Capture-recapture method for estimating misclassification errors: application to the measurement of vaccine efficacy in randomized controlled trials

Francois Simondon and Hayet Khodja

Background

The measure of efficacy is optimally performed by randomized controlled trials. However, low specificity of the judgement criteria is known to bias toward lower estimation, while low sensitivity increases the required sample size. A common technique for ensuring good specificity without a drop in sensitivity is to use several diagnostic tests in parallel, with each of them being specific. This approach is similar to the more general situation of case-counting from multiple data sources, and this paper explores the application of the capture-recapture method for the analysis of the estimates of efficacy.

Method

An illustration of this application is derived from a study on the efficacy of pertussis vaccines where the outcome was based on >21 days of cough confirmed by at least one of three criteria performed independently for each subject: bacteriology, serology, or epidemiological link. Log-linear methods were applied to these data considered as three sources of information.

Results

The best model considered the three simple effects and an interaction term between bacteriology and epidemiological linkage. Among the 801 children experiencing >21 days of cough, it was estimated that 93 cases were missed, leading to a corrected total of 413 confirmed cases. The relative vaccine efficacy estimated from the same model was 1.50 (95% confidence interval: 1.24–1.82), similar to the crude estimate of 1.59 and confirming better protection afforded by one of the two vaccines.

Conclusion

This method allows supporting analysis to interpret primary estimates of vaccine efficacy.

Keywords

Efficacy, misclassification, capture-recapture, log-linear models, pertussis, vaccine

Accepted 2 June 1998

Randomized controlled trials (RCT) have been acknowledged as the methodology of reference for measuring the efficacy of a treatment or of an intervention. The experimental structure allows for the comparability of the study groups, the unbiased measure of the outcome and predetermined power, and thus for causal interpretation. The exposure factor (the treatment) is usually well defined and the investigator can determine to which group the subject belongs. Classification errors are therefore minimized, or are known, as in intent-to-treat analysis. However, the diagnostic bias (misclassification of the judgement criteria) is related to the quality of the case definition. It is well known that low sensitivity increases the required sample size, while low specificity of the judgement criteria tends to bias toward lower estimation of the efficacy of the treatment. These so-called 'dilution effects' may be of particular importance in equivalence studies, leading to false conclusions of equivalence of treatments. Although use of a single judgement criteria has been recommended, a common method for minimizing misclassification of the disease status is to repeat the test, or more often to simultaneously consider different diagnostic tests of the disease, thus using a composite judgement criteria. The basic idea is to obtain information on different diagnostic tests from the same subject for the same disease, with each of the tests being specific, thus improving sensitivity while at the same time preserving specificity.
This approach is similar to the more general situation of case-counting from multiple and incomplete data. The objective of the present study was to discuss the application of the capture-recapture method for interpreting estimates derived from such RCT, where the different components of diagnostic criteria are considered as multiple sources of information. An illustration is given by the application of this method to data from the Senegal Pertussis Trial.

Materials and Methods

The Senegal Pertussis Trial

The development of acellular pertussis vaccines (AC) has followed concern about the safety and protective efficacy of an AC vaccine with reference to the WC vaccine routinely used in Senegal, and widely used in other developing and European countries. The methodology and results of the study have been described elsewhere. Briefly, our study began in 1990 in Niakhar, a rural area of Senegal of approximately 28,000 inhabitants under demographic and epidemiological surveillance since 1983. In this area, the residential unit is the compound where an extended family lives in one or more households. The median number of people living in a compound is 20. Of the 30 villages in the study area, each compound has been visited weekly since 1988 by a field worker who collects demographic and epidemiological information on migration, marriages, births, mortality, vaccinations, breastfeeding, and infections such as measles, diarrhea, and cough. Data were recorded in a centralized database and every person registered was given a unique identification number. The study population for the pertussis vaccine trial was comprised of eligible infants, i.e. those born between February 1990 and April 1994, to mothers residing in the study area, and whose parents gave their consent for participation. Inclusion criteria consisted of no previous history of pertussis, no previous pertussis vaccination and no severe or chronic disease.

Based on the central database, children due to be vaccinated were visited by a field worker the week before a monthly vaccination session. Before the first dose, enrolled infants were randomly assigned to one of the two vaccine groups. Active surveillance of pertussis cases of all children<15 years old living in the Niakhar area was initiated in 1988, i.e. before the first infants eligible for the study were born. Field workers visited all compounds weekly in search of children coughing for longer than 7 days. Every coughing child<15 years was examined weekly, until the end of all cough illnesses in the compound, by a physician blind to vaccination status. Culture specimens were collected at the first visit and one week later. Blood specimens for serology were collected at the first visit and approximately 6 weeks later.

Bacteriological confirmation was based on identification of colonies by morphology, Gram staining, oxidase and urease reactions and direct immunofluorescence. Serological confirmation was based on a twofold rise in antibody titres between acute and convalescent serum by enzyme-linked immunosorbent assay.

The case definition of pertussis included all children with ≥21 days of cough and confirmed by either a positive bacteriological result, a positive serological result or epidemiological linkage (Eplink). Eplink was defined as a child who had been in contact, within a compound, with another culture-confirmed child, and who started coughing within 28 days before or after the onset of illness in the culture-confirmed child. This case definition was very similar to the pertussis case definition for vaccine trials endorsed by the World Health Organization after the trial had begun. The difference concerned the type of cough, which was restricted to paroxysmal cough in the latter definition. The analysis considered all cases occurring ≥28 days after the third dose. For each child, surveillance ended either at the onset of pertussis or upon referral of investigation, additional pertussis immunization, emigration, or death. All surveillance ended 31 December 1994.

A comparable number of infants received a first dose of vaccine: 2089 in the WC group and 2092 in the AC group. A total of 3619 children received either the AC (1847) or the WC (1772) vaccine at 2, 4 and 6 months of age. Comparability between study groups for several background factors was evaluated at 28 days after the third dose, and no significant differences were found. In all, 320 pertussis cases were identified by the protocol definition of cases; 197 in the AC group and 123 in the WC one.

The crude relative efficacy, based on the ratio of pertussis incidence density of the AC group compared to the WC group (RRAC/WC) was 1.59 (95% confidence interval [CI]: 1.27–1.99), indicating lower protection provided by AC.

Capture-recapture methodology

The capture-recapture method, derived from its original application in the field of animal ecology, has been proposed and used in epidemiology to evaluate the completeness of registries or to correct case counting. It is based on the analysis of multiple and incomplete data sources. Four distinct conditions are usually required: (1) the sources should be independent; (2) the population is closed, i.e. no births, no deaths, no migrants, so the probability of identification is constant during the study; (3) the probability of identification within any source is equal for all individuals; (4) all identified individuals belong to the study population. The introduction of log-linear methods allowed adjustment for source dependencies when data are available from three or more sources. However, these conditions are still very restrictive and limit the application of this method. In our proposed application, where the different components of the diagnostic criteria were considered as multiple sources of information, the conditions were met. The independence of the sources, which is the major practical condition, was expected, since the diagnostic tests were systematically and independently performed for each suspected case. The population was considered as closed because all diagnostic tests were performed at the same time.

The analysis was based on the use of the log-linear model with BMDP 4E. The goodness-of-fit of the log-linear model was tested by the likelihood ratio statistic (G²) and the Pearson goodness-of-fit statistic. For a particular log-linear model, if the P-values were higher than the level of significance, then the model was considered to satisfactorily fit the data. Finally, the model selected was the simplest among those with satisfactory fit. A stepwise method was used to confirm the choice. The estimates of CI for the number of missed cases were based on the likelihood ratio statistic with application of iterative processes. The CI for the corrected RRAC/WC was calculated.
Results

Results are first presented concerning all participating children, and thereafter by vaccine group.

Among the 801 children experiencing ≥21 days of cough, 320 were defined as confirmed pertussis cases (8.8% of the population). They were confirmed by the combination of the three different sources (Table 1). Not all cases were detected in each of the three sources.

The choice of the model seemed unambiguous (Table 2): only three models had satisfactory fit at the 0.05 level of significance (S, BE; BS, BE; SE, BS, BE). The simplest one was the S, BE model with the three simple effects: bacteriology, serology, Epilink, and an interaction term between bacteriology and Epilink. Using this model, 93 infants were missed, leading to a corrected total of 413 affected children (95% CI: 375–466). Subsequently, the corrected proportion of affected infants, calculated from the size of total vaccinated infants (Table 3), increased to 11.3% (95% CI: 10.3–12.8%).

It was also possible to compute the sensitivity of the diagnostic tests against the capture-recapture corrected estimate. As an example, the sensitivity of the bacteriologic test against serology was 28.3% and against Epilink it was 26.5%. It decreased to 21.6% against the capture-recapture estimate.

Discussion

Several attempts at correcting estimates for misclassification effects based on different methodologies have been reported.\textsuperscript{18–21} As far as we know, the capture-recapture method has not been previously applied. The conditions of application of this method were satisfied in relation to the experimental design of RCT. Each of the three tests was performed independently for every suspect case. However, the interaction term between bacteriology and Epilink retained in the best fitted model express some degree of dependency between these two sources in the population, and account for it. The selection of the best model is a major question and in our example, the use of the Akaike information criteria, which has been shown to be preferable,\textsuperscript{12} has been tested and led to similar results.

A substantial number of missed cases (n = 93) was ascertained by the capture-recapture method, leading to a total of 413 confirmed cases. This estimate seems plausible as it is still lower than the potential number of cases found by the surveillance (n = 801). These missed cases are not identifiable, and consequently they cannot be included for subsequent analysis, such as adjustment for confounding. This under-ascertainment is biologically plausible, particularly among vaccinated children. Culture of the organism lack sensitivity,\textsuperscript{10} as well as serology\textsuperscript{23} and Epilink, which depends on bacteriology of vaccinated contacts.

---

Table 1 Results of three different tests (+ denotes positive response, − denotes negative response) for 320 infants

<table>
<thead>
<tr>
<th>Epilink</th>
<th>Bacteriology</th>
<th>Serology</th>
<th>Number of infants</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>26</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>−</td>
<td>46</td>
</tr>
<tr>
<td>+</td>
<td>−</td>
<td>+</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 2 Choice of the three-way log-linear model in which indices are bacteriology (B), serology (S) and Epilink (E)

<table>
<thead>
<tr>
<th>Model</th>
<th>d.f.</th>
<th>G²</th>
<th>P-value</th>
<th>$\chi^2$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>5</td>
<td>138.66</td>
<td>0.0000</td>
<td>164.25</td>
<td>0.0000</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
<td>194.53</td>
<td>0.0000</td>
<td>186.64</td>
<td>0.0000</td>
</tr>
<tr>
<td>E</td>
<td>5</td>
<td>139.00</td>
<td>0.0000</td>
<td>178.42</td>
<td>0.0000</td>
</tr>
<tr>
<td>B, S</td>
<td>4</td>
<td>46.78</td>
<td>0.0000</td>
<td>49.19</td>
<td>0.0000</td>
</tr>
<tr>
<td>B, E</td>
<td>4</td>
<td>109.20</td>
<td>0.0000</td>
<td>115.17</td>
<td>0.0000</td>
</tr>
<tr>
<td>S, E</td>
<td>4</td>
<td>17.15</td>
<td>0.0000</td>
<td>17.81</td>
<td>0.0000</td>
</tr>
<tr>
<td>B, S, E</td>
<td>3</td>
<td>13.94</td>
<td>0.0000</td>
<td>15.62</td>
<td>0.0035</td>
</tr>
<tr>
<td>BS</td>
<td>3</td>
<td>22.50</td>
<td>0.0000</td>
<td>36.49</td>
<td>0.0000</td>
</tr>
<tr>
<td>BE</td>
<td>3</td>
<td>76.25</td>
<td>0.0000</td>
<td>72.23</td>
<td>0.0000</td>
</tr>
<tr>
<td>SE</td>
<td>3</td>
<td>91.89</td>
<td>0.0000</td>
<td>85.52</td>
<td>0.0000</td>
</tr>
<tr>
<td>S, SE</td>
<td>2</td>
<td>9.91</td>
<td>0.0071</td>
<td>9.74</td>
<td>0.0077</td>
</tr>
<tr>
<td>S, BE</td>
<td>2</td>
<td>5.23</td>
<td>0.0733</td>
<td>5.47</td>
<td>0.0648</td>
</tr>
<tr>
<td>B, BS</td>
<td>2</td>
<td>11.49</td>
<td>0.0032</td>
<td>11.11</td>
<td>0.0039</td>
</tr>
<tr>
<td>BS, BE</td>
<td>1</td>
<td>1.02</td>
<td>0.3121</td>
<td>0.97</td>
<td>0.3236</td>
</tr>
<tr>
<td>BS</td>
<td>1</td>
<td>3.12</td>
<td>0.0778</td>
<td>3.41</td>
<td>0.0320</td>
</tr>
<tr>
<td>SE</td>
<td>1</td>
<td>8.83</td>
<td>0.0030</td>
<td>7.94</td>
<td>0.0048</td>
</tr>
<tr>
<td>SE, BS</td>
<td>0</td>
<td>0.00</td>
<td>1.0000</td>
<td>0.00</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

\(\text{d.f.}\): Degree of freedom. 

\(\text{P-value}\): Likelihood ratio statistic. 

\(\text{P-value}\): The higher order effect of the hierarchical model contains all lower-order effects that are included in the higher effect.

---

Table 3 Estimates of the missed infants for the total population and for each group. Derived estimates of the per cent failure rates of vaccines

<table>
<thead>
<tr>
<th>Subjects affected</th>
<th>Total population</th>
<th>Whole cell</th>
<th>Acellular</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>123</td>
<td>197</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>41</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>55–146</td>
<td>18–85</td>
<td>27–92</td>
<td></td>
</tr>
<tr>
<td>413</td>
<td>164</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>375–466</td>
<td>141–208</td>
<td>224–289</td>
<td></td>
</tr>
<tr>
<td>10.3–12.8</td>
<td>7.9–11.7</td>
<td>12.1–15.6</td>
<td></td>
</tr>
</tbody>
</table>

\% failure rates: 11.3, 9.2, 13.4

95% CI: 10.3–12.8

11.3% (95% CI: 10.3–12.8%).

13.4% (95% CI: 10.3–12.8%).
An implicit assumption was that each of the diagnostic tests was 100% specific. This may not necessarily be true. However, this assumption also belongs to the classical analysis and to other correction attempts.24

The sensitivity of each of the sources can be computed against the capture-recapture estimate. It can be shown that derived sensitivity, as based on the assumption of 100% specificity, is underestimated.24

The point estimate of relative efficacy was nearly unchanged. This is not surprising, as the same model fit best in both vaccine groups. This suggests that there is no differential diagnostic bias between vaccine groups, and is an important finding for the capture-recapture estimate. It can be shown that derived sensitivity, as based on the assumption of 100% specificity, is underestimated.24

The application of the capture-recapture method can be recommended as supporting analysis to interpret primary estimates of vaccine efficacy with respect to misclassification. When relevant, it should be taken into account in the preparation of trials, in order to ensure that several diagnostic tests are performed systematically and independently for the same subject.23

Acknowledgements
The authors are indebted to L Papoz and K Simondon, who commented on our paper before submission.

References
International Journal of Epidemiology

Contents

Leading Articles

1 Traditional reviews, meta-analyses and pooled analyses in epidemiology
M Blettner, W Sauerbrei, B Schienhölzer, T Scheuchepflug and C Friedenreich

10 Evaluation designs for adequacy, plausibility and probability of public health programme performance and impact
JPH Habicht, CG Victoria and JP Vaughan

Original Articles

19 Cancer mortality in Russia and Ukraine: validity, competing risks and cohort effects
VM Shkolnikov, M McKee, J Vallin, E Aksel, D Leon, L Chenet and F Meslé

30 Alcohol consumption as a major risk factor for the rise in liver cancer mortality rates in Japanese men
K Makimoto and S Higuchi

35 A mobile unit: an effective service for cervical cancer screening among rural Thai women
W Swaddiwudhipong, C Chaovakratipong, P Nguntra, P Mahasakpan, Y Tatip and C Boonnak

40 Trends in mortality from cutaneous malignant melanoma in Belgium
L Bleyen, D De Bacquer, K Myny, L Brochez, JM Naeyaert and G De Backer

46 Socioeconomic status and cardiovascular risk factors in the Czech Republic
M Bobak, C Hertzman, Z Skodova and M Marmot

53 Physical activity assessment in population surveys: can it really be simplified?
F Domínguez-Berjón, C Borrell, M Nebot and A Plasència

58 Inequalities in mortality according to educational level in two large Southern European cities
C Borrell, R Regidor, L-C Arias, P Navarro, R Pulipinós, V Domínguez and A Plasència

64 Classification differences and maternal mortality: a European study
B Salanave, M-H Bouvier-Colle, N Vannoux, S Alexander, A Maclaran and the MOMS Group

70 Maternal mortality in Guinea-Bissau: the use of verbal autopsy in a multi-ethnic population
L Høj, J Stensballe and P Aaby

77 The effects of wine and tobacco consumption on cognitive performance in the elderly: a longitudinal study of relative risk
D Leibovici, K Ritchie, B Ledésert and J Touchon

82 The comparison of mixed distribution analysis with a three-criteria model as a method for estimating the prevalence of iron deficiency anaemia in Costa Rican children aged 23 months
JH Cohen and JD Haas

continued on outside back cover