Use of capture-recapture analysis in data sources

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.6 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.1,3-5,8-10 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.2 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.11 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>
<thead>
<tr>
<th>Model</th>
<th>( G^2 )</th>
<th>P-value</th>
<th>df</th>
<th>AIC</th>
<th>( N_{est} )</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent model</td>
<td>8.16</td>
<td>0.043</td>
<td>3</td>
<td>2.16</td>
<td>10,322</td>
<td>10,115 ; 10,547</td>
</tr>
<tr>
<td>SINAN*SIM interaction</td>
<td>7.40</td>
<td>0.025</td>
<td>2</td>
<td>3.40</td>
<td>10,292</td>
<td>10,077 ; 10,525</td>
</tr>
<tr>
<td>SINAN*SIH interaction</td>
<td>2.56</td>
<td>0.278</td>
<td>2</td>
<td>-1.44</td>
<td>11,130</td>
<td>10,417 ; 12,006</td>
</tr>
<tr>
<td>SIM*SIH interaction</td>
<td>2.91</td>
<td>0.234</td>
<td>2</td>
<td>-1.09</td>
<td>10,071</td>
<td>10,071 ; 10,503</td>
</tr>
<tr>
<td>SINAN<em>SIM, SINAN</em>SIH interactions</td>
<td>0.75</td>
<td>0.385</td>
<td>1</td>
<td>-1.25</td>
<td>11,923</td>
<td>10,550 ; 13,891</td>
</tr>
<tr>
<td>SIM<em>SIH, SINAN</em>SIH interactions</td>
<td>0.45</td>
<td>0.503</td>
<td>1</td>
<td>-1.55</td>
<td>10,842</td>
<td>10,120 ; 11,754</td>
</tr>
<tr>
<td>SIM<em>SIH, SINAN</em>SIM interactions</td>
<td>1.95</td>
<td>0.163</td>
<td>1</td>
<td>-0.05</td>
<td>10,246</td>
<td>10,031 ; 10,475</td>
</tr>
<tr>
<td>Saturated model</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11,342</td>
<td>9,861 ; 13,712</td>
<td></td>
</tr>
</tbody>
</table>
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.7

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

REFERENCES

AUTHOR’S REPLY

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, but the infrastructure of health care services and availability of reliable information systems should be considered for facilitating its application. In a study 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”. 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 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. 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


