the center hi the probability $p_{hi} = c_{hi}\Big/\sum_{j=1}^{K_h} c_{hj}$ of being selected; without replacing the first drawn element, (say h, i_0), it give the other element the probability:

$$p_{hi \mid h, i_0} = Q_{hi}\Big/\left(Q_h - Q_{h, i_0}\right).$$

The variance $\mathrm{var}(N)$ can be calculated by

$$\mathrm{var}(N) = E\left(\mathrm{var}\left(N \mid N_{hi}, i = 1, \cdots, k_h\right)\right)$$
$$+ \mathrm{var}\left(E\left(N \mid N_{hi}, i = 1, \cdots, k_h\right)\right).$$

Considering that all N_{hi} are estimated independently, we could have

$$\mathrm{var}\left(E\left(N \mid N_{hi}, i = 1, \cdots, k_h\right)\right)$$
$$\approx \sum_{h=1}^{L} \frac{K_h}{k_h}\sum_{i=1}^{k_h} \mathrm{var}\left(N_{hi}\right).$$

By the Taylor linearization technique, an approximation of the variance of the estimator N for the size of injecting drug users who use NEP in the period of the three months is

$$\mathrm{var}(N) = \sum_{h=1}^{L}\left[\frac{1}{2}\sum_{i=1}^{k_h}\sum_{l=1}^{k_h}\left(\pi_{hi}\pi_{hl}/\pi_{hil}-1\right)\left(\frac{N_{hi}}{\pi_{hi}}-\frac{N_{hl}}{\pi_{hl}}\right)^2 \right.$$
$$\left. + \frac{K_h}{k_h}\sum_{i=1}^{k_h} \mathrm{var}\left(N_{hi}\right)\right].$$