

ESTIMATION OF PROBLEM DRUG USERS IN PRAGUE IN 2011 FROM LOW-THRESHOLD DATA: MODIFIED CAPTURE-RECAPTURE METHOD, ADJUSTED FOR CLIENTS AVOIDING ANY IDENTIFICATION (NON-CODED CLIENTS)

Bruno Sopko^{1,2}, Kateřina Škařupová³, Vlastimil Nečas^{1,5}, Viktor Mravčík^{1,4,5}

¹National Monitoring Centre for Drugs and Drug Addiction, Prague, Czech Republic

²Department of Medical Chemistry and Clinical Biochemistry, 2nd Faculty of Medicine, Charles University and Motol University Hospital, Prague, Czech Republic

³Institute for Research on Children, Youth and the Family, Faculty of Social Studies, Masaryk University, Brno, Czech Republic

⁴Department of Addictology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic

⁵National Institute of Mental Health, Prague, Czech Republic

SUMMARY

Aim: Local prevalence estimates of problem drug use (PDU) are crucial in the process of assessment of drug situation and trends and for the planning of evidence-based policy responses. The aims of our study are twofold: to estimate the number of problem drug users (PDUs) in the Czech capital city in 2011, and to examine the usability of the capture-recapture method (CRM) modified for data from low-threshold programmes (LTPs) for drug users.

Method: Six independent LTPs provided data for analysis (SANANIM, Drop-in and Progressive, each of these providing one drop-in centre and one outreach programme). After adjustment of the standard CRM formula for cases without an individual identifier, the overlaps between programmes were calculated and the size of hidden population was estimated.

Results: In total, it was estimated that there were 10,754 PDUs in Prague in 2011. The current estimate is in line with estimates obtained previously using another indirect standard approach – the multiplier method.

Conclusion: The modified version of CRM thus proved a reliable method for local PDU estimates.

Key words: drug use, low-threshold programmes, population, capture-recapture, estimate

Address for correspondence: B. Sopko, Department of Medical Chemistry and Clinical Biochemistry, 2nd Faculty of Medicine, Charles University and Motol University Hospital, V Úvalu 84, 150 06 Prague 5, Czech Republic. E-mail: bruno.sopko@lfmotol.cuni.cz

<http://dx.doi.org/10.21101/cejph.a4096>

INTRODUCTION

Heavy drug use is often stigmatized or associated with social exclusion or crime and therefore of a hidden nature. A substantial proportion of the drug-using population is not in contact with any kind of service. Therefore, prevalence estimates of problem drug use (PDU) are crucial for the assessment of drug situation and trends, the public health burden of problem drug use and planning of evidence-based policy responses (1, 2). For the estimation of PDU, indirect methods are preferred since direct methods such as cross-sectional questionnaire surveys cannot reach hidden subpopulations sufficiently. The most frequently used methods are the multiplier method and the capture-recapture method, which are based on the extrapolation of information on the portion of population which is captured in available data sources and registers (3, 4).

Traditionally, the PDU estimates for Prague have been calculated annually using the multiplier method applied to the aggregated number of clients of low-threshold services (LTSs), with yearly results ranging from 8,400 to 11,500 PDUs in the period between 2005 and 2011 (5). The multiplier method, however, does not account for the probable situation in which one client uses more than one service within the year for which the estimate is being calculated (i.e. double counting). Our study aimed to estimate the number of problem drug users in the Czech capital city, Prague, by applying a modified version of the capture-recapture method (CRM) to data from the low-threshold facilities (i.e. outreach programmes and drop-in centres providing harm reduction (HR) services for injecting drug users) and to assess overlaps in the clientele between programmes operating in the geographical area of Prague.

MATERIALS AND METHODS

Definition

We adopted the European definition of PDU that includes long-term, regular, and/or injecting use of opioids, cocaine, and amphetamines (6–8). Cocaine use is excluded from the working definition because of the close-to-zero prevalence of the drug in the Czech Republic; amphetamines are mainly represented by methamphetamine, locally known as pervitin (9).

Capture-recapture Log-linear Analysis

Capture-recapture log-linear analysis (CRM) is a recommended method for the estimation of number of individuals in hard-to-reach groups where a census or direct estimation is difficult and therefore is well suited for the local estimation of PDU prevalence (10–13). CRM seeks overlaps between two or more data sources in order to estimate the size of hidden part of the population. However, the method assumes the independence of all data sources and requires a unique identifier for each individual in the sample (8). These requirements represent a challenge for the application of CRM to data from the low-threshold programmes that provide services anonymously and where clients' participation in more than one programme is assumed but the exact pattern is unknown.

Data

We worked with two types of data. Routine monitoring data from the low-threshold facilities contained only information on individual clients in contact with each programme in 2011. Since the services are provided anonymously, all LTPs offer voluntary registration of the clients using an anonymised individual identifier (called the HR code) for monitoring purposes. The HR code is created as follows: the first three characters of first name of the client's mother, followed by two digits representing the client's day of birth, then the first three characters of client's first name and two digits representing the client's month of birth (e.g. MAR12BRU06 for the first author). The size of clientele without an HR code is generally not known. The information on the proportion of non-coded clients and information on the structure of drug-using population in terms of basic demographics (age and gender) and drug-using patterns (primary drug and polydrug use) were obtained in a specialised survey among clients of the respective LTPs in 2012.

Six low-threshold programmes that operate independently in the city of Prague (SANANIM, Drop-in and Progressive, each of these providing one drop-in centre and one outreach programme) provided data for the analysis. All the programmes run independently, with separate budget managements and teams in close and sometimes overlapping areas (the total number of coded individuals was 6,786).

No other data was collected by LTSs and thus none was gathered for the analysis. In order to supply additional information (primary drug, sex and drugs used) and to assess the proportion of non-coded individuals (HR codes are offered on a voluntary basis) among the clients of HR services, a special survey was carried out. The survey was designed as a more detailed form of

data collection from clients who contacted the programmes during a period of 1–2 weeks in September 2012. Data on the gender and age of clients as well as their primary drug was collected (altogether 634 individuals were questioned). Data was provided on the aggregated level.

Adjustment for clients without an HR code was performed. We assumed that, despite their reluctance to receive a code, the behaviour of the non-coded clients would follow a similar pattern to that of the coded clients. Therefore, the non-coded clients can be proportionally added to numbers in the distribution table.

Adding Non-coded Clients to Distribution

The problem with the non-coded clients can be described as being that there is a part of the PDU population who choose to be anonymous as far as possible. Using the data supplied by LTSs, we are able to determine the sets of coded clients and their overlap and the overall population of PDUs in contact with LTSs (both coded and non-coded and all clients). From these data we can derive the following "correction" of the standard capture-recapture method.

Two Sample Cases

Considering two low-threshold programmes, supplying the following data (Fig. 1):

X – number of clients attending low-threshold programme 1 (coded and non-coded)

Y – number of clients attending low-threshold programme 2 (coded and non-coded)

x – number of coded clients attending low-threshold programme 1 only

y – number of coded clients attending low-threshold programme 2 only

z – number of coded clients attending both low-threshold programmes, 1 and 2

The overall population size (S) is usually determined as:

$$S = \frac{(x+z) \times (y+z)}{z}$$

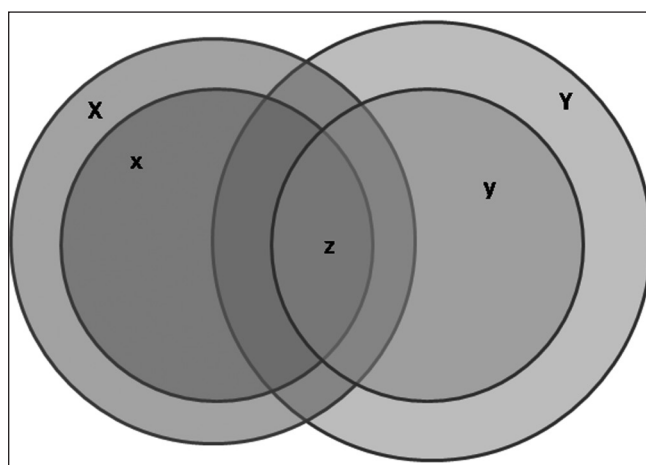


Fig. 1. Venn diagram of the intersection of the two populations. Coded-only clients are in darker shades, all clients (coded plus non-coded) are in lighter shades. The proportion of intersection remains in the same ratio as in case of coded-only clients.

However, some of the clients avoid being given any code and therefore the number of coded clients is smaller than the total number. Assuming that the non-coded clients exhibit the same behaviour as the coded ones (except their unwillingness to receive any possible identification) we can correct the equation for estimating the overall population size (S_{corr}):

$$S_{corr} = \frac{X \times Y}{\frac{z \times (X + Y)}{(x + z) \times (y + z)}} = \frac{X \times Y \times (x + z) \times (y + z)}{z \times (X + Y)}$$

Corrected Method

Let us presume that N_i is the number of coded clients of programme i only, N_{ij} is the number of coded clients reported by programmes i and j together, and NT_i is the sum of all coded clients reported by programme i . P_i is the sum of coded and non-coded clients reported by programme i .

The corrected distribution number (F_i , $i \in (1; n)$) can thus be calculated using the following equation:

$$F_i = \frac{N_i}{\sum_{i=1}^n NT_i} \times \sum_{i=1}^n P_i$$

where $i \in (1; n)$

Table 1. Numbers of coded clients and corrected numbers of all clients reported by six low-threshold programmes in Prague in 2011

Programme	Number of coded clients	Corrected number of all clients
CC SANANIM	1,668	2,285
OP SANANIM	1,455	2,469
CC Drop-in	401	539
OP Drop-in	1,142	1,905
CC Progressive	977	1,331
OP Progressive	1,143	1,902
Total	6,786	10,431

The Recapture package was used for the statistical analysis. As our data had been shown not to be fully independent, we used the bias-correcting function closedp.mX; the Akaike Information Criterion (AIC) was used to determine the model of the probable interaction (14, 15).

Table 2. Frequency distribution of coded and all clients by the number of programmes entered

Number of programmes used	Frequency (coded clients)	Corrected frequency (all clients)
$i = 1$	2,722	4,071
$i = 2$	476	726
$i = 3$	290	463
$i = 4$	296	469
$i = 5$	178	277
$i = 6$	28	43
Total	3,990	6,049

i = number of programmes registering the code (client)

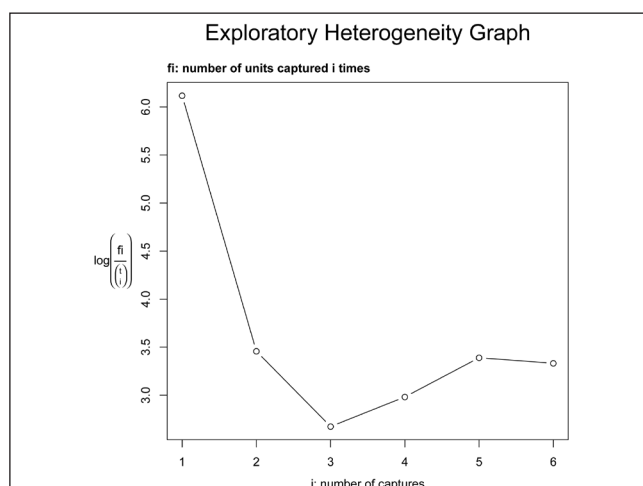


Fig. 2. The heterogeneity of distribution of codes among programmes (according to Baillargeon and Rivest, 2007).

Table 3. The results of special survey carried out by HR programmes in 2012

Service provider	Type of service	Number of clients		Primary drug										Women (%)	Average age men	Average age women	Average age
		Total	Coded clients	Heroin	Methamphetamine	Cocaine	Cannabinoids	MDMA	Volatiles	Buprenorphine	Methadone	Alcohol	Polydrug users				
Progressive	CC	264	217	24	98	0	5	-	-	129	2	4	23	25	32.69	27.21	31.58
	OP	774	436														
Drop-in	CC	259	213	38	173	3	23	-	-	63	41	4	4	25	37.89	33.11	36.66
	OP	759	428														
SANANIM	CC	271	218	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	OP	794	438														

The numbers in italics were missing and have been obtained as an average of the other corresponding values.

RESULTS

The six low-threshold programmes reported 6,786 coded clients in total (Table 1).

When the code lists supplied by low-threshold programmes were compared, 3,990 unique codes were identified; of these 2,722 (68.2%) were reported by one programme only, while 1,268 (31.8%) were reported by two or more programmes; 28 clients with codes had been in contact with all six programmes (Table 2). Table 6 shows the overlap of codes (clients) between the HR programmes and corrected numbers as well.

The high statistical probability of interaction between data sources is indicated in Fig. 2 (the Heterogeneity graph is far from linear, and the probability of multiple reporting, i.e. 4–6 reports, is significantly increased). Therefore, CRM log-linear analysis was run with the following model: $Source_1 + Source_2 + Source_3 + Source_4 + Source_5 + Source_6 + Source_1 * Source_2 * Source_3 * Source_4 * Source_5 + Source_1 * Source_2 * Source_4 * Source_5 * Source_6 + Source_1 * Source_2 * Source_3 * Source_4 * Source_5 * Source_6$, where “*” denotes the interaction among sources and Source x means xth source of client codes.

The results of special survey are shown in Table 3 and the overall results together with PDU estimates in Table 4.

Using the results of special survey, we calculated the overall distribution of drugs among PDUs (Table 5). Please note that users could report more than one drug in the special survey and thus groups by drugs overlap in CRM, while in MM just the primary drug is collected.

DISCUSSION

The number of clients unwilling to receive an HR code is important in the estimation of the size of PDU population. These clients would probably resist using the code all the time and in contact with all of HR programmes. Therefore, there are two populations of PDUs: the coded clients and non-coded clients.

Table 4. The results of the capture-recapture method applied to coded-only clients and those corrected for no codes

CRM Prague 2011	Estimation of PDU population
	Population (95% CI)
Coded clients only	7,280 (6,983; 7,603)
Coded + non-coded clients	10,754 (10,405; 11,127)

Table 5. The overall results (central values) obtained by the capture-recapture method including the special survey in comparison with the multiplier method (all estimates refer to the year 2011)

Source	Total PDUs	Men	Women	Heroin	Methamphetamine	Cocaine	Buprenorphine	Methadone
CRM	10,754	8,056	2,689	2,581	5,592	64	4,732	860
MM	10,900	n.a.	n.a.	2,200	5,400	n.a.	3,300	n.a.

Running CRM on the coded population then only leads to underestimation of the total PDU population as the random samples are drawn from the “codable” population only, and the non-coded population is omitted (clients without codes are actually looking for contact with LTSs but avoid registered contact involving identification). Assuming that, apart from their unwillingness to obtain the code, the overall contact behaviour of non-coded clients follows the same pattern as in the case of coded clients, we have adjusted the number of coded clients to the proportion of non-coded clients as shows the equation above. This equation conserves all ratios in the distribution table (Table 6).

The data used were supplied by six low-threshold programmes in Prague but there are two more LTSs in Prague – the NGOs Sance and Eset-Help that did not participate in our study. The first one does not focus primarily on drug users and the second one has a limited coverage and neither of them belongs among the major service providers in Prague. Therefore, we can assume that the absence of these two low-threshold programmes does not play a crucial role in the estimation and that a substantial proportion of PDU population is in contact with the participating programmes.

Using the log-linear modelling in the Recapture package (14), we found the interaction model to be $Source_1 + Source_2 + Source_3 + Source_4 + Source_5 + Source_6 + Source_1 * Source_2 * Source_3 * Source_4 * Source_5 + Source_1 * Source_2 * Source_4 * Source_5 * Source_6 + Source_1 * Source_2 * Source_3 * Source_4 * Source_5 * Source_6$, where “*” denotes the interaction among sources and Source x means xth source of the client codes (based on AIC). As can be seen from Table 5, the estimate of PDU population based on coded clients only is approximately 7,000. This number is an underestimation when compared to 10,900 estimated for the same year by the multiplier method (5). After correction, the PDU population was determined to be 10,754 (10,754–11,127), which is in accordance with the previous estimates obtained via the multiplier method (10,900).

The data shows that the clients of LTSs are predominantly in contact with just one programme (68%). On the other hand, only 0.7% of clients were in contact with all six programmes. The overlap among the programmes is not too large; for instance, only 1.8% of clients were in contact with both the drop-in centre and outreach programme of the same provider. The only noticeable overlap was detected among outreach programmes (48.2% of clients – both coded and non-coded – were in contact with at least two OPs; coded clients only accounted for 44.7% of those in contact with multiple OPs). As reported in Table 5, which shows a comparison between the results obtained with corrected CRM with the results obtained with annually applied MM, we can see that the data obtained by corrected CRM is in accordance with the estimates obtained via MM.

Our modification of the CRM method solves the problem of clients avoiding any identification and is therefore suitable for estimating the number of PDUs from the data of low-threshold programmes (LTPs) in large metropolitan areas where more than one LTP service operates.

CONCLUSION

The results of our study show that the data obtained from several low-threshold HR programmes can be used for the successful estimation of PDU prevalence using capture-recapture

Table 6. The frequency of code distribution among low-threshold programmes (“0” denotes the absence and “1” the presence of client in the programme)

CC SANANIM (c1)	OP SANANIM (c2)	CC Drop-in (c3)	OP Drop-in (c4)	CC Progressive (c5)	OP Progressive (c6)	Frequency (coded clients)	Corrected frequency (all clients)
1	0	0	0	0	0	815	1,013
0	1	0	0	0	0	553	1,003
0	0	1	0	0	0	225	274
0	0	0	1	0	0	369	654
0	0	0	0	1	0	399	486
0	0	0	0	0	1	361	641
1	1	0	0	0	0	53	80
1	0	1	0	0	0	10	12
1	0	0	1	0	0	31	45
1	0	0	0	1	0	113	139
1	0	0	0	0	1	36	53
0	1	1	0	0	0	11	19
0	1	0	1	0	0	67	120
0	1	0	0	1	0	14	22
0	1	0	0	0	1	62	111
0	0	1	1	0	0	4	7
0	0	1	0	1	0	11	13
0	0	1	0	0	1	3	5
0	0	0	1	1	0	12	18
0	0	0	1	0	1	31	55
0	0	0	0	1	1	18	27
1	1	1	0	0	0	0	0
1	1	0	1	0	0	45	71
1	1	0	0	1	0	19	27
1	1	0	0	0	1	43	68
1	0	1	1	0	0	3	4
1	0	1	0	1	0	6	7
1	0	1	0	0	1	3	4
1	0	0	1	1	0	12	17
1	0	0	1	0	1	17	26
1	0	0	0	1	1	16	22
0	1	1	1	0	0	3	5
0	1	1	0	1	0	5	8
0	1	1	0	0	1	3	5
0	1	0	1	1	0	15	25
0	1	0	1	0	1	84	150
0	1	0	0	1	1	8	13
0	0	1	1	1	0	3	4
0	0	1	1	0	1	0	0
0	0	1	0	1	1	1	1
0	0	0	1	1	1	4	6
1	1	1	1	0	0	7	11
1	1	1	0	1	0	2	3
1	1	1	0	0	1	1	2

Continued on the next page

Continued from the previous page

CC SANANIM (c1)	OP SANANIM (c2)	CC Drop-in (c3)	OP Drop-in (c4)	CC Progressive (c5)	OP Progressive (c6)	Frequency (coded clients)	Corrected frequency (all clients)
1	1	0	1	1	0	32	48
1	1	0	1	0	1	145	235
1	1	0	0	1	1	42	64
1	0	1	1	1	0	2	3
1	0	1	1	0	1	3	5
1	0	1	0	1	1	7	10
1	0	0	1	1	1	13	19
0	1	1	1	1	0	3	5
0	1	1	1	0	1	6	10
0	1	1	0	1	1	2	3
0	1	0	1	1	1	27	45
0	0	1	1	1	1	4	6
1	1	1	1	1	0	3	4
1	1	1	1	0	1	19	30
1	1	1	0	1	1	6	9
1	1	0	1	1	1	133	207
1	0	1	1	1	1	3	4
0	1	1	1	1	1	14	23
1	1	1	1	1	1	28	43
Total						3,990	6,049

log-linear analysis. However, the anonymous nature of provision of low-threshold services and the consequent insufficient and incomplete unique identification of clients represents a challenge for the method, which can be addressed using small-scale surveys among clients and correction for the anonymous clients.

Acknowledgements

Internal Grant Agency of the Ministry of Health (grant number NT/14064), project "National Institute of Mental Health (NIMH-CZ)", grant number CZ.1.05/2.1.00/03.0078 and the European Regional Development Fund. The managing authorities had no role in the design of review and collection, analysis, and interpretation of data.

Computational resources were provided by the MetaCentrum under the LM2010005 programme and CERIT-SC under the Centre CERIT Scientific Cloud programme, part of the Research and Development for Innovations Operational Programme, Reg. no. CZ.1.05/3.2.00/08.0144.

REFERENCES

- Degenhardt L, Hall W. Extent of illicit drug use and dependence, and their contribution to the global burden of disease. *Lancet*. 2012 Jan 7;379(9810):55-70.
- Thanki D, Vicente J. PDU (Problem drug use) revision summary. Lisbon: European Monitoring Centre for Drugs and Drug Addiction; 2013.
- Kraus L, Augustin R, Frischer M, Kümmler P, Uhl A, Wiessing L. Estimating prevalence of problem drug use at national level in countries of the European Union and Norway. *Addiction*. 2003 Apr;98(4):471-85.
- Rehm J, Room R, van den Brink W, Kraus L. Problematic drug use and drug use disorders in EU countries and Norway: an overview of the epidemiology. *Eur Neuropsychopharmacol*. 2005 Aug;15(4):389-97.
- Mravčík V, Grohmannová K, Chomynová P, Nečas V, Grolmusová L, Kiššová L, et al. Annual report: The Czech Republic – 2011 drug situation. Prague: Office of the Government of the Czech Republic; 2012.
- European Monitoring Centre for Drugs and Drug Addiction. An overview of the problem drug use (PDU) key indicator. Lisbon: EMCDDA; 2009.
- European Monitoring Centre for Drugs and Drug Addiction. Methodological guidelines to estimate the prevalence of problem drug use on the local level. Lisbon: EMCDDA; 1999.
- European Monitoring Centre for Drugs and Drug Addiction. Key epidemiological indicator: prevalence of problem drug use. Lisbon: EMCDDA; 2004.
- Mravčík V, Buresova Z, Popov P, Miovisky M. Sobering-up stations in the Czech Republic in the context of analogous models of care for acute intoxications in Europe. *Cas Lek Ces*. 2013;152(3):129-34. (In Czech.)
- Otis DL, Burnham KP, White GC, Anderson DR. Statistical-inference from capture data on closed animal populations. *Wildlife Monogr*. 1978;(62):7-135.
- Wittes J, Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. *J Chronic Dis*. 1968 Aug;21(5):287-301.
- Estimating the size of a closed population. discrete multivariate analysis: theory and practice. In: Bishop Y, Fienberg S, Holland P. *Discrete multivariate analysis theory and practice*. New York: Springer; 2007. p. 229-56.
- Tilling K, Sterne JA. Capture-recapture models including covariate effects. *Am J Epidemiol*. 1999 Feb 15;149(4):392-400.
- Baillargeon S, Rivest L. Rcapture: loglinear models for capture-recapture in R. *J Stat Softw*. 2007 Apr;19(5):31.
- Team RC. R: a language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2013 [cited 2016 Jan 15]. Available from: <https://www.r-project.org/>.

Received September 9, 2014
Accepted in revised form December 21, 2015