

5-1-2011

Fisher's Exact Test for Misclassified Data

Tze-San Lee

Western Illinois University, leetzesan@gmail.com

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Lee, Tze-San (2011) "Fisher's Exact Test for Misclassified Data," *Journal of Modern Applied Statistical Methods*: Vol. 10 : Iss. 1 , Article 14.

DOI: 10.22237/jmasm/1304223180

Fisher's Exact Test for Misclassified Data

Tze-San Lee
 Western Illinois University
 Macomb, IL USA

Fisher's exact test is adapted to handle the misclassified data arising from comparing two binomial populations. The bias-adjusted odds ratio is proposed to account for misclassification errors. Its expected power depends in a nonlinear way on the true sensitivity and specificity of the classification method. The data taken from the no conviction rate of criminality for two types of twin populations was used to illustrate how to calculate true sensitivity and specificity and the expected power of the adjusted odds ratio.

Key words: Fisher's exact test, misclassification, power function, odds ratio, sensitivity, specificity.

Introduction

Fisher's (1946) exact test is used when the sample size is less than five. However, the issue on how to adapt Fisher's exact test if the data are misclassified has not been addressed. It is the aim of this article to adapt Fisher's exact test to account for misclassification errors.

Methodology

Consider two independent binomial random variables X and Y with parameters (n_X, p_X) and (n_Y, p_Y) , respectively, where both n_X and n_Y are less than 5. A classical problem is to find an exact test for the null hypothesis $H_0: p_X = p_Y$ against an alternative hypothesis $H_1: p_X > p_Y$, or equivalently, $H_0: \gamma = 1$ against $H_1: \gamma > 1$, where γ is the odds ratio defined by (Fleiss, Levin & Paik, 2003)

$$\gamma = \frac{p_X q_Y}{p_Y q_X}, \quad (1)$$

Let the number of successes be x and y , respectively, among n_X and n_Y subjects. Assume that among x and y successes there were possible misclassified cases. Before showing how to adapt Fisher's exact test to deal with misclassified data, a depiction of Fisher's exact test is provided.

By conditioning that $x + y = z$ is fixed, the [conditional] distribution of $X = x$ is given by the extended (or non-central) hypergeometric distribution under the alternative hypothesis (Gart, 1971; Harkness, 1965; Levin, 1984)

$$\Pr(X = x | z; \gamma) = \frac{\binom{n_X}{x} \binom{n_Y}{z-x} \gamma^x}{\sum_{j=0}^z \binom{n_X}{j} \binom{n_Y}{z-j} \gamma^j}, \quad (2)$$

where $x = \max(0, z - n_Y), \dots, \min(z, n_X)$, or the [conditional] distribution of $Y = y$ is given by

$$\Pr(Y = y | z; \gamma) = \frac{\binom{n_X}{z-y} \binom{n_Y}{y} \gamma^{-y}}{\sum_{j=0}^z \binom{n_X}{z-j} \binom{n_Y}{j} \gamma^{-j}}, \quad (3)$$

Tze-San Lee is presently working at the Centers for Disease Control and Prevention, Chamblee, GA. Email him at: tjl3@cdc.gov.

where $y = \max(0, z - n_x), \dots, \min(z, n_y)$. In practical applications, equation 2 is used if $x < y$; otherwise, equation 3 is used. Note that equations 2 and 3 were first introduced by Fisher (1935). Equation 2 (or 3) can be used to devise significance tests or confidence intervals on any value of γ .

If $x > y$, then the p -value of Fisher's exact test $H_0: \gamma = 1$ against $H_1: \gamma > 1$ is given by

$$p\text{-value} = \sum_{k=y}^z \Pr(Y = k | z; \gamma = 1), \quad (4)$$

where $\Pr(Y = k | z; \gamma = 1)$ is given by equation 3 which yields an ordinary hypergeometric distribution as follows:

$$\Pr(Y = y | z; \gamma = 1) = \frac{\binom{n_x}{z-y} \binom{n_y}{y}}{\binom{n_x + n_y}{z}},$$

$$y = 0, 1, \dots, z, \quad (5)$$

on which the Fisher-Irwing exact test is based.

For small frequencies, the critical value of Y has been provided by choosing y_c for one-sided alternatives ($p_x > p_y$) such that

$$\Pr(Y \leq y_c | z) = \sum_{y=0}^{y_c} \Pr(Y = y | z; \gamma = 1) \leq \alpha$$

and

$$\Pr(Y \leq y_c + 1 | z) > \alpha. \quad (6)$$

For nominal levels of significance $\alpha = 0.05, 0.025, 0.01, 0.005$, x_c has been tabulated for $n_y \leq n_x \leq 25$ (Bennett & Hsu, 1960). For two-sided alternatives ($p_x \neq p_y$), the tabular exact probabilities are doubled accordingly.

If $x > y$, the (conditional) power function of the exact test is then given by

$$\beta(\gamma | z) = \Pr(y \leq y_c | H_1)$$

$$= \sum_{y=0}^{y_c} \Pr(Y = y | z; \gamma) \quad , \quad (7)$$

$$= \frac{\sum_{y=0}^{y_c} \binom{n_x}{z-y} \binom{n_y}{y} \gamma^{-y}}{\sum_{j=z_1}^{z_2} \binom{n_x}{z-j} \binom{n_y}{j} \gamma^{-j}}$$

where

$$z_1 = \max(0, z - n_x)$$

and

$$z_2 = \max(z, n_y).$$

Note that $\beta(\gamma | z)$ of equation 7 is a rational function in γ , that is, a ratio of two polynomial functions in γ .

If $x > y$, the expected power of the exact test is given by

$$\beta(\gamma) = \sum_z \beta(\gamma | z) \Pr(Z = z)$$

$$= q_x^{n_x} q_y^{n_y} \left\{ \sum_z \left(\frac{p_x}{q_x} \right)^z \left[\sum_{y=z_1}^{y_c} \binom{n_x}{z-y} \binom{n_y}{y} \gamma^{-y} \right] \right\}, \quad (8)$$

where $\Pr(Z = z)$ representing the distribution of Z is given by

$$\Pr(Z = z) = \sum_{i=L}^U \binom{n_x}{z-i} p_x^{z-i} q_x^{n_x-z+i} \binom{n_y}{i} p_y^i q_y^{n_y-i}, \quad (9)$$

where $L = \max(0, z - n_x)$, $U = \min(z, n_y)$ and the summation in z is over all significance pairs of points on the diagonals, $z = x + y$, in the (x, y) sample space at level of significance equals to α at most (Bennett & Hsu, 1960; Casagrange, Pike & Smith, 1978a). To facilitate a calculation of the expected power of equation 7 a FORTRAN program was written by Casagrange, Pike and Smith (1978b). If $n_x = n_y$, Conlon & Thomas (1993) presented an algorithm which was feasible for very large sample sizes.

FISHER'S EXACT TEST FOR MISCLASSIFIED DATA

If classification errors exist on the number of successes for both X and Y , the question becomes how to test the hypotheses previously postulated. Let E^* be the surrogate classification variable for E and let φ_Z and ψ_Z , $Z = X, Y$, be the sensitivity and specificity for classifying the status of the outcome among samples from the case and the control populations, respectively, that is, for $Z = X$ or Y ,

$$\begin{aligned} \varphi_Z &= \Pr(E^* = 1 | E = 1), \\ \text{and} \\ \psi_Z &= \Pr(E^* = 0 | E = 0). \end{aligned} \quad (10)$$

It is known that, for the unknown p_X and p_Y , the following maximum likelihood estimators are no longer unbiased:

$$\hat{p}_X = x/n_X \text{ and } \hat{p}_Y = y/n_Y, \quad (11)$$

assume that $\hat{p}_Z \sim \text{Binomial}(n_Z, p_Z)$ for $Z = X, Y$. The crude odds ratio (COR) defined by

$$\hat{\gamma} = \frac{\hat{p}_X \hat{q}_Y}{\hat{p}_Y \hat{q}_X}, \quad (12)$$

as a point estimator for the true odds ratio γ of equation 1 can have substantial bias (Kleinbaum, et al., 1982).

To account for the misclassification bias, the bias-adjusted [point] estimators for the prevalence of success/failure p_Z and q_Z are given by (Lee, 2009)

$$\begin{aligned} \bar{p}_Z &= (\psi_Z - \hat{q}_Z) / \Delta_Z \\ \text{and} \\ \bar{q}_Z &= (\varphi_Z - \hat{p}_Z) / \Delta_Z, \end{aligned} \quad (13)$$

where $\hat{q}_Z = 1 - \hat{p}_Z$, $Z = X, Y$, and Δ_Z is given by

$$\Delta_Z \equiv \varphi_Z + \psi_Z - 1. \quad (14)$$

Conditioned on that φ_Z and ψ_Z are given, it is easily shown that equation 13 is an unbiased estimator for p_Z and q_Z , respectively. The

bias-adjusted estimators \bar{p}_Z and \bar{q}_Z (equation 13) are said to be plausible if \bar{p}_Z and \bar{q}_Z lie between 0 and 1. In order for \bar{p}_Z and \bar{q}_Z to be plausible, the following constraints are imposed: for $Z = X, Y$,

$$\varphi_Z > \hat{p}_Z, \psi_Z > \hat{q}_Z \text{ and } \Delta_Z > 0. \quad (15)$$

A set of φ_Z and ψ_Z is said to be feasible if equation 15 holds. Furthermore, a set of feasible φ_Z and ψ_Z is said to be admissible if for these feasible φ_Z and ψ_Z , \bar{p}_Z and \bar{q}_Z are plausible.

The bias-adjusted odds ratio (BAOR) for γ which accounts for misclassification bias is then given by

$$\begin{aligned} \bar{\gamma} &= \frac{\bar{p}_X \cdot \bar{q}_Y}{\bar{p}_Y \cdot \bar{q}_X} \\ &= \frac{(\psi_X - \hat{q}_X)(\phi_Y - \hat{p}_Y)}{(\phi_X - \hat{p}_X)(\psi_Y - \hat{q}_Y)} \end{aligned} \quad (16)$$

and its asymptotic variance is given by

$$\text{var}(\ln(\bar{\gamma})) = \sum_{Z=X}^Y \frac{\text{var}(\hat{p}_Z)}{(\Delta_Z p_Z q_Z)^2}, \quad (17)$$

where $\text{var}(\hat{p}_Z)$ is given by

$$\text{var}(\hat{p}_Z) = \frac{(p_Z \Delta_Z + 1 - \psi_Z)(q_Z \Delta_Z + 1 - \phi_Z)}{n_Z}. \quad (18)$$

Using equations 16-17 to find a 100% $\times(1 - \alpha)$ confidence interval (LCL, UCL) for the true γ (equation 1), where LCL and UCL are abbreviations denoting for lower and upper confidence limit, respectively, and $0 < \alpha < 1$ as follows:

$$\text{(LCL, UCL):} = \exp(\ln(\bar{\gamma}) \mp z_{1-\frac{\alpha}{2}} \times \sqrt{\text{var}(\ln(\bar{\gamma}))}). \quad (19)$$

Note that Fisher's exact test, which accounts for misclassification errors, is exactly the same as

that shown previously, that is, the p-value for the significance test of $H_0: \gamma = 1$ against $H_1: \gamma > 1$ does not depend upon the sensitivity and specificity of the classification method at all. But the conditional power of equation 7 or the expected power of equation 8 at $\gamma = \bar{\gamma}$ depends on the sensitivity and specificity of the classification method for both populations.

Results

The Lange’s data on criminality among twin brothers/sisters of criminals (Fisher, 1946) was used for analysis. Table 1 shows the numbers of twin brothers/sisters of criminals who have been convicted, separately for dizygotic (= X) (but like-sexed) and monozygotic twins (= Y). Because $\hat{p}_X = 13/15 > 3/13 = \hat{p}_Y$, the phrase not convicted is taken as success. Inspection of Table 1 shows $x = 13$, $y = 3$, $n_X = 15$, and $n_Y = 13$. The COR of equation 12 was obtained as $\hat{\gamma} = 21.7$ with $p = 0.001$ by using the SAS software with a specification to Fisher’s exact test (Stokes, Davis & Koch, 2000). As a result, the null hypothesis is rejected. This means that the deviation from proportionality in Table 1 is significant to provide evidence that criminality is more frequent among monozygotic twins of criminals than among dizygotic twins of criminals (Finney, 1948).

Table 1: Lange’s Data on Criminality among Twin Brothers/Sisters of Criminals

	Dizygotic (= X)	Monozygotic (= Y)	Row Total
Not convicted (Success)	13	3	16
Convicted (Failure)	2	10	12
Column Total	15	13	28

Suppose that classification error exists in the observed data shown in Table 1. Because

validation data was not available and because it was not possible to know what true classification is, all possible re-arrangements of Table 1 were considered and then each re-arranged table was treated as if it were a true table. Thus, it was possible to calculate true sensitivity and specificity for both of the two populations. For the population of dizygotic twins, there are a total of 13 possible truly classified tables. Hence, in theory, there are a total of 13 possible pairs of true sensitivities and specificities for this targeted population. However, after checking the feasibility constraints (equations 14-15), only four pairs of sensitivity and specificity were feasible (Table 2). Similarly, five out of a total of eleven possible pairs of sensitivity and specificity were feasible for the population of monozygotic twins (Table 2).

Because it was not possible to know which pair was the true sensitivity and specificity for either one of the two populations of twins, it was necessary to calculate the BAOR, $\bar{\gamma}$ (equation 16), for all 20 (= 4×5) possible combinations of feasible pairs of sensitivity and specificity for the two targeted populations of twins. The calculation was organized as follows. One pair of feasible sensitivity and specificity was fixed from the population of dizygotic twins and then combined with all five pairs of feasible sensitivity and specificity for the population of monozygotic twins in order to calculate $\bar{\gamma}$ (equation 16). This procedure was then repeated by changing only the pair from the population of dizygotic twins until all four feasible pairs were used (Table 3). As shown in Table 3, only three BAORs from the 2nd to the 4th entries were found to be significant for cases i-ii, whereas none of the BAORs were significant for cases iii-iv. If the COR is credible, then this implies that to under-misclassify two or three pairs of dizygotic twins in the convicted category is implausible. If only one pair of dizygotic twins is over-misclassified (comparing case i with the correctly classified pair of dizygotic twins in Table 2), the COR ($\hat{\gamma} = 21.7$) over-estimated the true γ because the BAORs were 17.2 and 19.5 when one pair of monozygotic twins was under- and over-misclassified in the convicted category, while under-estimated the true γ because the BAOR

FISHER'S EXACT TEST FOR MISCLASSIFIED DATA

Table 2: Pairs of Feasible Sensitivity and Specificity for Two Types of Twins under the Assumption that True Classifications Are Known

Dizygotic [$\hat{p}_x = 0.867$, $\hat{q}_x = 0.133$]				Monozygotic [$\hat{p}_y = 0.23$, $\hat{q}_y = 0.77$]			
Convicted		φ_x	ψ_x	Convicted		φ_y	ψ_y
No	Yes			No	Yes		
14	1	0.963	0.667	1	12	0.5	0.909
12	3	0.960	0.800	2	11	0.8	0.952
11	4	0.917	0.667	4	9	0.857	0.947
10	5	0.870	0.571	5	8	0.75	0.889
				6	7	0.667	0.824

was 24.0 when two pairs of monozygotic twins was over-misclassified in the convicted category. Similarly, If only one pair of dizygotic twins is under-misclassified (case ii), the COR under-estimates the true γ because the BAORs were given by 22.2, 25.1, and 31.0 corresponding to when one pair was either under- or over-misclassified or two pairs of monozygotic twins were over-misclassified, respectively.

To calculate the expected power for either $\hat{\gamma}$ (equation 12) or $\bar{\gamma}$ (equation 16), the crude/bias-adjusted point estimator is substituted for p_i and q_i , namely, the COR or BAOR for γ in equation 8; thus the expected power of equation 8 (Table 4) were obtained. Note that the results shown in the first row of Table 4 correspond to the COR because the COR can be viewed as a special case of the BAOR with perfect classification, that is, both sensitivity and specificity equal to one. If $n_x = 15$, $n_y = 13$, and $x = 13$, then $y_c = 6$ from the table of critical values for y (Finney, 1948; p. 154); this is used in determining all possible z -values in equation 8.

The results of the expected power for the COR and six admissible BAORs are given in Table 4. It is not surprising to see that the COR has the highest expected power (= 0.45) because both the sensitivity and specificity equals one, whereas the expected power of the six BAORs varies. It seems that the expected power of the BAOR depends on the values of the sensitivity

and specificity: the higher the specificity across the board, the larger the expected power. Indeed, the expected power is higher for $(\varphi_x, \psi_x) = (0.96, 0.8)$ than for $(\varphi_x, \psi_x) = (0.963, 0.6667)$, that is, 0.40, 0.36, and 0.29 compared to 0.33, 0.29, and 0.23, respectively (see Table 4). Also, the larger expected power 0.40 (or 0.33) corresponds to the highest specificity ($\psi_y = 0.9524$) for the monozygotic population. In terms of the type of misclassification the highest power corresponds to that exactly one pair of twins is under-misclassified in the category of convicted for both populations.

Conclusion

Fisher's exact test was adapted to handle a scenario where data are misclassified. The bias-adjusted odds ratio was proposed to account for the misclassification errors. Because a validation sample is not available, all possible pairs of true sensitivity and specificity were calculated from the observed data by assuming that a true table is known. Although the p-value is not affected by the true sensitivity and specificity of the classification method, the expected power of Fisher's exact test depends on these in a nonlinear way. The data regarding whether the no-conviction rate are the same between the dizygotic and monozygotic twins of brothers/sisters was used to illustrate how to calculate true sensitivity and specificity, the bias-adjusted odds ratio and their expected power accordingly.

Table 3: The Point Estimate, Standard Error of Logarithm of Adjusted Odds Ratio and 95% CI for 20 Possible Combinations of True Sensitivity and Specificity of Two Twin Populations

(φ_Y, ψ_Y)	$\bar{\gamma}$	$s.e.(\ln(\bar{\gamma}))$	LCL	UCL
(i) $(\varphi_X, \psi_X) = (0.963, 0.667)$				
(0.5, 0.909)	10.7	1.66	0.41	278.4
(0.8, 0.952)	17.2	1.37	1.18	251.1
(0.857, 0.947)	19.5	1.37	1.34	283.8
(0.75, 0.889)	24.0	1.61	1.02	567.7
(0.667, 0.824)	44.5	2.65	0.25	7993.5
(ii) $(\varphi_X, \psi_X) = (0.96, 0.8)$				
(0.5, 0.909)	13.8	1.66	0.53	357.0
(0.8, 0.952)	22.2	1.36	1.53	321.7
(0.857, 0.947)	25.1	1.36	1.74	363.6
(0.75, 0.889)	31.0	1.61	1.32	727.9
(0.667, 0.824)	57.3	2.65	0.32	10271.6
(iii) $(\varphi_X, \psi_X) = (0.917, 0.667)$				
(0.5, 0.909)	20.5	2.30	0.23	1869.6
(0.8, 0.952)	33.2	2.10	0.54	2020.8
(0.857, 0.947)	37.5	2.10	0.62	2284.9
(0.75, 0.889)	46.3	2.27	0.55	3921.5
(0.667, 0.824)	85.6	3.09	0.20	36497.9
(iv) $(\varphi_X, \psi_X) = (0.87, 0.571)$				
(0.5, 0.909)	291.0	30.5	3.0×10^{-24}	2.7×10^{28}
(0.8, 0.952)	469.8	30.5	5.1×10^{-24}	4.3×10^{28}
(0.8571, 0.947)	531.6	30.5	5.8×10^{-24}	4.9×10^{28}
(0.75, 0.889)	656.0	30.5	7.0×10^{-24}	6.1×10^{28}
(0.667, 0.824)	1213.6	30.6	1.1×10^{-23}	1.3×10^{29}

Table 4: The Expected Power of the Crude/Adjusted Odds Ratio

(φ_X, ψ_X)	(φ_Y, ψ_Y)	p_X	p_Y	γ	$\beta(\gamma z)$
(1.0, 1.0)	(1.0, 1.0)	0.87	0.23	21.7	0.45
(0.963, 0.667)	(0.8, 0.952)	0.85	0.24	17.2	0.33
(0.963, 0.667)	(0.857, 0.947)	0.85	0.22	19.5	0.29
(0.963, 0.667)	(0.75, 0.889)	0.85	0.19	24.0	0.23
(0.96, 0.8)	(0.8, 0.952)	0.88	0.24	22.2	0.40
(0.96, 0.8)	(0.857, 0.947)	0.88	0.22	25.1	0.36
(0.96, 0.8)	(0.75, 0.889)	0.88	0.19	31.0	0.29

FISHER'S EXACT TEST FOR MISCLASSIFIED DATA

Acknowledgement

The author would like to acknowledge the use of the Microsoft's EXCEL spread sheet to facilitate numerical calculations for Tables 2-4.

References

- Agresti, A. (1992). A survey of exact inference for contingency tables. *Statistical Science*, 7, 131-153.
- Agresti, A. (2002). *Categorical data analysis* (2nd Ed.). New York: Wiley.
- Bennett, B. M., & Hsu, P. (1960). On the power function of the exact test for the 2×2 contingency table. *Biometrika*, 47, 393-398.
- Casagrande, J. T., Pike, M. C., & Smith, P. G. (1978a). The power function of the exact test for comparing two binomial distribution. *Applied Statistics*, 27, 176-180.
- Casagrande, J. T., Pike, M. C., & Smith, P. G. (1978b). Algorithm AS 129: The power function of the exact test for comparing two binomial distribution. *Applied Statistics*, 27, 212-219.
- Colon, M., & Thomas, R. G. (1993). Algorithm AS: 280: The power function for Fisher's exact test. *Applied Statistics*, 42, 258-260.
- Finney, D. J. (1948). The Fisher-Yates test of significance in 2×2 contingency tables. *Biometrika*, 35, 145-156.
- Fisher, R. A. (1935). The logic of inductive inference. *Journal of the Royal Statistical Society, Series A*, 98, 39-82.
- Fisher, R. A. (1946). *Statistical methods for research workers* (10th Ed.). Edinburgh: Oliver and Boyd.
- Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical methods for rates and proportions* (3rd Ed.). New York: Wiley.
- Gart, J. J. (1971). The comparison of proportions: a review of significance tests, confidence intervals and adjustments for stratification. *Review of the International Statistical Institute*, 39, 148-169.
- Harkness, W. L. (1965). Properties of extended hypergeometric distribution. *Annals of Mathematical Statistics*, 36, 938-945.
- Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research: principles and quantitative methods*. New York: Wiley.
- Lee, T-S. (2009). Adjusted odds ratio for the exposure misclassification in a case-control study. *The Internet Journal of Epidemiology*, 6(2). Accessed from <http://www.ispub.com/journal/the-internet-journal-of-epidemiology/volume-6-number-2/bias-adjusted-exposure-odds-ratio-for-misclassified-data-1.html>.
- Levin, B. (1984). Simple improvements on Cornfield's approximation to the mean of a noncentral hypergeometric random variable. *Biometrika*, 71, 630-632.
- Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical data analysis using the SAS system* (2nd Ed.). Cary, North Carolina: SAS Institute, Inc.