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

Department of Microbiology and Immunology, New York Medical College, Valhalla, New York; Laboratory of Populations, The Rockefeller University, New York, New York and Columbia University, New York, New York; St. Giles Laboratory of Human Genetics of Infectious Diseases, The Rockefeller University, New York, New York; Department of Medicine, Georgetown University, Washington, DC

Abstract. Several risk groups are known for the rare but serious, frequently fatal, viscerotropic reactions following live yellow fever virus vaccine (YEL-AVD). Establishing additional risk groups is hampered by ignorance of the numbers of vaccinees in factor-specific risk groups thus preventing their use as denominators in odds ratios (ORs). Here, we use an equation to calculate ORs using the prevalence of the factor-specific risk group in the population who remain well. The 95% confidence limits and *P* values can also be calculated. Moreover, if the estimate of the prevalence is imprecise, discrimination analysis can indicate the prevalence at which the confidence interval results in an OR of ~1 revealing if the prevalence might be higher without yielding a non-significant result. These methods confirm some potential risk groups for YEL-AVD and cast doubt on another. They should prove useful in situations in which factor-specific risk group denominator data are not available.

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[2,7] (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[8] 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.[9] 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

$$\text{OR} = (N_1/D_1)/(N_2/D_2), \tag{1}$$

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

$$D_1 = f \times D \tag{2}$$

and

$$D_2 = D - D_1$$

———————
*Address correspondence to Stephen J. Seligman, 48 Locust Road, Briarcliff Manor, New York 10510. E-mail: Stephen_seligman@nymc.edu

Using Eq. (2) and substituting for D_1

$$D_2 = D - f \times D = (1 - f) \times D. \qquad (3)$$

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

$$OR = (N_1/N_2) \times (1 - f)/f. \qquad (4)$$

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 ($SE_{LN[OR]}$) can be approximated by

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

The 95% confidence limits (CL) around the logarithm of the OR are the logarithms of the OR ± 1.96 times the $SE_{LN(OR)}$, and the antilogarithms of these logarithmic limits are the 95% CIs 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 (NP). In most situations, NP >> D_1 and NP ≈ 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

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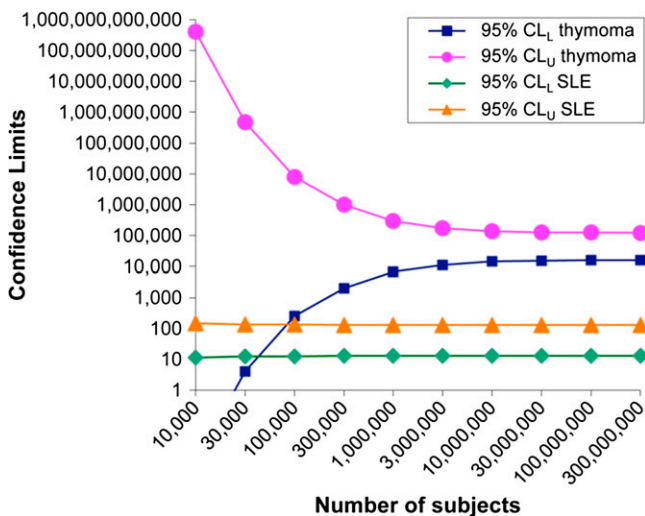
and

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

FIGURE 1. Variation of confidence limits (CLs) around the odds ratio (OR) with number of subjects with rare (thymectomy for thymoma) and more common systemic lupus erythematosus (SLE) potential risk factors.

Evaluating variations in the estimate of the prevalence of the risk group. Because f is not usually known with precision, it is of interest to estimate how much f can vary and still be associated with 95% CIs whose lower limit is > 1. Such calculations are facilitated by using an iterative procedure, such as the "goal seek" feature in Microsoft Excel (Microsoft Corp., Redmond, WA). In the goal seek setup box, the "set cell" is the cell containing the initial calculated value for the 95% CL_L, the "to value" is set to 1, and the "by changing cell" is set to the cell with the initial value for f. The first and third of these cell values will change so that it is advisable to have the initial values preserved in other cells in the worksheet. If CLs for the estimate of f are available (none is in the current examples), the new value of f is acceptable if it remains within the original CLs.

Estimating P values from 95% CIs. Using a numerical approximation for the normal probability distribution, the following method for estimating P values from 95% CLs has been developed[10]:

$$SE_{LN(OR)} = (LN[CL_U] - LN[CL_L])/(2 \times 1.96)$$

$$z = ABS(LN[OR]/SE_{LN[OR]})$$

$$p \approx e^{-0.717 \times z - 0.416 \times z \times z}$$

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

TABLE 1

Calculations of odds ratios using estimates of the prevalence of the risk group in the vaccinated population who did not become ill

Risk group	Prevalence of risk group in the vaccinated population who did not become ill, f	N_1	N_2	OR = (N_1/N_2) × $(1-f)/f$
Men ≥ 56	0.061	26	38	10
Women 19–34	0.10	15	49	2.8
Autoimmune disease	0.03	10	54	6.0
Thymectomy for thymoma	0.0000015	4	19	140,000
Systemic lupus erythematosus (SLE)	0.0012	3	61	41
Pernicious anemia	0.016	1	63	0.99

N_1 = the number of subjects with the risk factor who became ill; N_2 = the number of subjects without the risk factor who became ill; OR = odds ratio. Frequency estimates for men ≥ 56 years of age and women between 19 and 34 years of age were obtained using the average from three references assuming equal numbers of male and female vaccinees. The remaining frequencies were those used in previous calculations by Monath (Table 3 in Reference [2]). For the risk group, individuals thymectomized as treatment of thymoma, N_2 was reduced to 19 because following the report of that association in 2004,[3] vaccination of such individuals was no longer recommended. Accordingly, the chance for accrual of additional thymectomized individuals became remote. In addition to the three cases of systemic lupus erythematosus (SLE), the two cases of Addison's disease, and the one case of pernicious anemia, one case each of autoimmune disease and YEL-AVD occurred with ulcerative colitis, myasthenia gravis, the combination of polymyalgia with hypothyroidism, and Crohn's disease (recent analysis raises the possibility that the latter entity may be an innate immune defect of macrophages and not autoimmune[11]).

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

Risk group	$D = NP - (N_1 + N_2)$	$D_1 = D \times f$	$D_2 = D \times (1 - f)$	LN(OR)	$SE_{LN(OR)}$	95% CI $= e^{(LN[OR] \pm 1.96 \times SE_{LN[OR]})}$
Men $\geq$ 56	2,999,936	183,596	2,816,340	2.4	0.25	6.4–17
Women 19–34	2,999,936	275,236	2,701,200	1.0	0.30	1.6–4.9
Autoimmune disease	2,999,936	89,998	2,909,938	1.8	0.34	3.0–12
Thymectomy for thymoma	2,999,977	4	2,999,973	12	0.72	34,000–580,000
SLE	2,999,936	3,600	2,996,336	3.7	0.59	13–130
Pernicious anemia	2,999,936	47,399	2,952,537	0.010	1.0	0.14–7.1

D = total number exposed who did not become ill; NP = total number of exposed subjects; D_1 = number of exposed subjects with the risk factor who remained well; D_2 = number of exposed subjects without the risk factor who remained well; $SE_{LN(OR)}$ = standard error of the natural logarithm of the odds ratio; CI = confidence interval; LN = natural logarithm; N_1 = the number of subjects with the risk factor who became ill; N_2 = the number of subjects without the risk factor who became ill; f = frequency of the suspected risk factor; and OR = odds ratio.

for populations over 3 million (Figure 1). The total number of vaccinees after the occurrence of YEL-AVD was recognized is over 500 million[1]; however, case ascertainment particularly in Africa is frequently uncertain. The assertions that the ORs are > 1 for the suspected risk groups: men $\geq$ 56 years of age, women 19 to 34 years of age, patients with autoimmune disease, people thymectomized because of thymoma, and SLE (Table 1) are supported by the CL_L being appreciably > 1.

The data for pernicious anemia do not support its inclusion as a risk group. However, because the SE approximation is inaccurate for numbers $\leq$ 4, and $N_1 = 1$ 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 N_1 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 $\geq$ 56 years of age, the prevalence increases 4.8-fold, in women 19–34 years of age 1.5-fold, in patients with autoimmune disease 2.9-fold, with thymectomy for thymoma the increase is 32,000-fold, and for SLE, 13-fold. In all instances the ORs remain > 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% CL_L values $\geq$ 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,[11] P values for men $\geq$ 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% CL_L 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 $\geq$ 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.[1] 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 3
Effect of decreasing CL_L to 1 on estimates of the prevalence of the risk group in the general population assuming a hypothetical vaccinated population of 3 million

Risk group	Initial estimate of prevalence, f	f at which 95% CL_L is reduced to ~1	OR with altered f	Altered LN(OR)	$SE_{LN(OR)}$	New 95% CI
Men $\geq$ 56	0.061	0.29	1.7	0.50	0.25	1.0–2.7
Women 19–34	0.10	0.15	1.8	0.58	0.30	1.0–3.2
Autoimmune disease	0.03	0.086	2.0	0.67	0.34	1.0–3.9
Thymectomy for thymoma	0.0000015	0.048	4.1	1.4	0.72	1.0–17
SLE	0.0012	0.015	3.2	1.2	0.59	1.0–10

CL_L = lower confidence limit; f = frequency of the suspected risk factor; LN = natural logarithm; $SE_{LN(OR)}$ standard error of the natural logarithm of the odds ratio; OR = odds ratio; CI = confidence interval.

<div align="center">TABLE 4</div>
<div align="center">Calculations of P values from confidence interval (CI)</div>

Risk group	LN(OR)	$SE_{LN(OR)} = (LN[CL_U] - LN[CL_L])/(2 \times 1.96)$	$z = Abs(LN[OR]/SE_{LN[OR]})$	$P = e^{(-0.717 \times z - 0.416 \times z \times z)}$
Men ≥ 56	2.4	0.25	9.2	< 0.0001
Women 19–34	1.0	0.30	3.5	0.0006
Autoimmune disease	1.9	0.33	5.7	< 0.0001
Thymectomy for thymoma	12	0.72	16	< 0.0001
SLE	3.7	0.59	6.3	< 0.0001
Pernicious anemia	−0.01	1.0	0.01	0.99

CL_U upper CI, CL_L lower CI; Abs = absolute value; LN = natural logarithm; OR = odds ratio; SLE = systemic lupus erythematosus.

in CCR5 and RANTES genes.[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,[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

$OR = (N_1 / N_2) \times (1-f)/f$

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)

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

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

Where

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

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 **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

$$SE_{LN(OR)} = (LN(CL_U) - LN(CL_L)) / (2 \times 1.96)$$

$$z = ABS(LN(OR) / SE_{LN(OR)})$$

$$p = e^{(-0.717 \times z - 0.416 \times z \times z)}$$

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.

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

1. Monath TP, Gershman M, Staples JE, Barrett AD, 2012. Yellow fever vaccine. Plotkin SA, Orenstein WA, and Offit PA, eds. *Vaccines*. Sixth edition. China: Saunders, 870–968.
2. Monath TP, 2012. Review of the risks and benefits of yellow fever vaccination including some new analyses. *Expert Rev Vaccines 11:* 427–448.
3. Barwick R, 2004. History of thymoma and yellow fever vaccination. *Lancet 364:* 936.
4. Lindsey NP, Schroeder BA, Miller ER, Braun MM, Hinckley AF, Marano N, Slade BA, Barnett ED, Brunette GW, Horan K, Staples JE, Kozarsky PE, Hayes EB, 2008. Adverse event reports following yellow fever vaccination. *Vaccine 26:* 6077–6082.
5. Khromava AY, Eidex RB, Weld LH, Kohl KS, Bradshaw RD, Chen RT, Cetron MS, 2005. Yellow fever vaccine: an updated assessment of advanced age as a risk factor for serious adverse events. *Vaccine 23:* 3256–3263.
6. Seligman SJ, 2011. Yellow fever virus vaccine-associated deaths in young women. *Emerg Infect Dis 17:* 1891–1893.
7. Gershman MD, Staples JE, Bentsi-Enchill AD, Breugelmans JG, Brito GS, Bastos Camacho LA, Cottin P, Domingo C, Durbin A, Gascon J, Guenaneche F, Hayes EB, Jelenik Z, Khromava A, Martins RD, Wilson MM, Massy N, Nasidi A, Niedrig M, Sherwat A, Tsai T, Vilella A, Wilson ME, Kohl KS, 2012. Viscerotropic disease: case definition and guidelines for collection, analysis, and presentation of immunization safety data. *Vaccine 30:* 5038–5058.
8. Isenman H, Burns A, 2012. A case of yellow fever vaccine-associated disease. *NZ Med J 125:* 92–95.
9. Breugelmans JG, Lewis RF, Agbenu E, Veit O, Jackson D, Domingo C, Bothe M, Perea W, Niedrig M, Gessner BD, Yactayo S, 2013. Adverse events following yellow fever preventive vaccination campaigns in eight African countries from 2007 to 2010. *Vaccine 31:* 1819–1829.
10. Altman DG, Bland JM, 2011. How to obtain the *P* value from a confidence interval. *BMJ 343:* d2304.
11. Casanova JL, Abel L, 2009. Revisiting Crohn's disease as a primary immunodeficiency of macrophages. *J Exp Med 206:* 1839–1843.
12. Pulendran B, Miller J, Querec TD, Akondy R, Moseley N, Laur O, Glidewell J, Monson N, Zhu T, Zhu H, Staprans S, Lee D, Brinton MA, Perelygin AA, Vellozzi C, Brachman P Jr, Lalor S, Teuwen D, Eidex RB, Cetron M, Priddy F, del Rio C, Altman J, Ahmed R, 2008. Case of yellow fever vaccine-associated viscerotropic disease with prolonged viremia, robust adaptive immune responses, and polymorphisms in CCR5 and RANTES genes. *J Infect Dis 198:* 500–507.
13. Casanova JL, Jouanguy E, Lamhamedi S, Blanche S, Fischer A, 1995. Immunological conditions of children with BCG disseminated infection. *Lancet 346:* 581.
14. Alcais A, Quintana-Murci L, Thaler DS, Schurr E, Abel L, Casanova JL, 2010. Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? *Ann N Y Acad Sci 1214:* 18–33.