

# Use of capture-recapture analysis in data sources

## Uso de análise de captura-recaptura em fontes de dados

Rotterdam, January 22nd 2008.

Dear Editors,

I have read with much interest your recent article on estimating the incidence of visceral leishmaniasis with a relatively simple two-source capture-recapture analysis through Chapman's Nearly Unbiased Estimator, which is only valid when data sources are independent, among other underlying assumptions.<sup>6</sup> Despite describing most assumptions in detail, regarding one of the most important assumptions, namely the source independence assumption, in the Discussion the authors only mention that they have "verified" that the data sources, as used in this study, are generally independent but do so without further explanation. They also state that it is not possible to determine the level of pair-wise interdependence (i.e. the probability of being observed in one register is affected by being –positive dependence- [or not being –negative dependence-] observed in another register) in a two-source capture-recapture analysis. I assume the authors mean that the data, as presented in the Venn-diagram in the figure, allow for investigation of possible pair-wise interdependencies by means of three-source capture-recapture analysis through log-linear modeling.<sup>1,3-5,8-10</sup> The use of death certificates as a third source of information allows for the probability of patients being reported to the notification and/or hospital admission register, as shown in the figure, although

intuitively it could create a negative interdependency with the other registers. The preferred log-linear model (Table), using the (lowest) Akaike Information Criterion, and incorporating interaction terms between SIM\*SIH and SINAN\*SIH, gives an estimate of 10,842 (95% confidence interval 10,120;11,754) patients with visceral leishmaniasis. Internal validity analysis by the two-source capture-recapture method shows that there is some relevant positive interaction between the SINAN and SIH registers as the estimate of 10,207 cases is slightly lower.<sup>2</sup> Two-source capture-recapture analysis using the SIM and SIH registers estimates 12,288 (95% CI 10,358;14,395) cases, indicating a relevant negative interaction between these registers, resulting in an overestimate. The two-source analysis, using the interaction term not identified as significant in the three-source analysis, namely between SINAN and SIM, estimates 10,691 cases, very close to the selected three-source capture-recapture estimate. An alternative population estimation model, the truncated binomial model, estimates 12,197 cases, but the high coefficient of variation of the data source probabilities (0,71) indicates violation of the equiprobability assumption underlying the binomial model, analytically resulting in overestimation.<sup>11</sup> The coverage of SINAN according to the three-source capture-recapture analysis (5,896/10,842 cases) is 54,3%, close to the author's estimate of 55%, as expected. In general one should be

**Table.** The eight possible three-source log-linear capture recapture models with the likelihood ratio ( $G^2$ ), P-value, degrees of freedom (df), Akaike Information Criterion (AIC) and estimated number of visceral leishmaniasis patients ( $N_{est}$ ) and the 95% confidence interval (95%CI).

Model	$G^2$	P-value	df	AIC	$N_{est}$	95% CI
Independent model	8,16	0.043	3	2,16	10,322	10,115 ; 10,547
SINAN*SIM interaction	7,40	0.025	2	3,40	10,292	10,077 ; 10,525
SINAN*SIH interaction	2,56	0.278	2	-1,44	11,130	10,417 ; 12,006
SIM*SIH interaction	2,91	0.234	2	-1,09	10,071	10,071 ; 10,503
SINAN*SIM, SINAN*SIH interactions	0,75	0.385	1	-1,25	11,923	10,550 ; 13,891
SIM*SIH, SINAN*SIH interactions	0,45	0.503	1	-1,55	10,842	10,120 ; 11,754
SIM*SIH, SINAN*SIM interactions	1,95	0.163	1	-0,05	10,246	10,031 ; 10,475
Saturated model	0		0	0	11,342	9,861 ; 13,712

cautious relying on two-source capture-recapture estimates alone for estimating the completeness of reporting of infectious diseases, as relevant (often positive) interdependencies between data sources used, such as notification, laboratory or hospital registers, could result in (untested and therefore unobserved) biased estimates.

This further, less mysterious, explanation of the preferred methodology for this study, i.e. three-source capture-recapture analysis, should be useful to the readers. Capture-recapture analysis, as a method to estimate infectious disease incidence and completeness of registration, is often not the cheap, quick, simple and reliable method as once advocated.<sup>7</sup>

Rob van Hest

Consultant Tuberculosis Control Physician/  
Epidemiologist Tuberculosis Control Section  
Municipal Public Health Service  
Rotterdam-Rijnmond, Netherlands  
vanhestr@ggd.rotterdam.nl

## REFERENCES

1. Fienberg SE. The multiple-recapture census for closed populations and the  $2^k$  incomplete contingency table. *Biometrika*. 1972;59:591-603.
2. Hook EB, Regal RR. Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. *Am J Epidemiol*. 1995;142(9 suppl):S48-52.
3. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev*. 1995;17(2):43-63.
4. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record estimation I: History and theoretical development. *Am J Epidemiol*. 1995;142(10):1047-58.
5. International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record estimation II: Applications in human diseases. *Am J Epidemiol*. 1995;142(10):1059-68.
6. Maia-Elkhoury ANS, Carmo EH, Sousa-Gomes ML, Mota E. Análise dos registros de leishmaniose visceral pelo método de captura-recaptura. *Rev Saude Publica*. 2007;41(6):931-7.
7. Van Hest NA. Capture-recapture methods in surveillance of tuberculosis and other infectious diseases [thesis]. Rotterdam: Erasmus University; 2007.
8. Van Hest NA, Smit F, Verhave JP. Underreporting malaria notification in the Netherlands: results from a capture-recapture study. *Epidemiol Infect*. 2002;129(2):371-7.
9. Van Hest NA, Smit F, Baars HW, Vries G de, Haas P de, Westenenk PJ, et al. Completeness of notification of tuberculosis in the Netherlands: how reliable is record-linkage and capture-recapture analysis? *Epidemiol Infect*. 2007;135(6):1021-9.
10. Van Hest NA, Hoebe CJ, Den Boer JW, Vermunt JK, IJzerman EP, Boersma WG, et al. Incidence and completeness of notification of Legionnaires' disease in the Netherlands: covariate capture-recapture analysis acknowledging geographical differences. *Epidemiol Infect*. 2007; Published on-line 22 June 2007; doi: 10.1017/S0950268807008977.
11. Van Hest NA, Grant D, Smit F, Story A, Richardus JH. Estimating infectious diseases incidence: validity of capture-recapture analysis and truncated models for incomplete count data. *Epidemiol Infect*. 2008;136(1):14-22.

## AUTHOR'S REPLY

Brasilia, February 25th 2008.

Dear Editors,

Capture-recapture analysis has been applied in epidemiology and disseminated mainly since 1990s with increasing number of articles published on different diseases using this same approach. In Brazil, its application is still punctual and quite limited to health care services.

Some authors find this methodology a less expensive and simpler alternative for disease monitoring,<sup>1,4</sup> but the infrastructure of health care services and availability of reliable information systems should be considered

for facilitating its application. In a study<sup>5</sup> conducted in Brazil to assess underreporting of AIDS cases in the city of Belo Horizonte, state of Minas Gerais (Southeastern Brazil), the author concludes that "capture-recapture analysis is a perfectly feasible approach when based on information systems available in public health services in Brazil". She also states it is "relatively easy, simple, and not costly especially when compared to active surveillance".<sup>5</sup> In particular on the application of this approach in the study of visceral leishmaniasis, the authors restate that capture-recapture analysis is a cheap and simple alternative to be applied by surveillance systems in Brazil, mainly because these information systems are widely spread in Brazilian municipalities and are used and managed by local and state health departments and the Brazilian Ministry of Health.

It is important to stress the need to take into account the assumptions required for the application of capture-recapture analysis (close population, effective link of registers, diagnosis validity, and source independence). When these assumptions are violated, i.e., when any assumption cannot be fully met, information should be detailed and discussed. Because of this condition, many investigators question the application of capture-recapture analysis in epidemiological studies.

In the study of registers of visceral leishmaniasis information, the authors proposed a combination of pair-wise sources using a simple approach for the intended estimates since data sources analyzed were the same as used in studies of other diseases in Brazil<sup>5</sup> and due to easiness of its application in health care services. They also pointed out that as they chose to study pair-wise

sources, they did not determine the level of source interdependence. But the authors are aware that log-linear models and Bernoulli census are approaches for the treatment of source interdependence.<sup>2,3</sup> In his letter to the editor, Van Hest presents a demonstration of source interdependence based on a log-linear model. It should be noted though that an analysis without the adequacy of data sources can lead to inaccurate conclusions.

It should also be underlined that the article "Analysis of visceral leishmaniasis reports by the capture-recapture method" had no intention to either exhaust this subject or claim that capture-recapture analysis is the best choice for disease monitoring by epidemiological surveillance systems. Yet the article provides relevant information for leishmaniasis surveillance in Brazil and promotes the debate on the application of capture-recapture analysis in epidemiology.

Ana Nilce Silveira Maia Elkhoury

Departamento de Vigilância Epidemiológica,  
Secretaria de Vigilância em Saúde do  
Ministério da Saúde  
Brasília, DF, Brasil  
Ana.Elkhoury@saude.gov.br

## REFERENCES

1. Corrão G, Bagnardi V, Vittadini G, Favilli S. Capture-recapture methods to size alcohol related problems in a population. *J Epidemiol Community Health*. 2000;54(8):603-10.
2. Hook EB, Albright SG, Cross PK. Use of Bernoulli census and log-linear methods for estimating the prevalence of spina bifida in live births and the completeness of vital reports in New York State. *Am J Epidemiol*. 1980;112(6):750-8.
3. Hook EB, Reagal RR. Capture-recapture methods in epidemiology: Methods and Limitations. *Epidemiol Rev*. 1995;17(2):243-64.
4. Laporte RE. Assessing the human condition: Capture-recapture techniques. *BMJ*. 1994;308(6920):5-6.
5. Oliveira MTC A subnotificação de casos de Aids em Belo Horizonte, Minas Gerais: uma aplicação da técnica de captura-recaptura [dissertação de mestrado]. Belo Horizonte: Faculdade de Medicina da Universidade Federal de Minas Gerais 2000.