บทความวิจัย

การเปรียบเทียบและการประเมินแบบจำลองการจับ-จับใหม่ สำหรับการประมาณจำนวนผู้ที่มีความผิดปกติของการใช้โอปิออยด์

Model Comparison and Assessment for Closed Capture-recapture Models for Estimate Opioid Use Disorders

รัชฎาภรณ์ อึ้งเจริญ[1] ลี่ลี อิงศรีสว่าง[1*] แมททิว ลอว์[2]

Ratchadaporn Ungcharoen[1], Lily Ingsrisawang[1*] and Matthew Law[2]

บทคัดย่อ

วิธีการจับ-จับใหม่ (Capture-recapture Methods: CRC) เป็นวิธีที่ได้รับความนิยมในการประมาณขนาดของประชากรที่ไม่ทราบขนาดซึ่งวิธีนี้ได้มีการประยุกต์ใช้ในการปรับข้อมูลทางระบาดวิทยา การศึกษานี้มีวัตถุประสงค์เพื่อเปรียบเทียบตัวประมาณค่าของการจับ-จับใหม่ สำหรับการประมาณจำนวนผู้ที่มีความผิดปกติในการใช้โอปิออยด์ (OUD) โดยใช้ข้อมูล OUD ในปี 2014 ในรัฐนิวเซาท์เวลส์ ประเทศออสเตรเลีย วิธี CRC ใช้ในการประมาณจำนวนผู้ที่มีความผิดปกติ OUD โดยใช้ข้อมูลจาก 3 แหล่ง ประกอบด้วยหอผู้ป่วย ห้องฉุกเฉิน และทะเบียนการตาย วิเคราะห์ข้อมูลด้วย Rcapture โมเดลที่นำมาใช้สำหรับการเปรียบเทียบสำหรับการประมาณค่าจำนวนผู้ป่วย OUD ได้แก่ M_0 M_t M_h และ M_b และใช้เกณฑ์ AIC สำหรับการคัดเลือกโมเดลที่ดีที่สุด โดยใช้ข้อมูลจาก 3 แหล่งได้แก่ หอผู้ป่วย จำนวน 87 คน, ห้องฉุกเฉิน จำนวน 407 คน และ ทะเบียนการตาย จำนวน 15 คน และเมื่อจับคู่ข้อมูลที่ซ้ำกันทั้ง 3 แหล่งข้อมูลพบว่ามีจำนวน 54 คน ผลจากการประมาณค่าโดยใช้โมเดล ดังนี้ โมเดล M_0 และ M_h Chao (LB) จำนวน 666 คน โมเดล M_t จำนวน 465 คนโมเดล M_h Poisson 2 จำนวน 433 คน โมเดล M_h Gamma 3.5 จำนวน 351 คน และโมเดล M_b จำนวน 503 คน โดยพบว่าโมเดล M_t มีค่า AIC น้อยที่สุด (AIC = 51.962) ด้วยโมเดล M_t มีความเหมาะสม ในการประมาณจำนวนผู้ที่มีความผิดปกติในการใช้โอปิออยด์ โมเดลนี้ควรที่จะนำไปใช้กับการประมาณค่าอื่นที่มีบริบทที่คล้ายคลึงกันนี้ วิธีการประมาณนี้เป็นวิธีการที่ง่าย รวดเร็วสำหรับการประมาณค่าจำนวนผู้ที่มีความผิดปกติในการใช้โอปิออยด์

คำสำคัญ: วิธีการจับ-จับใหม่ การประมาณค่า ระบาดวิทยา โอปิออยด์

---

[1] นิสิตบัณฑิตศึกษา ปร.ด.(สถิติ) ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ 10903 ประเทศไทย

[1] ผู้ช่วยศาสตราจารย์, ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ กรุงเทพฯ 10903 ประเทศไทย

[2] ศาสตราจารย์ และหัวหน้าหน่วยชีวสถิติ และฐานข้อมูล มหาวิทยาลัยนิวเซาท์เวลล์ 2052 ประเทศออสเตรเลีย

[1] Graduate Student, Ph.D.(Statistics), Department of Statistics, Department of Statistics, Faculty of Sciences, Kasetsart University, Bangkok, 10903, Thailand

[1] Asst. Prof., Department of Statistics, Faculty of Sciences, Kasetsart University, Bangkok, 10903, Thailand

[2] Prof., Head of Biostatistics and Databases Program,University of New South Wales, New South Wales, 2052, Australia

[*] Corresponding author: E-mail: ingsri.lily@gmail.com

(Received: January 31, 2020; Revised: March 16, 2020; Accepted: March 30, 2020)

Abstract

A popular method for dealing with an unknown population size is the capture-recapture method (CRC). Then capture-recapture models were applied in the adjustment of epidemiology data. The aim of study was to compare the estimator of capture-recapture models in the closed population for estimation the number of opioid use disorders (OUD). The data of opioid use disorders in 2014, in New South Wales, Australia. The CRC was use to estimate the number of OUD cases based on three data sources, including patient department, emergency department, and national death index. The data were analyzed using the Rcapture package. The model used M_0, M_t, M_h, and M_b to compare the assessment of OUD data. The minimum Akaike's Information Criterion (AIC) value was used to select the best model. The three data sources were: 87 patients at the Patient Department, 407 cases from Emergency Department, and 15 cases from the National Death Index. The overlapping of the three data sources involved 54 cases. The results showed that the estimates obtained were 666 cases from M_0 model and M_h Chao (LB), 465 cases from M_t model, 433 cases from M_h Poisson 2, 351 cases from M_h Gamma 3.5 and 503 cases from M_b model. The smallest AIC was obtained for the M_t model (AIC = 51.962). With the M_t model more suitable for OUD were estimated. This model should be able to apply to any setting with similar context. The method provided a simple, quick method to estimate the numbers of OUD.

Keywords: Capture-recapture, Estimation, Epidemiology, Opioid

Introduction

A popular method for dealing with an unknown population size is the capture-recapture method, which was developed to estimate the number of wild animals in a population [1]. The method was then extended to other fields including software testing [2], epidemiology and public health [3-4]. This method is being increasingly used to estimate the number of people with a given disease including the completeness of cancer registries [5], homeless people[6], human immunodeficiency virus-positive population [7-8], HIV-infected injection drug users in Bangkok [9], and the number of drug users in Bangkok in 2001[10-11]. The closed capture-recapture method is one variety of this model that includes dependence among samples and can take the form of an ecological model, a log-linear model, and a sample-coverage model [12].

Log-linear models are popular for the estimation of unknown populations size in epidemiology and public health data. The most general log-linear model was presented by Fienberg[13]. Generally, epidemiological and health data apply a log-linear model for the estimation of the unknown population [14-16]. The interaction between sources reduces the dependence between sources. These are not considered to be sources of variation but do reduce the dependence between sources when using interaction terms. In an ecological model this allows relaxation of the assumption of equal probability of capture [17-18].

Log-linear ecological models can be in the form M_{tbh} (time varying, behavioral response and heterogeneity model) or one of its sub-forms depending on interpretation and context [19]. The approach may specify various forms of capture probabilities based on empirical investigations of animal ecology. These models relax the equal catchability assumptions [12,17].

The current study estimated the total number of opioid use disorders based on sources of variation models using data from the National Drug and Alcohol Research Centre (NDARC), University of New South Wales, Australia. The model combined up to three sources of variation among capture probabilities: M_t (time varying), M_b (behavioral response), and M_h (heterogeneity)[17, 20-21]. A useful structure for models has been introduced by Otis et al. [21].

The current study used ecological models for comparison among models of sources of variation for the assessment of the number of unknown Opioid Use Disorders (OUD) who had not registered using any of the three incomplete registries. The data of opioid use disorders in 2014, in New South Wales, Australia is retrieved from NDARC. The aim of study was to compare the estimator of capture-recapture models in the closed population for estimation the number of OUD.

Methods

Data for Analysis

The number of OUD was estimated using capture-recapture methodology based on three data sources (patient department, emergency department and the national death index in New South Wales, Australia 2014). The data were matched using the NDARC then arranged as shown in Table 1.

Table 1 Distribution of cases by combination of three data sources in 2014, New South Wales, Australia.

Sources			Number of Observations
Patient Department	**Emergency Department**	**National Death Index**	
0	0	0	Unknown Cases
1	0	0	87
0	1	0	407
0	0	1	15
1	1	0	657
1	0	1	0
0	1	1	11
1	1	1	54

Sources: The National Drug and Alcohol Research Centre (NDARC), University of New South Wales, Australia

Statistical Analysis

Capture-recapture models for closed populations were developed as useful classifications based on an observation model. The basic models of no heterogeneity were M_0 (no variation), M_t, M_b, and M_h. The classification models are a useful way of structuring capture-recapture models for closed populations in which the assumption of equal probability of capture is released [17-18]. These models were generalized by Otis *et al*. [21] to accommodate three types of heterogeneity in capture probability. Table 2 shows the model of variation among the capture probabilities: a temporal effect, heterogeneity between units and a behavioral effect. A temporal effect causes the capture probabilities to vary among capture occasions; heterogeneity causes the capture probabilities to vary among units. A behavioral effect means that the first capture changes the behavior of a unit, so the capture probability differs before and after the first capture.

Table 2 Summary of capture-recapture models for inspections.

Model	Assumptions on Detectability	Estimators
M_0	No variation exists.	M_0
M_t	Allows capture probabilities to vary by time.	M_t
M_h	Allows heterogeneous capture probabilities.	M_h Chao
		M_h Poisson 2
		M_h Darroch
		M_h Gamma 3.5
M_b	Allows behavioral responses to capture.	M_b

The models were: 1) M_0 which had one capture probability that was constant across individual and time; 2) M_t where the capture probability varied by sampling occasion only and all individuals in the population were equally catchable on any occasion and this model had a different capture probability for each sample occasion which was constant across individuals; and 3) M_b which had two capture probabilities: p if an individual has never been caught and c if an individual has been caught before, where this model allowed for the study to affect the behavior for the subjects and that if the capture process was a bad experience for the subject, we get a trap-shy response with $p > c$. Note that these responses can also be induced by the study design used. For example, if our sampling targets areas where subjects were previously caught, we are likely to observe a trap-happy response. Model M_h had a different capture probability for each individual in the population which was constant through time. The historically popular approach to estimate *N* for model M_h has been to use the non-parametric jackknife [22]. Model M_h assumed that each individual had its own unique probability that remained constant over samples. Choa's (or LB) models estimate a lower bound for the abundance (M_h Chao). The estimate obtained under M_h Chao is Chao's moment estimator [23]. Chao's lower bound models contain t-2 parameters, called eta parameters, for the heterogeneity. These parameters should theoretically be greater or equal to zero. The degrees of freedom of Chao's model increase when eta parameters are set to zero. Models for heterogeneity were defined as: Poisson $2(2^k-1)$, Darroch: $k^2/2$, and Gamma 3.5(-log(3.5+k)+log(3.5))[24], where k is the number of captures, Poisson and Gamma models with alternatives to the parameter defaults values of 2 and 3.5 can be fitted using the "colsedp" functions. Darroch's models for M_h was considered by Darroch [24-25].

The analyzed by capture-recapture methods to finding the best fitting model and estimating the OUD cases. The R program was used to find the number of OUD cases. The Rcapture package [26] was used to fit a model to a closed population dataset. The smallest Akaike Information Criteria (AIC) and degrees of freedom are useful tools to compare models and to assess the goodness of their fit.

Results

Table 1 shows the numbers of participants enrolled in 2014 for the dataset: Patient Department (n = 87), Emergency Department (n = 407), and National Death Index (n = 15). The overlapping of the three data sources was 54 cases.

The results of the capture-recapture method are shown in Table 3. The smallest AIC model was M_t. The estimates of the total number of OUDs from the 465 cases based on the AIC value were in ascending order: M_h Poisson 2, 433 (SE = 53.0); M_h Gamma 3.5, 351 (SE = 16.8);and M_b model, 503 (SE = 51.4); M_h Chao(LB) model, 666 (SE = 52.8); and M_0 model, 666 (SE = 52.8), respectively. The M_t model was the best estimated the number of OUD cases by using AIC.

Table 3 Estimates of population size OUD cases in New South Wales, Australia 2014.

Model	$\hat{N}$	SE	Deviance	df	AIC
M_0	666	52.8	391.991	5	425.454
M_t	465	26.8	14.500	3	51.962
M_h Chao (LB)	666	52.8	391.991	5	425.454
M_h Poisson 2	433	53.0	383.205	4	418.668
M_h Darroch	372	33.8	383.205	4	418.668
M_h Gamma 3.5	351	16.8	383.205	4	418.668
M_b	503	51.4	387.242	4	422.704

$\hat{N}$: Number Estimates of OUD Cases, SE: Standard Error,

df : Degree of Freedom, AIC: Akaike Information Criterion

Discussion and Conclusion

In the present study, we used three sources (patient department, emergency department, and the national death index in 2014), to compare the sources of variation in closed population models which used data form 2014 sourced from New South Wales, Australia. To the best of our knowledge, this has been the first study to apply capture-recapture analysis to estimate the number of OUD cases based on the sources of variation in closed population models. Moreover, the sources of variation were compared and the results from the best fitting model were determined based on the smallest AIC value.

The closed capture-recapture method is generally used to estimate hidden populations and commonly uses a log-linear regression model for estimation. In this study, we used an ecological model approach on the OUD data. Two commonly used forms have been proposed: the multiplicative (or log-linear) form and the logistic form [12]. The package Rcapture was used with Poisson regressions to estimate parameters in our capture-recapture models. A temporal effect causes the capture probabilities to vary among capture occasions, while heterogeneity causes the capture probabilities to vary among units. A behavioral effect indicates that that the first capture changes the behavior of a unit, so the capture probability differs before and after the first capture. There are sources of variation in closed population models (M_0, M_t, M_h, M_b). The AIC value was used for model selection with the smallest AIC being for the M_t, M_b, M_h, and M_0 models, respectively. The M_t model was estimated the population size OUD of 465 cases. The best fit the model was considered based on the smallest of AIC and the M_t model had the smallest of AIC [27]. With the M_t model more suitable OUD were estimated. This model should be able to apply to any setting with similar context. The method provided a simple, quick method to estimate the numbers of OUD.

Capture-recapture is a simple method to estimate the size of an unknown population size and is a useful and practical approach to estimate hard-to-reach populations, due to the assumptions and limitations of the method. In addition, ecological models are easy to use to provide an estimation of data sources. For estimation based on epidemiology data, the M_t model explained more than the other models as well as having the smallest AIC. Therefore, the M_t models are appropriate for estimation based on epidemiology or public health data. The recapture was easy to use and was appropriate for considering epidemiological data for the source effect or variation.

The limitation of this study was the three sources of data which we received based on linked data from the NDARC. Therefore, we focused the analysis on the overall population. The findings presented here show that the comparison of the sources of variation models and the number of unknown OUD cases. Although direct capture-recapture is used in few areas in New South Wales, datasets suitable for this purpose are likely to exist in most areas. Further studies of this nature would provide other sources of data with essential information and enable further exploration of OUD cases. Moreover, the data should be considered based on factors of individuals for analysis.

Acknowledgements

This study was supported by Kasetsart University Chalermprakiet Sakon Nakhon Province Campus. We acknowledge the National Drug and Alcohol Research Centre, UNSW, Australia for the data and for management data and thank the data team from the Biostatistics and Data Bases Program, the Kirby Institute, UNSW.

References

[1] Seber, G. A. F. (1982). *The estimation of animal abundance: and related parameters* (2nd Ed.). London: C. Griffin & Co., Ltd.,

[2]. Wohlin, C., Runeson, P., & Brantestam, J. (1995). An experimental evaluation of capture-recapture in software inspections. *Journal of Software: Testing, Verification and Reliability*, *5*(4), 213–232. DOI:10.1002/stvr.4370050403.

[3] Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y., & Chao, D. Y. (2001). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine, 20*(20), 3123–3157.

[4] Tsay, P.K., & Chao, A. (2001). Population size estimation for capture - recapture models with applications to epidemiological data. *Journal of Applied Statistics*, 28(1), 25–36.

[5] Kim, D.-S., Lee, M.-S., Kim, D.-H., Bae, J.-M., Shins, M.-H., Lee, C.-M., . . . Ahn, Y.-O. (1999). Evaluation of the completeness of cancer case ascertainment in the Seoul male cohort study: application of the capture-recapture method. *Journal of Epidemiology*, *9*(3), 146–154. DOI:10.2188/jea.9.146.

[6] Coumans, A. M., Cruyff, M., Van der Heijden, P. G. M., Wolf, J., & Schmeets, H. J. S. I. R. (2017). Estimating homelessness in the Netherlands using a capture-recapture approach. *Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement*, *130*(1), 189–212. DOI:10.1007/s11205-015-1171-7.

[7] Poorolajal, J., Haghdoost, A. A., Mahmoodi, M., Majdzadeh, R., Nasseri-Moghaddam, S., & Fotouhi, A. (2010). Capture-recapture method for assessing publication bias. *Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences, 15*(2), 107–115.

[8] Abeni, D. D., Brancato, G., & Perucci, C. A. (1994). Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. *Epidemiology, 5*(4), 410–414.

[9] Mastro, T. D., Kitayaporn, D., Weniger, B. G., Vanichseni, S., Laosunthorn, V., Uneklabh, T.,... Limpakarnjanarat, K. (1994). Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. *American Journal of Public Health*, *84*(7), 1094–1099.

[10] Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W., & Viwatwongkasem, C. (2004). Estimating the number of drug users in Bangkok 2001: A capture–recapture approach using repeated entries in one list. *European Journal of Epidemiology*, *19*(12), 1075. DOI:10. 1007/s10654-004-3006-8.

[11] Comiskey, C. M., & Barry, J. M. (2001). A capture recapture study of the prevalence and implications of opiate use in Dublin. *European Journal of Public Health*, *11*(2), 198–200. DOI:10.1093/eurpub/11. 2.198.

[12] Chao, A. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics, 6*(2), 158–175.

[13] Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete 2k contingency tables. *Biometrika*, *59*(3), 591–603. DOI:10.1093/biomet/59.3.591.

[14] Braeye, T., Verheagen, J., Mignon, A., Flipse, W., Pierard, D., Huygen, K., . . . Hens, N. (2016). Capture-recapture estimators in epidemiology with applications to pertussis and pneumococcal invasive disease surveillance. *PLOS ONE*, *11*(8), e0159832. DOI: 10.1371/ journal.pone.0159832.

[15] Van Hest, N. A. H., Grant, A. D., Smit, F., Story, A., & Richardus, J. H. (2008). Estimating infectious diseases incidence: validity of capture-recapture analysis and truncated models for incomplete count data. *Epidemiology and Infection*, *136*(1), 14–22.

[16] Choi, Y. H., & Comiskey, C. M. (2003). Methods for providing the first prevalence estimates of opiate use in Western Australia. *International Journal of Drug Policy*, *14*(4), 297–305. DOI: https://doi.org/10.1016/S0955-3959(03)00080-X.

[17] Pollock, K. H. (1974). *The assumption of equal catchability of animals in tag-recapture experiments.* Unpublished Ph.D. Dissertation, Cornell University.

[18] Sutherland, J. M. (2003). *Multi-list methods in closed populations with stratified or incomplete information.* Ph.D. Dissertation, Simon Fraser University.

[19] Kurtz, Z. T. (2014). *Local log-linear models for capture-recapture.* Ph.D. Dissertation, Carnegie Mellon University.

[20] Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y., & Chao, D. Y. (2001b). The applications of capture-recapture models to epidemiological data. *Statistics in Medicin*e, *20*(20), 3123–3157. DOI: https://doi.org/10.1002/sim.996.

[21] Otis, D. L., Burnham, K. P., White, G. C., & Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs*, (62), 3–135.

[22] Burnham, K. P., & Overton, W. S. (1979). Robust estimation of population size when capture probabilities vary among animals. *Ecology*, *60*(5), 927–936. DOI:10.2307/1936861.

[23] Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, *43*(4), 783–791.

[24] Darroch, J. N., Fienberg, S. E., Gary, F. V. G., & Junker, B. W. (1993). A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *Journal of the American Statistical Association*, *88*(423), 1137–1148. DOI:10.2307/2290811.

[25] Agresti, A. (1994). Simple capture-recapture models permitting unequal catchability and variable sampling effort. *Biometrics*, *50*(2), 494–500. DOI:10.2307/2533391.

[26] Baillargeon, S., & Rivest, L.-P. (2007). Rcapture: Loglinear models for capture-recapture in R. *Journal of Statistical Software*, *19*(5), 1–31. DOI:10.18637/jss.v019.i05.

[27] Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*(2), 101–108. DOI:https://doi.org/10.1016/j.tree.2003.10.013.

Appendix

The log-linear models for multiple capture-recapture experiments [26]

Model M_0

$$\log(\mu_\omega) = \log(N) + X_1 \log(p) + (t - X_1)\log(1-p)$$

where $\omega_j = 1$ if the unit is captured at the j^{th} occasion and, 0 if not $j=1,...,t$

Model M_t

$$\log(\mu_\omega) = X_0\gamma + \sum_{j=1}^{t} X_j B_j$$

where $p_i = 1$ is stand for that of occasion j, 0 if not and $j=1,...,t$

$$\gamma = \log\left\{N\prod(1-p_j)\right\}$$

$$\beta_j = \log\left\{\log(p_i/(1-p_i)\right\}, j=1,...,t.$$

Model M_b

$$\log(\mu_\omega) = X_0\log(N_p) + X_1\log(1-p) + X_2\left\{\frac{\log_c}{1-c}\right\} + (t - X_1 - 1)\log(1-c).$$

where $c =$ nuisance parameter

$\mu_j =$ the number of units that are first caught at the j^{th} capture occasion.

$\mu_j; j=1,...,t$

Note: For more information, see [26]

Model selection criteria

Akaike Information Criterion (AIC)

$$AIC = -2\ln[(\hat{\theta}_p)|Y] + 2p$$

Note: For more information, see [27]