

# Estimation of the Frequency of Intravenous Drug Users in Hamadan City, Iran, Using the Capture-recapture Method

Salman Khazaei<sup>1</sup>, Jalal Poorolajal<sup>2</sup>, Hossein Mahjub<sup>2</sup>, Nader Esmailnasab<sup>3</sup>, Mohammad Mirzaei<sup>4</sup>

<sup>1</sup>Department of Epidemiology & Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan; <sup>2</sup>Research Center for Health Sciences, Department of Epidemiology & Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Hamadan; <sup>3</sup>Department of Community Medicine, School of Medicine, Kurdistan University of Medical Sciences, Kurdistan; <sup>4</sup>Center for Disease Control & Prevention, Vice-chancellor of Health Services, Hamadan University of Medical Sciences, Hamadan, Iran

**OBJECTIVES:** The number of illicit drug users is prone to underestimation. This study aimed to use the capture-recapture method as a statistical procedure for measuring the prevalence of intravenous drug users (IDUs) by estimating the number of unknown IDUs not registered by any of the registry centers.

**METHODS:** This study was conducted in Hamadan City, the west of Iran, in 2012. Three incomplete data sources of IDUs, with partial overlapping data, were assessed including: (a) Volunteer Counseling and Testing Centers (VCTCs); (b) Drop in Centers (DICs); and (c) Outreach Teams (ORTs). A log-linear model was applied for the analysis of three-sample capture-recapture results. Two information criteria were used for model selection including Akaike's Information Criterion and the Bayesian Information Criterion.

**RESULTS:** Out of 1,478 IDUs registered by three centers, 48% were identified by VCTCs, 32% by DICs, and 20% by ORTs. After exclusion of duplicates, 1,369 IDUs remained. According to our findings, there were 9,964 (95% CI, 6,088 to 17,636) IDUs not identified by any of the centers. Hence, the real number of IDUs is expected to be 11,333. Based on these findings, the overall completeness of the three data sources was around 12% (95% CI, 7% to 18%).

**CONCLUSION:** There was a considerable number of IDUs not identified by any of the centers. Although the capture-recapture method is a useful and practical approach for estimating unknown populations, due to the assumptions and limitations of the method, the results must be interpreted with caution.

**KEY WORDS:** Capture-recapture, Log-linear models, Poisson regression, Intravenous drug users, Iran

## INTRODUCTION

Illicit drug use is a major public health problem worldwide [1]. It is estimated that about 5% of the world's adult population,

or 230 million people, have used an illicit drug at least once in 2010. The number of drug users worldwide is estimated to be about 27 million, which is 0.6% of the world's adult population [2]. About 15.9 million people worldwide may inject drugs, of whom about three million may be human immunodeficiency virus (HIV) positive [3]. Injecting drug use is the main route of transmission for about 10% of HIV infections in the world and 30% of infections outside of sub Saharan Africa. Preventing HIV transmission through injecting drug use is one of the key challenges in reducing the burden of HIV worldwide [1].

It is estimated that between 1.2 and 2 million people are drug abusers in Iran of about 20% to 25% have injected at least once in their lifetime. Approximately 1% of the Iranian population use heroin, of which 40% inject it. Therefore, it is estimated that there are around 280 thousand injecting drug abusers in the country [4].

Correspondence: Jalal Poorolajal, MD, PhD  
Research Center for Health Sciences, Department of Epidemiology & Biostatistics, School of Public Health, Hamadan University of Medical Sciences, Shahid Fahmideh Ave., Hamadan 6517838695, Iran  
Tel: +98-811-838-0090, Fax: +98-811-838-0509,  
E-mail: poorolajal@umsha.ac.ir

Received: Jul 9, 2012, Accepted: Sep 19, 2012, Published: Oct 31, 2012

This article is available from: <http://e-epih.org/>.

© 2012, Korean Society of Epidemiology

© This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Hamadan City, the center of Hamadan Province, stands on the north west of Iran and has a population of approximately 600 thousand people [5]. According to the data recorded in the District Health Network database, 1,369 intravenous drug users (IDUs) were identified in the city in 2011, which seemed much less than that expected according to the national estimates.

Illicit drug use is an illegal and stigmatized behavior and most drug users are reluctant to be identified as such. Hence, the number of illicit drug users is prone to underestimation [3]. One method which is commonly used to estimate hard-to-reach populations is the so-called capture-recapture approach. This method was first used in ecology to estimate the unknown size of wild animal populations [6]. It was subsequently used as a potentially useful method in epidemiology for estimating the completeness of ascertainment of disease registers [7,8] and incomplete reporting [9,10]. Actually, the capture-recapture method can principally be applied to any situation where there are two or more incomplete lists.

A high prevalence of HIV [4] and hepatitis C virus (HCV) [11] infections among IDUs emphasizes the importance of quantifying the dynamics of this population. Estimating the size of IDU populations is fundamental for predicting the future burden of HIV and HCV diseases and for planning and providing health care services for this group of individuals [12]. Furthermore, statistically valid and reasonable estimates of IDU prevalence can help policymakers to estimate probable outbreaks of infections associated with drug injection [13]. Populations with negative traits are usually underestimated. Several statistical methods have been suggested for the estimation of such unknown populations [14-16]. A straightforward and easy to use method is the capture-recapture approach. In this study, we used this method as a statistical procedure and a practical means for measuring IDU prevalence by estimating the number of unknown IDUs not registered by any of the three incomplete registries.

## MATERIALS AND METHODS

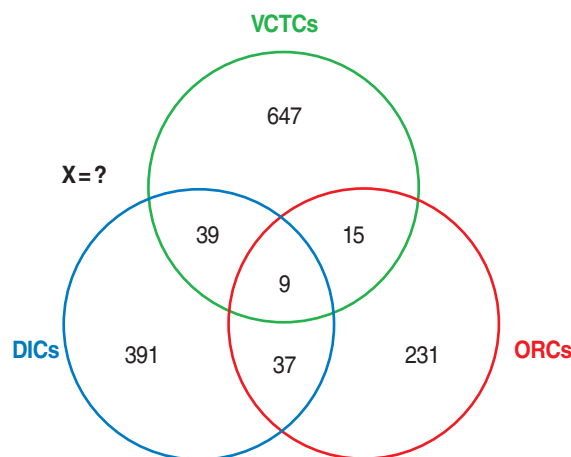
This study was approved by the Human Subject Review Board of Hamadan University of Medical Sciences and conducted in Hamadan City, the center of Hamadan Province in the west of Iran, in 2012. We extracted the data on IDUs registered in at least one the three centers that provide health services to IDUs in this city, from April 2011 to March 2012. IDUs were defined as people who inject narcotic substances into the body with a hollow needle and a syringe which is pierced through the skin into the body usually intravenously, but also intramuscularly or subcutaneously [17].

There are three centers for providing health care services to

IDUs in Iran: (a) Volunteer Counseling and Testing Centers (VCTCs); (b) Drop in Centers (DICs); and (c) Outreach Teams (ORTs). VCTCs provide consultant and educational services to IDUs in order to improve their knowledge of high-risk behaviors and harm reduction methods. In addition, these centers provide diagnostic tests for IDUs and refer the suspected IDUs to specialized medical centers for diagnostic and treatment services. DICs provide social support such as serving food, clothes, and bathing for IDUs. These centers also give agonist and maintenance medications. They also plan educational programs for IDUs in order to improve their knowledge of high-risk behaviors and healthy sexual behaviors. ORTs include a well-trained team of health workers who actively look for the IDUs not referred to VCTCs or DICs. They distribute safety boxes among IDUs including sterile syringes, disinfecting material and condoms. They also encourage IDUs to adopt safe and low-risk behaviors. The IDUs may be identified and registered by any of these centers. Some IDUs may be identified by more than one center. Nonetheless, none of these centers has a complete list of IDUs. We used the capture-recapture method in order to statistically estimate the approximate number of IDUs not identified by these registries.

For this purpose, the list of IDUs recorded in these three data sources were extracted and compared with each other, in order to identify common names listed in more than one database. Since the IDUs' national identification codes were not recorded in the data sources, we used several demographic characteristics of the IDUs for comparison, including first name, second name, age, marital status, and region of residence. Then we arranged the data as shown in Figure 1.

The simplest capture-recapture model is the so-called two-sample model [6]. Another capture-recapture method, which



**Figure 1.** Distribution of intravenous drug users registered by three centers: Volunteer Counseling and Testing Centers (VCTCs); Outreach Teams (ORTs); and Drop in Centers (DICs).

was used in this study, is the so-called three-sample model, in which the three incomplete data sources of IDUs were used. In this approach, the capture-recapture method becomes more complicated and includes the following eight possible models as follows (Figure 1):

1. Number of IDUs identified by VCTCs only (A)
2. Number of IDUs identified by DICs only (B)
3. Number of IDUs identified by ORTs only (C)
4. Number of IDUs identified by A and B but not by C (AB)
5. Number of IDUs identified by A and C but not by B (AC)
6. Number of IDUs identified by B and C but not by A (BC)
7. Number of IDUs identified by all three centers (ABC)
8. Number of IDUs studies identified by none of the three centers (X)

There are many elaborate statistical models available for the analysis of three-sample capture-recapture results. The log-linear model, or Poisson regression, is a simple model which easily accommodates the three sources of data and is able to explore dependence between sources and adjust for dependence by including interaction terms in the model [18]. For this purpose, we prepared a dataset with four variables as follows: (a) “var\_A”, with values 0 or 1, that described belonging to list A; (b) “var\_B”, with values 0 or 1, that described belonging to list B; (c) “var\_C”, with values 0 or 1, that described belonging to list C; and (d) “var\_freq”, which was a non-negative variable that described the frequency of observations in the combination of lists given by variables A, B and C. The unknown frequency of cases occurring in none of the lists was considered as missing. This missing value was estimated by the Poisson model.

The interaction terms were used to model dependence. The basic assumption for the capture-recapture model is that there is no three-list interaction term for three lists. In other words, the third-order interaction vanishes ( $ABC=0$ ) [18]. By accommodating the three sources of data as shown above, the log-linear model can estimate the number of IDUs not identified by any of the three centers (X) and hence the total population of IDUs (N).

There are two different information criteria which were proposed for model selection: Akaike’s Information Criterion (AIC) and the Bayesian Information Criterion (BIC) [19]. The AIC is calculated as follows:

$$AIC = G^2 - [2 \times (df)]$$

Where  $G^2$  is the likelihood ratio statistic associated with the fit of any model to the data and  $df$  denotes the degrees of freedom of the model. The model giving the smallest value of AIC is the one selected. The second criterion, BIC, which is usually preferred to AIC in some applications, is calculated as follows:

$$BIC = G^2 - [\ln(Nobs/2\pi)] \times (df)$$

Where  $G^2$  and  $df$  are defined as above, and where  $\ln(Nobs)$  is natural logarithm of the number of parameters in the model.

As above, the model giving the smallest value of BIC is the one selected.

All analyses were performed at the 0.05 significance level using statistical software Stata version 11.0 (StataCorp, College Station, TX, USA).

## RESULTS

The general characteristics of the three data sources are shown in Table 1. The majority of IDUs were single men aged 30 to 44 years. Out of 1,478 IDUs identified by three centers, 710 (48%) were identified by VCTCs, 476 (32%) by DICs, and 292 (20%) by ORTs. In addition, 100 (7%) of IDUs were identified by at least two centers and 9 (1%) by all three centers (Figure 1). After exclusion of duplicates, 1,369 IDUs remained.

The results of the log-linear analysis are shown in Table 2. The p-values show that there were significant differences between the saturated model (the 8th model) and all other reduced models. The sixth model (ABC AB BC) was the best fitting model with the smallest value of both AIC and BIC. According to these findings, there might be 9,964 (95% confidence interval [CI], 6,088, 17,636) IDUs who were not identified by any of the centers. Hence, the real number of IDUs is expected to be 11,333 (1,369 + 9,964). Accordingly, the combined completeness of the three data sources was almost 12% (95% CI, 7% to 18%). Among the three data sources, VCTCs were more complete than

**Table 1.** Distribution of the demographic characteristics of the intravenous drug users registered by three centers including Volunteer Counseling and Testing Centers (VCTCs); Outreach Teams (ORTs); and Drop in Centers (DICs)

Characteristics	VCTCs	DICs	ORTs
Gender			
Male	709 (99.86)	367 (77.10)	238 (81.51)
Female	1 (0.14)	109 (22.90)	54 (18.49)
Age (yr)			
<30	242 (34.08)	202 (42.44)	211 (72.26)
30-44	380 (53.52)	222 (46.64)	70 (23.97)
45-59	85 (11.97)	44 (9.24)	11 (3.77)
≥60	3 (0.42)	8 (1.68)	0 (0.00)
Marital status			
Married	246 (34.65)	134 (28.15)	44 (15.07)
Single	395 (55.63)	254 (53.36)	208 (71.23)
Divorced	68 (9.58)	64 (13.45)	30 (10.27)
Widow	1 (0.14)	24 (5.04)	10 (3.42)
Region			
Urban	701 (98.73)	470 (98.74)	288 (98.63)
Rural	9 (1.27)	6 (1.26)	4 (1.37)
Total	710 (100.00)	476 (100.00)	292 (100.00)

Values are presented number (%).

**Table 2.** Log-linear models fitted to three lists of intravenous drug users registered by three centers and estimated number of intravenous drug users which was not registered by the centers

Model	K	df	Log likelihood	G <sup>2</sup>	AIC	BIC	p-value	X	95% CI
1. A B C	3	3	-39.6971	36.43	30.43	30.59	0.001	4,761	3,852, 5,931
2. A B C AB	4	2	-38.5777	34.19	30.19	30.30	0.001	4,078	3,071, 5,491
3. A B C AC	4	2	-37.1667	31.37	27.37	27.47	0.001	4,108	3,229, 5,278
4. A B C BC	4	2	-25.2870	7.61	3.61	3.71	0.023	6,768	5,202, 8,948
5. A B C AB AC	5	1	-31.5662	20.16	18.16	18.22	0.001	2,441	1,717, 3,558
6. A B C AB BC	5	1	-23.5780	4.19	2.19	2.24	0.014	9,964	6,088, 17,636
7. A B C AC BC	5	1	-25.2076	7.45	5.45	5.50	0.006	6,487	4,682, 9,235
8. A B C AB AC BC	6	0	-21.4841	0.00	0.00	0.00	1.000	24,298	9,082, 62,706

K, number of parameters; df, degree of freedom; G<sup>2</sup>, deviance; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; X, unregistered; CI, confidence interval.

**Table 3.** Comparison of the completeness of detecting intravenous drug users by the three centers including Volunteer Counseling and Testing Centers (VCTCs); Outreach Teams (ORTs); and Drop in Centers (DICs)

Data source	n	Completeness	95% CI
VCTCs	710	0.07	0.04, 0.12
ORTs	476	0.05	0.03, 0.08
DICs	292	0.03	0.02, 0.05

the two other centers.

Table 3 shows the completeness of detecting IDUs by the three centers. The completeness of the VCTCs was the highest and that of DICs was the lowest. However, the completeness values for all centers were very low.

## DISCUSSION

Estimation of the size of IDU populations is essential for planning and providing health care services to this group of individuals [12]. In addition, a valid and reliable estimate of the IDU population is needed for estimating the probable outbreaks of blood born infections associated with drug injection such as HIV and HCV [13]. Our findings revealed that the majority of IDUs were not identified by any of the centers.

The p-values showed that there was a significant difference between the reduced models and the saturated one. We used AIC and BIC criteria in order to choose the best fitting model among the models, all of which were statistically significantly different to the saturated model. The sixth model with the smallest value of AIC or BIC was the one selected as the best fitting model. However, care must be taken when using AIC and BIC values for model selection. AIC and BIC do not provide a test of a model in the sense of testing a null hypothesis; i.e. AIC and BIC can tell nothing about how well a model fits the data in an absolute sense. If all candidate models fit the data poorly, these values will not give any warning of this [20].

One might try two-sample methods (AB, AC, and BC) and compare the results with the three-sample method (ABC). However, it is important to remember that the capture-recapture method requires a sufficiently high level of overlapping information to produce a reliable estimate of the unknown population [18]. Using the three-sample approach increases the overlapping area and thus produces more reliable estimates than that of the two-sample approach. Therefore, if possible, the three-sample approach is preferred.

The completeness of the three lists was low and different. The proportion of overlapping cases was relatively low; however, sufficiently high overlapping information is required to produce a reliable estimate of the unknown population [18]. This might result in overestimation of the pooled estimates. This may explain the low completeness of the data sources. Furthermore, the VCTCs, DICs and ORTs provide different services and IDUs may refer to these centers for different purposes. This might explain the difference observed between the completeness of the three centers.

Capture-recapture is a simple statistical approach and represents an attractive method to assess the completeness of different data sources in order to estimate the size of the unknown population, which could not be identified from the data sources. Although the capture-recapture method is a useful and practical approach to estimate the hard-to-reach populations, due to the assumptions and limitations of the method, the results must be interpreted with caution.

A number of studies used the three-sample capture-recapture method to estimate the population of IDUs in other countries. These studies revealed that registries underestimated the real size of the population of IDUs. Davies et al. [21] estimated the number of injecting drug users in the city of Edinburgh, Scotland, in 1999, using the capture-recapture method. The number of IDUs was estimated to be 1,770 from the registry data, while the capture-recapture method estimated the number of IDUs to be 2,070. Accordingly, the completeness of the registry data was 86%. Hickman et al. [22] assessed IDU prevalence in Bris-

tol in 2004. They reported that the total number of IDUs recorded by the three registries was 2,986, whereas the overall estimated number of IDUs using the capture-recapture method was 5,540. Based on these results, the overall completeness of the registries was 54%. Böhning et al. [23] and Goli et al. [16] used a different statistical method, the so-called mixture Poisson model, to estimate the prevalence of heroin consumption in Bangkok and cigarette smokers in Hamadan in the west of Iran, respectively.

Although the capture-recapture approach is a potentially useful method for estimating the size of an unknown population, this method, like any other statistical procedures, has its own limitations. A critical limitation of this method is that sufficiently high overlapping information is required to produce a reliable estimate of the unknown population size. Otherwise, the likelihood functions may become flat and the resulting estimates based on log-linear models may possibly become unstable [18]. Another limitation of the capture-recapture method using log-linear models for investigating an unknown population is that a relative large number of data sources is required to hold the assumption of the normal distribution within log-linear models. Moreover, the validity of capture-recapture results depends on some assumptions. If these assumptions are not considered, the estimates may not be reliable. A critical assumption of capture-recapture methods is the independence of the data sources, so that either positively or negatively dependent sources may cause either underestimation or overestimation of the pooled estimates, respectively [6]. Of course, log-linear models are able to handle dependence among data sources and adjust for this dependence by including interaction terms in the model [18].

Despite its limitations, the current study indicated that the estimated population of IDUs based on registries, and hence the prevalence estimates of IDUs, are considerably underestimated relative to the actual values. Therefore, the policymakers who plan and provide health care services for IDUs should take this issue into special consideration.

The results of this study indicated that the capture-recapture method is an easy to conduct and straightforward approach for the estimation of the size of the hard-to-reach population of IDUs. According to these findings, the majority of the IDUs were not identified by the registry centers. Although the capture-recapture method is a useful and practical approach to estimate unknown populations, due to the assumptions and limitations of the method, the results must be interpreted with caution.

## ACKNOWLEDGMENTS

We would like to thank the Vice-chancellor of Education as

well as the Vice-chancellor of Research and Technology, Hamadan University of Medical Sciences, who approved this study.

## CONFLICT OF INTEREST

The authors have no conflict of interest to declare for this study.

## REFERENCES

1. Joint United Nations Programme on HIV/AIDS. Global AIDS response progress reporting 2012 [cited 2012 Sep 13]. Available from: [http://www.unaids.org/en/media/unaids/contentassets/documents/document/2011/JC2215\\_Global\\_AIDS\\_Response\\_Progress\\_Reporting\\_en.pdf](http://www.unaids.org/en/media/unaids/contentassets/documents/document/2011/JC2215_Global_AIDS_Response_Progress_Reporting_en.pdf).
2. United Nations Office on Drugs and Crime. World drug report 2012 [cited 2012 Sep 13]. Available from [http://www.unodc.org/documents/data-and-analysis/WDR2012/WDR\\_2012\\_web\\_small.pdf](http://www.unodc.org/documents/data-and-analysis/WDR2012/WDR_2012_web_small.pdf).
3. Mathers BM, Degenhardt L, Phillips B, Wiessing L, Hickman M, Strathdee SA, et al. Global epidemiology of injecting drug use and HIV among people who inject drugs: a systematic review. *Lancet* 2008;372:1733-1745.
4. World Health Organization. Best practice in HIV/AIDS prevention and care for injecting drug abusers: the triangular clinic in Kermanshah, Islamic Republic of Iran. Cairo: Regional Office for the Eastern Mediterranean; 2004, p. 12.
5. Statistical Center of Iran. Population estimation. Tehran: Presidency of I.R.I; 2012 [cited 2012 Sep 13]. Available from: [http://www.amar.org.ir/Portals/0/Files/jamiat\\_hamedan.pdf](http://www.amar.org.ir/Portals/0/Files/jamiat_hamedan.pdf) (Iranian).
6. Capture-recapture Web-page. Capture-recapture web-pae. [cited 2012 Jun 30]. Available from: <http://www.pitt.edu/~yuc2/cr/main.htm>.
7. Amin-Esmaili M, Nedjat S, Motevalian A, Rahimi-Movaghar A, Majdzadeh R. Comparison of databases for Iranian articles; access to evidence on substance abuse and addiction. *Arch Iran Med* 2009;12:559-565.
8. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;17:243-264.
9. Bennett DA, Latham NK, Stretton C, Anderson CS. Capture-recapture is a potentially useful method for assessing publication bias. *J Clin Epidemiol* 2004;57:349-357.
10. Poorolajal J, Haghdoost AA, Mahmoodi M, Majdzadeh R, Nasserimoghaddam S, Fotouhi A. Capture-recapture method for assessing publication bias. *J Res Med Sci* 2010;15:107-115.
11. Hagan H. Hepatitis C virus transmission dynamics in injection drug users. *Subst Use Misuse* 1998;33:1197-1212.
12. Des Jarlais DC, Dehne K, Casabona J. HIV surveillance among injecting drug users. *AIDS* 2001;15 Suppl 3:S13-S22.
13. Brady JE, Friedman SR, Cooper HL, Flom PL, Tempalski B, Gostnell K. Estimating the prevalence of injection drug users in the U.S. and in large U.S. metropolitan areas from 1992 to 2002. *J Urban Health* 2008;85:323-351.
14. Hall WD, Ross JE, Lynskey MT, Law MG, Degenhardt LJ. How many dependent heroin users are there in Australia? *Med J Aust* 2000;173:528-531.
15. van der Heijden PGM, Bustami R, Cruyff MJLF, Engbersen G, van Houwelingen HC. Point and interval estimation of the population

- size using the truncated Poisson regression model. *Stat Modelling* 2003;3:305-322.
16. Goli S, Mahjub H, Moghimbeigi A, Poorolajal J, Heidari Pahlavian A. Application of mixture models for estimating the prevalence of cigarette smoking in hamadan, iran. *J Res Health Sci* 2010;10:110-115.
  17. Wikipedia contributors. Drug injection [cited 2012 Sep 13]. Available from: [http://en.wikipedia.org/wiki/Drug\\_injection](http://en.wikipedia.org/wiki/Drug_injection).
  18. Chao A, Tsay PK, Lin SH, Shau WY, Chao DY. The applications of capture-recapture models to epidemiological data. *Stat Med* 2001;20:3123-3357.
  19. Hook EB, Regal RR. Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. *Am J Epidemiol* 1997;145:1138-1144.
  20. Wikipedia contributors. Akaike information criterion [cited 2012 Sep 13]. Available from: [http://en.wikipedia.org/wiki/Akaike\\_information\\_criterion](http://en.wikipedia.org/wiki/Akaike_information_criterion).
  21. Davies AG, Cormack RM, Richardson AM. Estimation of injecting drug users in the City of Edinburgh, Scotland, and number infected with human immunodeficiency virus. *Int J Epidemiol* 1999;28:117-121.
  22. Hickman M, Hope V, Coleman B, Parry J, Telfer M, Twigger J, et al. Assessing IDU prevalence and health consequences (HCV, overdose and drug-related mortality) in a primary care trust: implications for public health action. *J Public Health (Oxf)* 2009;31:374-382. .
  23. Böhning D, Suppawattanabodee B, Kusolvitkul W, Viwatwongkasem C. Estimating the number of drug users in Bangkok 2001: a capture-recapture approach using repeated entries in one list. *Eur J Epidemiol* 2004;19:1075-1083.