

A capture-recapture technique to estimate the size of the injecting drug user population attending syringe exchange programs: AjUDE-Brasil II Project

Método de captura-recaptura para estimar o tamanho da população de usuários de drogas injetáveis atendidos por programas de redução de danos: Projeto AjUDE-Brasil II

Sueli Aparecida Mingoti ¹
Waleska Teixeira Caiaffa ²
Projeto AjUDE-Brasil II ³

Abstract

This paper presents the results of a study with a two-occasion capture-recapture design. The data are part of the AjUDE-Brasil II Project, carried out in 2000-2001. Estimation of the size of the IDU population attending a syringe-exchange program (SEP) in São José do Rio Preto, Salvador, and Porto Alegre, Brazil, was performed using Chao's model. Capture probabilities were also estimated. For Porto Alegre a comparison of the results from the AjUDE-Brasil I and AjUDE-Brasil II Projects was performed. Results are also presented for error rates secondary to the choice of matching criteria.

Harm Reduction; Intravenous Substance Abuse; Street Drugs; Syringe-Exchange Programs

Introduction

Estimation of population size using capture-recapture methods ^{1,2} has been an issue in many studies over the years. Although these models are frequently used to estimate the size of animal populations, more recently they have become popular in other areas such as epidemiology and social sciences ^{3,4}. Interesting studies include McKeganey et al. ⁵ and Mastro et al. ⁶, where capture and recapture methods were used to estimate the number of HIV-infected drug users in Bangkok, Thailand, and the size of the commercial sex street-working population and HIV infection in Glasgow, Scotland. Capture and recapture methods have also been used to estimate the prevalence and underreporting of certain diseases such as AIDS and diabetes ^{7,8} and to perform adjustments for census undercount ⁹. Although practical interest currently focuses on models that allow for heterogeneity and trap response, which require more than just one recapture, the two-occasion capture-recapture design is still widely used in situations where high cost, extended time, and difficult data collection preclude more than one recapture. A good example relates to surveys of hard-to-reach populations. The AjUDE-Brasil I Project assessed injection drug users (IDUs) in five syringe exchange programs (SEPs) in Brazil ^{10,11}. The city of Porto Alegre in southern Brazil was used as the pilot for the capture-

¹ Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

² Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.

³ Other members listed at the end of the paper.

Correspondence

S. A. Mingoti
Departamento de Estatística,
Instituto de Ciências Exatas,
Universidade Federal
de Minas Gerais.
Av. Antonio Carlos 6627,
Belo Horizonte, MG
31270-970, Brasil.
sueli@est.ufmg.br

recapture methodology to estimate the size of the IDUs population attending the local SEP. The IDUs population was very difficult to interview, and interviewers required special training. The overall cost of the survey was high. Time-consuming data collection procedures would have been significantly increased if a multiple capture-recapture design was used. Of interest, the results obtained with just one recapture were very reasonable, indicating that there was probably no need to conduct additional recaptures.

Another version of the AJUDE-Brasil Project was conducted in 2000-2001¹². The capture-recapture technique was used again to estimate the size of the IDU population attending the SEPs located in São José do Rio Preto, Salvador, and Porto Alegre. The results of the previous study are presented in this paper.

Capture-recapture methodology

If the population study consists of N elements, N being unknown and finite, capture-recapture methods can be applied to estimate the size of N . Statistical models have been developed considering two basic situations: (1) The population is assumed to be closed in the sense that if a certain number of elements leave the population during the sampling period, a similar number of elements enters the population during the same period¹³. Thus, N remains unchanged over time. (2) Immigration, death, and birth are allowed. The population is then classified as open, and the statistical models consider parameters describing the possible changes in population size^{14,15,16,17,18}. A good overview of capture-recapture methods for closed populations is available in Chao's reference text¹⁹.

Multinomial probabilistic¹³ and log-linear regression methods^{19,20} have been used to develop sensitive estimators of population size. In both methodologies the estimators are derived on the basis of the presence of some variation factors in the capture occasions such as: time, environment, particular individual behavior, and effect of the three sources taken together. Data used to estimate the population size can be collected on two or more occasions (multiple captures^{3,13,14}). In this paper we focus only on the two-capture model for closed populations.

Two-occasion capture-recapture design

In the two-occasion capture-recapture design, two random samples are collected: the first sample (capture) has m distinct elements of the

population which are captured, tagged, and returned to the population. The second sample has n elements from which s were already observed in the first sample. The latter are called recaptured elements. It is assumed that the markers on the elements captured in the first sample are not lost during the time elapsed between collection of the two samples, so that the elements captured in the first sample can be identified unequivocally in the second sample¹³.

Estimation of population size N

A well-known estimator for size N of a closed population and a capture-recapture design with only two capture occasions is the Lincoln-Petersen estimator^{21,22}, first used by Laplace in 1786 to estimate the size of the French population. The estimator is simple and based on the fact that the proportion of marked elements in the second sample is an estimator of the marked elements in the population before the second sample is collected. Considering this approach, the Lincoln-Petersen estimator (\hat{N}) is expressed as:

$$\hat{N} = \frac{n m}{s}$$

The fewer the recaptured elements, the greater the value of \hat{N} , and if $s = 0$ then \hat{N} is infinite. Capture probabilities for each occasion are estimated by

$$\hat{p}_i = \frac{n_i}{\hat{N}},$$

where n_i is the number of elements sampled (captured) on occasion i , $i = 1, 2$.

For the two-occasion capture-recapture design time variation in capture probabilities is the only factor that can be incorporated into the model. The estimator $M(t)$ proposed by Darroch²³, with bias correction based on Chao²⁴, is then used to estimate N . It is also possible to calculate a confidence interval for the true value of N . A simple form is to use the normal distribution to construct the confidence intervals. However, a more precise method was proposed by Burnham et al.²⁵, based on the assumption that the number of individuals in the population not captured in the sample has a log-normal distribution. Chao's estimator with the correction for confidence intervals is found in special software developed for capture-recapture designs, such as the user-friendly free-ware Capture and Mark (<http://www.cnr.colostate.edu/~gwhite/mark/mark.htm>, accessed on 18/Aug/2004).

Sample data

The data discussed in this paper are part of the AjUDE-Brasil II Project¹². The population consists of IDUs attending the three SEPs, located in São José do Rio Preto, Salvador, and Porto Alegre, Brazil. The first capture took place from May to August 2000 and the second from September 2000 to February 2001. For both capture and recapture occasions the IDUs answered a series of questions designed to identify whether they were recaptured in the second sample. These questions were adapted from other studies considering some specificities of Brazilian culture and were tested in a similar study in a SEP located in Porto Alegre during the AjUDE-Brasil I Project in 1998.

The information collected on the two occasions for each SEP was entered twice into a computer database. After validation, the data were matched using a Turbo-Pascal program designed for the AjUDE-Brasil I Project, combined with database sorting and visual inspection. Three different criteria were tested to identify the recaptured IDUs. In the first procedure, the IDU's name and initials, sex, and date of birth and parents' initials were used as matching variables. In the second, the two data sets were compared by using only the IDU's name, sex, and date of birth. In the third, matching used only the IDU's sex and date of birth and parents' initials. Finally, matching was performed using all five variables jointly plus highly detailed sorting and visual inspection of the two data sets. For the visual inspection other variables available in the dataset were used, such as the interviewer's opinion about whether a given IDUs was a recapture. This was considered the "gold standard" for comparisons. For this study in particular, information on the IDU's name was available, which allowed us to compare the results of the three criteria and to

estimate the error rate for each. This was an important part of the research, because in most studies of this kind, IDU names are not available and initials are used instead (in combination with other variables) to perform matching.

Estimation of the IDU population attending the three SEPs used Chao's estimator²⁴, with correction for the confidence interval. The Capture software was used for the estimation.

Informed consent was obtained from all individuals, and the protocols were approved by the Institutional Review Board of the Federal University in Minas Gerais (ETIC number 168/99, 01/Mar/1999).

Results

A total of 624 IDUs were interviewed for the capture-recapture study, 329 in the first sample and 434 in the second, with 139 recaptures. Table 1 shows the number of captured and recaptured IDUs for each SEP. Table 2 provides the distribution of the captured-recaptured IDUs for each SEP, according to the criterion used to determine the matching persons in the capture and recapture data sets. Table 3 shows the estimated populations for each SEP and the respective capture probabilities. The estimates for the Porto Alegre SEP, using the data set from the AjUDE-Brasil II Project, were compared with those from the capture-recapture design conducted in 1998 for the AjUDE-Brasil I Project. The results are shown in Table 4. Considering the three cities together, Chao's model estimated that a total of 1,024 IDUs were attending the three SEPs, with a standard deviation of 54.15 and a 95% confidence interval (931;1,145). Estimated capture probability was 0.32 for the first occasion and 0.42 for the second.

Of all the interviewed IDUs, 14 in São José do Rio Preto and 13 in Porto Alegre reported

Table 1

Number of injection drug users (IDUs) in capture and recapture samples from three Syringe Exchange Programs (SEP) in Brazil.

SEP (city)	Number of IDUs (1 st sample)	Number of IDUs (2 nd sample)	Number of recaptures	Number of distinct IDUs
São José do Rio Preto	79	99	36	136
Salvador	119	134	42	207
Porto Alegre	131	201	61	271
Total	329	434	139	597

Table 2

Frequency of recaptures according to the matching criteria in three syringe-exchange programs, Brazil.

Matching criteria	Frequency of recaptures (%)		
	São José do Rio Preto*	Salvador**	Porto Alegre***
1. IDU name, sex, date of birth, parents' initials	23 (63.8)	19 (45.2)	19 (31.4)
2. IDU name, sex, date of birth	33 (91.7)	32 (76.2)	37 (60.6)
3. IDU initials, sex, date of birth, parents' initials	23 (63.8)	21 (50.0)	23 (37.7)
4. IDU name, initials, sex, date of birth, parents' initials, and visual inspection of data sets (gold standard)	36	42	61

Percentages relate to:

* $r = 36$;** $r = 42$;*** $r = 61$.

Table 3

Estimated size (N) of injection drug user (IDU) population attending each syringe-exchange program (SEP).

SEP (city)	Number of distinct IDUs (both samples)	Estimated size (\hat{N})	Standard deviation	95% confidence interval for true (N)	Estimated capture probabilities (1 st ;2 nd) occasions
São José do Rio Preto	142	215	20.68	(185; 268)	(0.37; 0.46)
Salvador	211	376	37.99	(317; 468)	(0.32; 0.36)
Porto Alegre	271	429	33.11	(377; 508)	(0.31; 0.47)

Table 4

Estimated size (N) of injection drug user (IDU) population attending Porto Alegre SEP. AjuDE-Brasil I Project (1998) and AjuDE-Brasil II Project (2000-2001).

Variables	AjuDE I	AjuDE II
Number of IDUs in each sample (1 st ;2 nd) occasions	(55; 99)	(131; 201)
Number of recaptures	17	61
Number of distinct IDUs	137	271
Estimated size (\hat{N})	317	429
Standard deviation of \hat{N}	57.17	33.11
95% confidence interval for N	(235; 467)	(377; 508)
Estimated capture probabilities (1 st ;2 nd) occasions	(0.17; 0.31)	(0.31; 0.47)

having participated in a similar study in 1998. They were probably part of the sample collected in the AjUDE-Brasil I Project, since this was the only study of its kind in the two cities that year.

Discussion

Comparing the three different matching criteria with the gold standard, criterion number two was the best for all three SEPs, although with high error rates for Salvador (23.8%) and Porto Alegre (39.4%). Criteria one and three presented similar results, suggesting that the use of IDU initials as a matching variable was as reliable as the use of IDU name. However, when compared to the gold standard, a large amount of information could be lost if a detailed analysis is not performed by the researcher, especially for Porto Alegre, where only 60.6% of recaptures were correctly identified by criterion two.

Comparison of criteria one and two suggested that inclusion of the IDU's parents' initials as a matching variable decreased the number of recaptures. Considering that the fewer the recaptured elements, the higher the estimated population size (N), a decrease in the number of recaptures due to failures in the matching criteria may overestimate the true value of N .

As depicted in Table 3, estimated capture probabilities for the first and second occasions for each SEP were more distinct in São José do Rio Preto and Porto Alegre than in Salvador. For São José do Rio Preto and Porto Alegre, the capture probability increased on the second occasion, suggesting a time variation factor. The smaller standard deviation shows that the population size was estimated more precisely in São José do Rio Preto.

Table 4 shows that the number of captured-recaptured individuals in Porto Alegre was 1.97 times that observed in the same city in the AjUDE-Brasil I Project. The number of IDUs observed on the first and second occasions was 2.4 and 2.03 times that of AjUDE Brasil I, and the number of recaptures was 3.6 times greater. The proportion of recaptures in relation to the number of distinct observed IDUs in the AjUDE-Brasil I and II Projects was 12.4 and 22.5%, respectively. This increased probably in the second study appears to be due to the fact that in the AjUDE-Brasil I, just one month was used to collect data for each occasion and the two samples were collected during very close periods (capture from April 1, 1998 to May 1,

1998; recapture from May 2, 1998 to June 6, 1998). In the AjUDE-Brasil II Project, three months were used to collect the capture sample and two months for the recapture sample. This allowed more accurate observation of the behavior of IDUs attending the Porto Alegre SEP in AjUDE-Brasil II as compared to the AjUDE-Brasil I. This fact also reflects the standard deviation in the population estimate (which was smaller for the AjUDE-Brasil II Project) as well as the confidence interval.

The difference in capture probabilities on the two occasions points to a time variation factor in the two surveys in Porto Alegre. Considering the population estimates shown in Table 4, the number of IDUs attending the Porto Alegre SEP increased 35.5% from early 1998 to early 2001. Thus, in Porto Alegre, the investment in labor and resources in the SEP activities was producing an important social return.

Caiaffa et al.¹¹ present an interesting discussion on the assumption of closed populations in this kind of study. Briefly, open population models cannot be applied with only two capture occasions. Thus, even if some questions are raised concerning the closed population assumption, they cannot be answered technically with the data set analyzed in this paper. Other studies would have to be performed, using multiple recaptures. However, more time and money would be expended to collect the data. In addition, as pointed out by Caiaffa et al.¹¹, for each SEP analyzed in this paper, the capture and recapture data used to estimate the population size may not be a truly random sample, since IDUs are a "*partially hidden population*". Questions could also arise concerning the true population being estimated, since some groups of IDUs may not be well represented in the samples, such as IDUs that are recent injectors or those that are ill¹¹. In any case, this study's results highlight the fact that the capture-recapture method is an important tool which can be combined with epidemiological information to help increase knowledge concerning hard-to-reach and difficult-to-count populations like as injecting drug users. The method can thus aid the decision-making process on planning and organizing public health programs to curb the spread of HIV and other blood-borne infections.

Resumo

O artigo apresenta os resultados de um delineamento de captura e recaptura em dois estágios realizado no período de 2000-2001, durante a execução do Projeto AJUDE-Brasil II. O número de usuários de drogas injetáveis atendidos pelos Programas de Redução de Danos (PRD), nas cidades de São José do Rio Preto, Salvador e Porto Alegre, Brasil, foi estimado utilizando-se o modelo de Chao. As probabilidades de captura também foram estimadas. Uma estimativa adicional foi calculada para a cidade de Porto Alegre, comparando-se os dados dos Projetos AJUDE-Brasil I e II. Alguns resultados sobre a taxa de erros, decorrentes da escolha do critério usado para emparelhamento dos dados e identificação dos recapturados são apresentados.

Redução do Dano; Uso Indevido de Drogas Parenterais; Drogas Ilícitas; Programas de Troca de Seringas

Contributors

S. A. Mingoti participated in the design of the questionnaire used to collect the data for the capture-recapture study, implementation of the capture-recapture statistical model, and analysis and discussion of the final results. W. T. Caiaffa participated in the design of the questionnaire used to collect the data for the capture-recapture study, supervision of the fieldwork, and discussion of the statistical results.

Other members of the AJUDE-Brasil II Project

T. Andrade, F. I. Bastos, R. T. Caiaffa, I. M. Cardoso, A. B. Carneiro-Proietti, M. A. B. Chagas, S. Deslandes, D. Doneda, R. E. M. Eller, E. M. A. Ferreira, D. Gandolfi, M. Hacker, N. Januário, R. Knoll, A. C. S. Lopes, A. C. Maia, M. Malta, R. C. R. Marinho, D. L. Matos, R. M. B. Mayer, H. F. Mendes, E. A. Mendonça, I. F. M. Picinim, F. A. Proietti, R. C. Silva, M. D. S. Sudbrack.

Acknowledgments

The AJUDE-Brasil Project II was conducted by the Universidade Federal de Minas Gerais (UFMG) with the technical and financial support of the cooperative project between the Brazilian National STD/AIDS Program (PNDST/AIDS) and the United Nations Office on Drugs and Crime (UNODC), project no. AD/BRA/EO2

We are grateful to the SEP coordinators, the outreach workers, and the IDUs who agreed to participate in this study.

Drs. W. T. Caiaffa, F. I. Bastos, and F. A. Proietti are recipients of Brazilian National Council for Scientific and Technological Development (CNPq) scholarships.

References

1. Seber GAF. A review of estimating animal abundance. *Biometrics* 1986; 42:267-92.
2. Seber GAF. A review of estimating animal abundance II. *Int Stat Rev* 1992; 60:129-66.
3. Oliveira MTC, Caiaffa WT, Mingoti SA, Fernandez H, Macedo LM, Mafra A, et al. Underreporting of AIDS cases in Brazil: applications of capture-recapture methods. In: XIII International AIDS Conference. Durban: International AIDS Society; 2000. p. 378.
4. Tsay PK, Chao A. Population size estimation for capture-recapture model with applications to epidemiological data. *J Appl Stat* 2001; 28:25-36.
5. McKeganey N, Barnard M, Leyland, A, Coote I, Follet E. Female street-working prostitution and HIV infection in Glasgow. *BMJ* 1992; 305:801-4.
6. Mastro TD, Kitayaporn D, Weniger BG, Vanichseni S, Laosunthorn V, Uneklabh T, et al. Estimating the number of HIV-infected drug users in Bangkok: a capture-recapture method. *Am J Public Health* 1994; 7:1094-9.
7. Bernillon P, Lievre L, Pillonel J, Laporte A, Costagliola D; The Clinical Epidemiology Group from CISIH. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993. *Int J Epidemiol* 2000; 29:168-74.
8. Ismail AA, Beeching NJ, Gill GV, Bellis MA. How many data sources are needed to determine diabetes prevalence by capture-recapture? *Int J Epidemiol* 2000; 29:536-41.
9. Bell WR. Using information from demographic analysis in post-enumeration survey estimation. *J Am Stat Assoc* 1993; 88:1106-18.
10. Caiaffa WT, Proietti ABC, Proietti F, Guimarães MD, Mingoti SA, Deslandes S, et al. Projeto AJUDE-Brasil: avaliação epidemiológica dos usuários de drogas injetáveis dos projetos de redução de danos apoiados pela CN-DST/AIDS. *Série Avaliação* 6. 2001. <http://www.aids.gov.br> (accessed on 18/Aug/2004).
11. Caiaffa WT, Mingoti SA, Proietti FA, Carneiro-Proietti AB, Silva RC, Lopes ACS. Estimation of the number of injecting drug users attending an outreach syringe-exchange program and infection with human immunodeficiency virus (HIV) and hepatitis C virus: the AJUDE-Brasil Project. *J Urban Health* 2003; 80:106-14.
12. Caiaffa WT, Bastos FI, Proietti FA, Reis ACM, Mingoti SA, Gandolfi D, et al. Practices surrounding syringe acquisition and disposal: effects of syringe exchange programmes from different Brazilian regions – the AJUDE-Brasil Project. *Int J Drug Policy* 2003; 14:365-71.
13. Otis DL, Burnhan KP, White GC, Anderson DR. *Statistical inference from capture data on closed populations*. Bethesda: The Wildlife Society; 1978. (Wildlife Monographs, 62).
14. Abuabara MAP, Petrere Jr. M. *Estimativas da abundância de populações animais: introdução à técnica de captura e recaptura*. Maringá: Editora da Universidade Estadual de Maringá; 1997.
15. Cormarck RM. Estimates of survival from the sighting of marked animals. *Biometrics* 1964; 51:429-38.

16. Huggins RM, Yip PSF. Estimation of the size of an open population from capture-recapture data using weighted martingale methods. *Biometrics* 1999; 55:387-95.
17. Jolly GM. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* 1965; 52:225-47.
18. Seber GAF. A note on the multiple recapture census. *Biometrika* 1965; 52:249-59.
19. Chao A. An overview of closed capture-recapture models. *Journal of Agricultural, Biological & Environmental Statistics* 2001; 6:158-75.
20. Bishop YMM, Fienberg SE, Holland PW. *Discrete multivariate analysis*. Cambridge: The MIT Press; 1988.
21. Lincoln FC. Calculating waterfowl abundance on the basis of banding returns. U.S. Department of Agricultural Circular 1930; 118:1-4.
22. Petersen GGJ. The yearly immigration of young plaice into the Limfjord from the German Sea. *Report of Danish Biology Statistics* 1896; 6:1-48.
23. Darroch JN. The multiple-recapture census I: estimation of a closed population. *Biometrika* 1958; 45:343-59.
24. Chao A. Estimating the population size for sparse data in capture-recapture experiments. *Biometrics* 1989; 45:427-38.
25. Burnham KP, Anderson DR, White GC, Brownie C, Pollock KH. *Design and analysis methods for fish survival experiments based on release-recapture*. Bethesda: American Fisheries Society; 1987. (American Fisheries Society Monographs, 5).

Submitted on 21/Jun/2005

Final version resubmitted on 27/Sep/2005

Approved on 27/Sep/2005