Defining Risk Groups to Yellow Fever Vaccine-Associated Viscerotropic Disease in the Absence of Denominator Data

Stephen J. Seligman,* Joel E. Cohen, Yuval Itan, Jean-Laurent Casanova, and John C. Pezzullo

INTRODUCTION

Before 2001, the live yellow fever virus vaccine was considered the world’s safest live virus vaccine.1 In that year reports started to appear of rare multisystemic, frequently fatal, reactions that are called yellow fever vaccine-associated viscerotropic disease (YEL-AVD) (for review see Reference 2). On the basis of four of the then known 23 cases, the first risk group for YEL-AVD identified was individuals thymectomized because of thymoma.3 A second risk group, identified using age— but not the combination of age-gender-specific denominator data, consists of people ≥ 60 years of age.4,5 Subsequently, it was recognized that the elderly with YEL-AVD were predominantly males.2 Another risk group is women in their prime child-bearing years.6 Other suspected risk groups include infants and children < 12 years of age, women between 39 and 49 years of age with systemic lupus erythematosus (SLE), and patients with Addison’s disease or other autoimmune disorders.7 (Seligman SJ, unpublished data). This study was initiated to assist in the documentation of other risk groups.

Odds ratios (ORs) are frequently used to identify risk groups, and to quantify the magnitude of the risk. Their calculation typically requires knowledge of the numbers of subjects who had an outcome and had or did not have the predisposing factor (numerator data) and the numbers of subjects in both categories who remained unaffected (denominator data). Consequently, if denominator data are not available, ORs cannot be calculated by the traditional method. The current report illustrates how estimates of the fraction of the exposed population with the suspected risk factor who remain well can be used instead to calculate the OR and confidence interval (CI), how the effect of uncertainty in the estimate of that fraction can be evaluated, and how P values can be calculated using the CI.

METHODS

Data sources. Tom Monath supplied an Excel file originally prepared by the Centers for Disease Control and Prevention (CDC) that contained data on 65 cases suspected of having YEL-AVD. The cases came from endemic regions in South America and from prospective travelers to yellow-fever prone areas in South America and Africa. No cases originated in Africa. Three of the cases in which the age and gender of the vaccinee were not known were excluded and two additional cases in prospective travelers (one from New Zealand and one from Peru [abstract presented in a poster by Turpo and others at the ASTMH meetings in Atlanta, GA, Nov. 14, 2012]) that came to our attention were added for a total of 64 cases. The five suspected cases of YEL-AVD listed in a report from Africa were not included because none was substantiated by virological or serological evidence.8 Cases were vaccinated in the years 1973–2011 and ranged in age from 10 months to 79 years. Suspected risk groups were identified by clusters of cases stratified by age, gender, and outcome.

Calculation of factor-specific odds ratios. A “case” is defined as a subject who becomes ill. The OR is defined as

\[
OR = \frac{N_1/D_1}{N_2/D_2},
\]

where \(N_1\) is the number of cases with the risk factor, \(N_2\) is the number of cases without the risk factor, \(D_1\) the number of exposed (e.g., vaccinated) subjects with the risk factor who remained well and \(D_2\) the number of exposed subjects without the risk factor who remained well. Denote the total number of exposed subjects as \(NP\). By definition

\[
NP = N_1 + N_2 + D_1 + D_2
\]

If \(D = D_1 + D_2\) is the total number exposed who did not become ill and \(f\) is the fraction of exposed subjects with the risk factor in the exposed population who did not become ill, then

\[
D_1 = f \times D
\]

and

\[
D_2 = D - D_1
\]
Using Eq. (2) and substituting for $D_1$

$$D_2 = D - f \times D = (1 - f) \times D.$$  \hspace{1cm} (3)

Substituting in Eq. (1) using Eq. (2) for $D_1$ and Eq. (3) for $D_2$ gives

$$OR = \frac{N_1}{N_2} \times \frac{1}{1 - f} / f.$$  \hspace{1cm} (4)

Accordingly, assuming that $N_1$ and $N_2$ are known, if the fraction $f$ of the exposed population with the suspected risk factor who remained well can be estimated, the OR can be calculated without knowing the denominators $D_1$ and $D_2$.

**Calculation of a 95% CI.** Under the “central limit theorem” assumption, the standard error of the natural logarithm of the OR ($SE_{\text{LN}(OR)}$) can be approximated by

$$SE_{\text{LN}(OR)} = \sqrt{\frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{D_1} + \frac{1}{D_2}}.$$

The 95% confidence limits (CL) around the logarithm of the OR are the logarithms of the OR ± 1.96 times the $SE_{\text{LN}(OR)}$, and the antilogarithms of these logarithmic limits are the 95% CLs around the OR itself. (The anti-natural-logarithm is simply the exponential function, $e^x$). The values for $D_1$ and $D_2$ depend upon the total number in the population ($N_P$). In most situations, $N_P \gg D_1$ and $N_P \approx D_2$. Even for very rare risk factors and a sufficiently large exposed population, the CLs approach limiting or “asymptotic” values that depend only on the size of $N_1$ and $N_2$ (Figure 1).

Accordingly, the lower and upper 95% CLs ($CL_L$ and $CL_U$, respectively) of the OR are given by

$$95\% CL_L = e^{\text{LN}(OR) - 1.96 \times SE}$$

and

$$95\% CL_U = e^{\text{LN}(OR) + 1.96 \times SE}.$$

**RESULTS**

**Calculation of ORs.** The age and gender of patients with YEL-AVD are known in 64 cases. Calculations of OR using Eq. (4) and estimated values of $f$ are shown (Table 1). With the exception of pernicious anemia, all of the suspected risk groups have ORs > 1, varying from 2.8 to 140,000.

**Calculation of CI.** The data for suspected risk groups are shown (Table 2). A vaccinated population with a hypothetical size of 3 million was chosen because even with rare frequency of the risk group, estimates of the 95% $CL_L$ and $CL_U$ closely approach the limiting values.
for populations over 3 million (Figure 1). The total number of vaccinees after the occurrence of YEL-AVD was recognized is over 500 million; however, case ascertainment particularly in Africa is frequently uncertain. The assertions that the ORs are particularly in Africa is frequently uncertain. The assertion to thymoma and SLE where the N1 values are 4 and 3, concerns also affect the calculations for thymectomy second-the calculated CLs for that variable are unreliable. Similar considerations also affect the calculations for thymectomy secondary to thymoma, and SLE (Table 1) are supported by the CLs appreciating > 1.

The data for pernicious anemia do not support its inclusion as a risk group. However, because the SE approximation is inaccurate for numbers ≤ 4, and N1 = 1 for pernicious anemia, the calculated CLs for that variable are unreliable. Similar concerns also affect the calculations for thymectomy secondary to thymoma and SLE where the N1 values are 4 and 3, respectively.

Discrimination analysis of prevalence estimates. The calculation results are listed in Table 3. Comparing the f values in Tables 1 with those in Table 3, for men ≥ 56 years of age, the prevalence increases 4.8-fold, in women 19–34 years of age, patients with autoimmune disease, people thymectomized because of thymoma, and SLE (Table 1) are supported by the CLs appreciating > 1. These results indicate the limits to which the change in prevalence is consistent with a statistically significant result as judged by being associated with 95% CLL values ≥ 1.

Calculations of P values. Results of the P value calculations are shown (Table 4). Although the precise P values calculated by this method are unreliable for values < 0.0001, P values for men ≥ 56 years of age, women between 19 and 34 years of age, people with autoimmune diseases, persons with thymectomy for thymoma, and SLE are all highly significant. The P values for pernicious anemia do not support its inclusion as a risk group.

DISCUSSION

Given an estimate of the frequency of a risk group in the exposed population who remained well, a frequency usually close to the frequency in the total population, calculations of ORs, CIs, and P values can easily be made (Box 1). Finding the prevalence that would reduce the 95% CLL to 1 illustrates the extent to which the prevalence may be increased with maintenance of a statistically significant result. Risk factor-specific denominator data are not necessary to obtain these estimates. As illustrated in the examples of the rare viscerotropic reactions following yellow fever vaccine, if the prevalence of the risk group in the total population that did not become ill can be estimated, these methods facilitate identification of risk groups when denominator data are not available.

The calculations confirm that people thymectomized because of thymomas, men ≥ 56 years of age, and women between 19 and 34 years of age are at increased risk for developing YEL-AVD and support the inclusion of patients with autoimmune disease and women 39 to 49 with SLE as being in a risk group. The data do not support the inclusion of patients with pernicious anemia in a risk group. Reasons for the susceptibility to YEL-AVD are unclear. Initially it was suspected that the vaccine virus may have regained virulence, but extensive sequencing of virus isolated from cases does not support this hypothesis. The current analysis indicates that a variety of autoimmune disease are associated with an increased risk (Tables 1–4). In a 64-year-old man who survived, limited sequencing of a few genes resulted in the identification of polymorphisms

### Table 2
Calculations of the 95% CI for a hypothetical vaccinated population of 3 million (NP)

<table>
<thead>
<tr>
<th>Risk group</th>
<th>D = NP – (N1 + N2)</th>
<th>D1 = D × f</th>
<th>D2 = D × (1 − f)</th>
<th>LN(OR)</th>
<th>SELN(OR)</th>
<th>95% CI = e^LN(OR) ± 1.96 × SELN(OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men ≥ 56</td>
<td>2,999,936</td>
<td>183,596</td>
<td>2,816,340</td>
<td>2.4</td>
<td>0.25</td>
<td>6.4–17</td>
</tr>
<tr>
<td>Women 19–34</td>
<td>2,999,936</td>
<td>275,236</td>
<td>2,701,200</td>
<td>1.0</td>
<td>0.30</td>
<td>1.6–4.9</td>
</tr>
<tr>
<td>Autoimmune disease</td>
<td>2,999,936</td>
<td>89,998</td>
<td>2,900,938</td>
<td>1.8</td>
<td>0.34</td>
<td>3.0–12</td>
</tr>
<tr>
<td>Thymectomy for thymoma</td>
<td>2,999,977</td>
<td>4</td>
<td>2,999,973</td>
<td>12</td>
<td>0.72</td>
<td>34,000–580,000</td>
</tr>
<tr>
<td>SLE</td>
<td>2,999,936</td>
<td>3,600</td>
<td>2,996,336</td>
<td>3.7</td>
<td>0.59</td>
<td>13–130</td>
</tr>
<tr>
<td>Pernicious anemia</td>
<td>2,999,936</td>
<td>47,399</td>
<td>2,952,537</td>
<td>0.010</td>
<td>1.0</td>
<td>0.14–7.1</td>
</tr>
</tbody>
</table>

D = total number exposed who did not become ill; NP = total number of exposed subjects; D1 = number of exposed subjects with the risk factor who remained well; D2 = number of exposed subjects without the risk factor who remained well; SELN(OR) = standard error of the natural logarithm of the odds ratio; CI = confidence interval; LN = natural logarithm; N1 = the number of subjects with the risk factor who became ill; N2 = the number of subjects without the risk factor who became ill; f = frequency of the suspected risk factor; and OR = odds ratio.

### Table 3
Effect of decreasing CLL to 1 on estimates of the prevalence of the risk group in the general population assuming a hypothetical vaccinated population of 3 million

<table>
<thead>
<tr>
<th>Risk group</th>
<th>Initial estimate of prevalence, f</th>
<th>f at which 95% CLL is reduced to 1</th>
<th>OR with altered f</th>
<th>Altered LN(OR)</th>
<th>SE LnOR</th>
<th>New 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men ≥ 56</td>
<td>0.061</td>
<td>0.29</td>
<td>1.7</td>
<td>0.50</td>
<td>0.25</td>
<td>1.0–2.7</td>
</tr>
<tr>
<td>Women 19–34</td>
<td>0.10</td>
<td>0.15</td>
<td>1.8</td>
<td>0.58</td>
<td>0.30</td>
<td>1.0–3.2</td>
</tr>
<tr>
<td>Autoimmune disease</td>
<td>0.03</td>
<td>0.086</td>
<td>2.0</td>
<td>0.67</td>
<td>0.34</td>
<td>1.0–3.9</td>
</tr>
<tr>
<td>Thymectomy for thymoma</td>
<td>0.00000015</td>
<td>0.048</td>
<td>4.1</td>
<td>1.4</td>
<td>0.72</td>
<td>1.0–17</td>
</tr>
<tr>
<td>SLE</td>
<td>0.0012</td>
<td>0.015</td>
<td>3.2</td>
<td>1.2</td>
<td>0.59</td>
<td>1.0–10</td>
</tr>
</tbody>
</table>

**CLL = lower confidence limit; f = frequency of the suspected risk factor; LN = natural logarithm; SE LnOR = standard error of the natural logarithm of the odds ratio; OR = odds ratio; CI = confidence interval.**
in CCR5 and RANTES genes.\textsuperscript{12} More extensive evaluation of the genomes of additional cases with a variety of risk factors is clearly indicated. As seen for example with live BCG adverse effects,\textsuperscript{13,14} identification of risk groups is an important guide in determining genetic immune defects. Such identification could lead to the prevention and cure of YEL-AVD resulting in safer administration of the live yellow fever virus vaccine.

### Odds Ratios

\[ OR = \frac{N_1}{N_2} \times (1-f)/f \]

\( N_1 \) is the \# of cases with the risk factor, \( N_2 \) is the \# without the risk factor, and \( f \) the prevalence of the risk factor in the population who remained well

#### 95% Confidence Limits (CL)

\[ 95\% \text{ CL}_L = e^{(\ln(OR) - 1.96 \times SE)} \]

\[ 95\% \text{ CL}_U = e^{(\ln(OR) + 1.96 \times SE)} \]

Where

\[ SE_{\ln(OR)} = \sqrt{\frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{D_1} + \frac{1}{D_2}} \]

\( NP = \) total number exposed

\( D_1 = \) \# exposed with the risk factor who remained well = \((NP - (N_1 + N_2))\times f\)

\( D_2 = \) \# exposed without the risk factor who remained well = \((NP - (N_1 + N_2))\times(1-f)\)

#### Setting 95% CL\(_L\) to 1

Use \textit{goal seek} in Microsoft Excel®

"set cell" is the cell containing the initial calculated value for the 95% CL\(_L\)

"to value" is set to 1

"by changing cell" is set to the cell with the initial value for \( f \).

#### \( p \) value from Confidence Intervals

\[ SE_{\ln(OR)} = (\ln(\text{CL}_U) - \ln(\text{CL}_L)) / (2 \times 1.96) \]

\[ z = \text{ABS}(\ln(OR)) / SE_{\ln(OR)} \]

\[ p = e^{(0.717 \times z - 0.416 \times z^2)} \]

---

Box 1. Formulas used to calculate odds ratios (ORs), confidence intervals (CIs), discrimination analysis of CIs, and \( P \) values using estimate of the frequency of the suspected risk factor. The \( P \) values are two-sided.
Received September 19, 2013. Accepted for publication November 9, 2013.

Published online January 6, 2014.

Acknowledgments: We thank Tom Monath for supplying the information on 65 cases of YEL-AVD used in the preparation of a book chapter. We also thank J. Martin Bland for reviewing the odds ratio formula and pointing out that f refers to the frequency of the suspected risk factor in the population that remained well and not to the frequency in the total population.

Financial support: This work was supported in part by U.S. National Science Foundation grants EF-1038337 and DMS-1225529 to JEC and in part by a Howard Hughes Medical Institute grant to J-LC.

Authors’ addresses: Stephen J. Seligman, Department of Microbiology and Immunology, New York Medical College, Valhalla, NY, E-mail: stephen_seligman@nymc.edu. Joel E. Cohen, Laboratory of Populations, The Rockefeller University, New York, NY, and Columbia University, New York, NY, E-mail: cohen@mail.rockefeller.edu. Yuval Itan and Jean-Laurent Casanova, St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University, New York, NY, E-mails: yitan@mail.rockefeller.edu and Jean-Laurent.Casanova@mail.rockefeller.edu. John C. Pezzullo, Department of Medicine, Georgetown University, Washington, DC, E-mail: pezzullo@georgetown.edu.

REFERENCES


