Journal of Modern Applied Statistical Methods

Volume 1 | Issue 2 Article 46

11-1-2002

Accounting For Non-Independent Observations In 2×2 Tables, With Application To Correcting For Family Clustering In Exposure-Risk Relationship Studies

Leslie A. Kalsih
New England Research Institutes, Watertown, MA

Katherine A. Riester
Biogen, Inc., Cambridge, MA

Stuart J. Pocock
London School of Hygiene and Tropical Medicine, London, UK

Journal of Modern Applied Statistical Methods
Fall 2002, Vol. 1, No 2, 379-386

1538 – 9472/02/$30.00

Accounting For Non-Independent Observations In 2×2 Tables, With Application To Correcting For Family Clustering In Exposure-Risk Relationship Studies

Leslie A. Kalish
New England Research Institutes
Watertown, MA

Katherine A. Riester
Biogen, Inc.
Cambridge, MA

Stuart J. Pocock
London School of Hygiene and
Tropical Medicine
London, UK

Participants in epidemiologic studies may not represent statistically independent observations. We consider modifications to conventional analyses of 2×2 tables, including Fisher's exact test and confidence intervals, to account for correlated observations in this setting. An example is provided, assessing the robustness of conclusions from a published analysis.

Key words: Chi-square test for independence, clustered data, confidence interval, correlation, epidemiologic methods, Fisher's exact test, odds ratio

Introduction

Participants in epidemiologic studies may not represent statistically independent observations. For instance, some individuals may belong to the same family. This will usually make simple statistical tests for exposure-risk relationships anti-conservative, i.e., the strength of evidence for a relationship will be exaggerated by ignoring the lack of independence. We consider a method to modify the standard statistical tests for 2×2 tables in this setting, in order to account for such non-independent observations.

For convenience and clarity, we describe the method in terms of an example comparison of "exposed" and "unexposed" children born to mothers enrolled in a study. Intra-family correlations may induce inter-dependence or clustering of outcomes between siblings. If the exposure of interest is a fixed characteristic of the mother, such as whether or not the mother is positive for a hereditary gene mutation, then all children of that mother will be concordant on their exposure. This will tend to induce positive correlations between outcomes in the siblings. Other exposures (e.g., gender of the child) may be concordant or discordant, and some exposures (e.g., birth-order) will always be discordant. Most previous research in this area has focused only on settings with no discordant exposures.

In this paper we provide a correction factor for the ordinary Pearson chi square test for independence, and for the construction of confidence intervals, and also propose a method for applying the correction factor to Fisher's exact test. The correction factor depends on the numbers of concordant pairs in each exposure group, the number of discordant pairs, and the intra-family correlation in outcome. We evaluate properties of the new tests using simulations.

An important application of these methods is in evaluating published epidemiologic findings based on a 2×2 table when correlated observations have been naively assumed to be independent. The methods in this paper can then be used to check the robustness of their findings after accounting for non-independence.

Leslie A. Kalish is Principal Research Scientist. Email: LesK@neri.org. Katherine A. Riester is Senior Biostatistician. This work was performed while she was employed at New England Research Institutes. Email: Katherine_Riester@biogen.com. Stuart Pocock is Professor of Medical Statistics. Email: Stuart.Pocock@lshtm.ac.uk.

Methodology

Suppose there are N_1 and N_2 subjects, respectively, in the exposed and unexposed groups ($N=N_1+N_2$). Let $\hat{\pi}_1$ and $\hat{\pi}_2$ denote the estimated probabilities of a binary disease outcome in the two groups. Assuming all observations are independent, under the null hypothesis of equal response probabilities, $H_0: \pi_1 = \pi_2$, the variance of $\hat{\pi}_1 - \hat{\pi}_2$ is

$$\text{var}\left[\hat{\pi}_1 - \hat{\pi}_2\right] = \pi\left(1-\pi\right)\left[1/N_1 + 1/N_2\right], \tag{1}$$

where $\pi = \pi_1 = \pi_2$.

Thus the normal approximation statistic for testing H_0 is

$$Z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\bar{\pi}\left(1-\bar{\pi}\right)\left[1/N_1 + 1/N_2\right]}}$$

where $\bar{\pi} = \left(N_1\hat{\pi}_1 + N_2\hat{\pi}_2\right)/\left(N_1 + N_2\right)$ denotes the overall estimate of response probability from both groups combined. Z has approximately a standardized normal distribution under H_0 when N_1 and N_2 are large, and Z^2 is the statistic from the ordinary Pearson chi square test for independence.

To account for lack of independence, let ρ be the within-family correlation of disease outcome (i.e., the correlation between binary variables), which is assumed known. Let S be the total number of sibling pairs. Note that each individual can be in more than one of the S pairs, for example four siblings would contribute six pairs to S. Let S_{11}, S_{12} and S_{22} denote the number of concordant exposed, discordant and concordant unexposed pairs, respectively (where "concordant exposed" means that both members of the pair are exposed and the other terms are defined similarly). Thus $S= S_{11}+S_{12}+S_{22}$. Using standard results for the variance of a linear combination of correlated variables, it can be shown that

$$\begin{aligned}\text{var}[\hat{\pi}_1 - \hat{\pi}_2] = \pi(1-\pi)\{1/N_1 + 1/N_2 + \\ 2\rho(S_{11}/N_1^2 - S_{12}/[N_1N_2] + S_{22}/N_2^2\}\end{aligned} \quad . \tag{2}$$

Expressions (1) and (2) suggest that the Pearson chi square statistic should be multiplied by a correction factor.

$$CF = \frac{1/N_1 + 1/N_2}{1/N_1 + 1/N_2 + 2\rho\left(S_{11}/N_1^2 - S_{12}/\left[N_1N_2\right] + S_{22}/N_2^2\right)}.$$

We refer to this as the modified chi square test for independence. In practice, ρ needs to be estimated, or a range of values used, because it is usually unknown.

It seems plausible that the correction factor can also be used to account for non-independence when performing Fisher's exact test, as would be appropriate in studies with small sample sizes. Suppose one wants an α=0.05 level Fisher's exact test. Rather than rejecting H_0 when the sum of probabilities of extreme tables is less than 0.05 (which corresponds to rejecting H_0 if Pearson's chi square statistic is greater than 3.84), one would use the nominal p-value which corresponds to the probability that the chi square distribution exceeds 3.84×CF. We refer to this as the modified Fisher's exact test.

The methods described so far have been in terms of hypothesis testing. By relaxing the null hypothesis assumption that $\pi_1=\pi_2$, one can extend the results so that confidence intervals can be constructed. Generalizing expression (2) by allowing $\pi_1\neq\pi_2$ yields the following formula for the variance of the risk difference, which accounts for correlations:

$$\mathrm{var}\left[\hat{\pi}_1-\hat{\pi}_2\right]=\frac{\pi_1(1-\pi_1)}{N_1}+\frac{\pi_2(1-\pi_2)}{N_2}$$
$$+2\rho\left[\begin{array}{l}\dfrac{\pi_1(1-\pi_1)S_{11}}{N_1^2}\\ -\dfrac{\sqrt{\pi_1(1-\pi_1)\pi_2(1-\pi_2)}S_{12}}{N_1N_2}\\ +\dfrac{\pi_2(1-\pi_2)S_{22}}{N_2^2}\end{array}\right].$$

In practice, π_1 and π_2 would be replaced by observed proportions from the data. Similarly, the familiar variance estimate for the log odds ratio (OR) based on a 2×2 table with cell entries {a,b,c,d}, where $\hat{\pi}_1$=a/(a+b) and $\hat{\pi}_2$=c/(c+d), is

$$\hat{\mathrm{var}}\left(\log\hat{OR}\right)=\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}.$$

With correlated observations this generalizes to:

$$\hat{\mathrm{var}}\left(\log\hat{\mathrm{OR}}\right)=\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}+$$
$$2\rho\left(\begin{array}{l}\dfrac{S_{11}}{N_1}\left[\dfrac{1}{a}+\dfrac{1}{b}\right]-\\ \dfrac{S_{12}}{\sqrt{N_1N_2}}\sqrt{\left[\dfrac{1}{a}+\dfrac{1}{b}\right]\left[\dfrac{1}{c}+\dfrac{1}{d}\right]}+\\ \dfrac{S_{22}}{N_2}\left[\dfrac{1}{c}+\dfrac{1}{d}\right]\end{array}\right).$$

These variance estimates can be used to construct confidence intervals for a risk difference or odds ratio based on a normal approximation.

We evaluate the true size of the modified Pearson chi square test for independence and modified Fisher's exact tests via simulations. For simplicity, in all simulations we assumed equal exposure group sample sizes (N_1=N_2) and a maximum number of siblings per family of two.

Letting θ denote the response probability, consider a fairly rare and a common outcome probability, θ = 0.1 and 0.5; small and large intra-family correlations, ρ={0.2, 0.8}; three total sample sizes, N=N_1+N_2={24, 100, 500}; and a low and high proportion of N which is made up of siblings, 2S/N≈{0.08, 0.64}. (A footnote to Table 3 below explains why we were not always able to achieve 2S/N=0.08 and 0.64 exactly.)

For each combination, we considered three ways that the S sibling pairs could be divided into concordant exposed, concordant unexposed and discordant pairs, as shown in Table 1. In configurations A and B all sibling pairs are concordant whereas in configuration C all pairs are discordant. Configuration A represents the extreme case where all concordant pairs are in a single exposure group. We did not consider cases with both concordant and discordant pairs because the signs on the S_{ij} terms in expression (2) show that these terms would tend to cancel each other out and the results would be intermediate between configurations considered.

All combinations of θ, ρ, N, S and configurations A-C were simulated (except for combinations with {N=24, θ=0.1}, which has a substantial probability of a zero marginal total because the study was too small). Thirty thousand simulations for each combination guaranteed that for a true rejection probability of 0.05, we would have a 95% chance of observing a rejection probability within [0.0475, 0.0525].

In the simulations, we used randomized critical regions (Cox and Hinkley, 1974) to correct for discreteness of the test statistic. Although this may not be used in practice, it makes the different procedures comparable by removing the inherent conservatism in Fisher's exact test (Agresti, 1996).

Example

Dickover et al. (1996) analyzed mother-to-child transmission of human immunodeficiency virus (HIV) in 97 mother-infant pairs, including two pregnancies resulting in twins and three mothers each having two singleton pregnancies. Thus, the 97 mother-infant pairs represented 95 pregnancies in 92 women.

Table 1. Three configurations for allocating siblings to concordant exposed, concordant unexposed and discordant pairs in the simulation study.

	Configuration								
	A			B			C		
	conc	disc		conc	disc		conc	disc	
exp	100	0	100	50	0	50	0	50	50
unexp	0	0	0	50	0	50	0	50	50
	100	0	100	100	0	100	0	100	100

Note: Numbers in each cell represent the percentage of the total number of siblings. (conc = concordant, disc = discordant, exp = exposed, unexp = unexposed).

One of the exposures considered is the use of the antiretroviral treatment zidovudine (ZDV) by the mother during pregnancy and/or during labor and delivery. In all, four of 43 ZDV exposed infants were HIV infected compared with 16 of 54 ZDV unexposed infants. The conventional Pearson χ^2 statistic without continuity correction is 6.043, corresponding to a two-sided p-value of 0.014. The two-sided Fisher's exact test p-value is 0.022.

Although we know $S_{11}+S_{12}+S_{22}=5$, we have only partial information on the values of S_{11}, S_{12} and S_{22} from the paper. Clearly, ZDV exposure within each of the twin pairs must be concordant, although we do not know if each pair is exposed or unexposed, leading to the restriction $S_{11}+S_{22}\geq 2$. The paper states that ZDV was used in both pregnancies by at least one of the mothers with two singleton births, yielding $S_{11}\geq 1$.

Given these restrictions, the most extreme allocations of $\{S_{11},S_{12},S_{22}\}$ result from setting $\{S_{11}=5,S_{12}=0,S_{22}=0\}$, or at the other extreme, $\{S_{11}=1,S_{12}=3,S_{22}=1\}$. Table 2 shows for both these extremes, the p-values for the modified Pearson χ^2 and the modified Fisher's exact test over a range of values for ρ from –1.0 to 1.0. The $\rho=0$ column corresponds to the naïve analysis. The true (unknown) correlation is plausibly small and positive, although there are not sufficient data to evaluate this. However, even at the theoretical extremes ($\rho=\pm 1.0$) the p-values change very little, illustrating that the presence of a small number of correlated observations in this data set has only minimal impact on the statistical findings.

The estimated odds ratio relating HIV infection to ZDV exposure is 0.243 with 95% confidence interval (CI), assuming independence, of (0.075, 0.795). Assuming $\{S_{11}=5,S_{12}=0,S_{22}=0\}$ and a correlation of $\rho=.20$, the CI becomes (0.073, 0.812). With a correlation of $\rho=1.0$ the CI becomes (0.068, 0.879). Again, the correlation has only minimal impact on statistical findings.

Table 2. Modified Pearson chi square test for independence square p-value (top entry) and modified Fisher's exact test p-value (bottom values) for mother-to-child HIV transmission example.

	ρ						
$\{S_{11},S_{12},S_{22}\}$	-1.0	-0.5	-0.2	0.0	0.2	0.5	1.0
{5,0,0}	.008	.011	.013	.014	.015	.017	.021
	.014	.018	.020	.022	.024	.026	.031
{1,3,1}	.015	.014	.014	.014	.014	.013	.013
	.023	.023	.022	.022	.022	.021	.021

Simulation

Simulation results for configurations B and C are shown in Table 3. Because $N_1=N_2$, the properties of the different tests are nearly invariant to any allocation of concordant siblings to the exposed and unexposed groups, and hence results for configuration A (not shown) are very similar to configuration B. Both the modified tests perform well, although the modified Fisher's exact test appears to correct for correlation better than the modified Pearson chi square test in most situations studied.

Table 3. Simulation Results.

					Actual Test Size for Nominal α=.05 Test			
N[1]	θ[2]	Config[3]	ρ[4]	2S/N[5]	Pearson	Modified Pearson	Fisher	Modified Fisher
100	.1	B	.2	Low	.0493	.0470	.0496	.0473
100	.1	B	.8	Low	.0598	.0462	.0594	.0520
100	.1	B	.2	High	.0639	.0414	.0635	.0485
100	.1	B	.8	High	.1164	.0519	.1117	.0504
500	.1	B	.2	Low	.0523	.0523	.0537	.0523
500	.1	B	.8	Low	.0557	.0494	.0576	.0503
500	.1	B	.2	High	.0666	.0515	.0673	.0517
500	.1	B	.8	High	.1094	.0498	.1110	.0507
24	.5	B	.2	Low	.0687	.0687	.0541	.0502
24	.5	B	.8	Low	.0804	.0499	.0624	.0489
24	.5	B	.2	High	.0829	.0533	.0647	.0495
24	.5	B	.8	High	.1422	.0729	.1178	.0520
100	.5	B	.2	Low	.0571	.0571	.0500	.0480
100	.5	B	.8	Low	.0650	.0444	.0562	.0492
100	.5	B	.2	High	.0733	.0486	.0650	.0505
100	.5	B	.8	High	.1188	.0481	.1077	.0489
500	.5	B	.2	Low	.0557	.0451	.0508	.0489
500	.5	B	.8	Low	.0624	.0507	.0579	.0499
500	.5	B	.2	High	.0667	.0453	.0614	.0472
500	.5	B	.8	High	.1168	.0541	.1096	.0504

Table 3. (Continued)

					Actual Test Size for Nominal α=.05 Test			
N[1]	θ[2]	Config[3]	ρ[4]	2S/N[5]	Pearson	Modified Pearson	Fisher	Modified Fisher
100	.1	C	.2	Low	.0491	.0578	.0489	.0501
100	.1	C	.8	Low	.0440	.0529	.0440	.0517
100	.1	C	.2	High	.0373	.0532	.0377	.0523
100	.1	C	.8	High	.0065	.0635	.0079	.0585
500	.1	C	.2	Low	.0480	.0514	.0499	.0509
500	.1	C	.8	Low	.0403	.0474	.0413	.0487
500	.1	C	.2	High	.0357	.0504	.0367	.0511
500	.1	C	.8	High	.0053	.0502	.0053	.0498
24	.5	C	.2	Low	.0614	.0614	.0487	.0513
24	.5	C	.8	Low	.0577	.0577	.0443	.0535
24	.5	C	.2	High	.0456	.0488	.0352	.0519
24	.5	C	.8	High	.0058	.0497	.0057	.0582
100	.5	C	.2	Low	.0537	.0537	.0471	.0488
100	.5	C	.8	Low	.0489	.0490	.0428	.0501
100	.5	C	.2	High	.0412	.0424	.0364	.0507
100	.5	C	.8	High	.0070	.0614	.0061	.0510
500	.5	C	.2	Low	.0544	.0544	.0501	.0521
500	.5	C	.8	Low	.0508	.0508	.0474	.0549
500	.5	C	.2	High	.0392	.0507	.0361	.0509
500	.5	C	.8	High	.0055	.0465	.0047	.0495

[1]N: Total sample size

[2]θ: Probability of disease outcome

[3]Config: Configuration of concordant exposed, concordant unexposed and discordant sibling pairs (see Table 1)

[4]ρ: Within-family correlation

[5]2S/N: Number of siblings as a proportion of total sample size. Target low and high values of 2S/N are 0.08 and 0.64. With a small total sample size of N=24, it was not possible to achieve 2S/N=0.08 or 0.64 exactly. For example, in configuration A (Table 1), with one concordant pair, 2S/N=2/24=0.08333 instead of 0.08. Similarly, the actual values of 2S/N for configurations B and C were 0.1667 and 0.0833, respectively. Instead of 0.64, the values of 2S/N were 0.50, 0.6667 and 0.50, respectively, for configurations A, B and C. With N=100 or 500, the only combination where it was impossible to achieve the target values of 2S/N was for {2S/N=0.64, Configuration A}, where allocating 64% of the sample to the exposed group would make N_1 exceed N/2. Thus we used 2S/N=0.50 here.

As expected, the conventional tests tend to be anti-conservative when there are concordant siblings (configurations A and B) and conservative when there are discordant siblings. (configuration C). The conventional Pearson chi square test for independence and Fisher's exact tests perform well when there are <10% siblings in the data set ($2S/N \approx 0.08$), even with correlation as high as 0.8. The magnitude of conservatism or anti-conservatism increases with the correlation (ρ) and as the number of sibling pairs in the data set (S) increases.

Conclusion

We have presented modifications to the ordinary Pearson χ^2 test for independence and to Fisher's Exact test for the analysis of 2×2 tables when some of the observations are correlated. The methods achieve the desired properties across a wide range of possible data sets even with quite small sample sizes. Formulae for constructing modified confidence intervals are also provided.

Previous work has focused on unstratified and stratified 2×2 tables and clustered data where it is assumed that exposure status is common to all units in a cluster (i.e., no discordant pairs) (Donald & Donner, 1987; Donner, 1989; Rao & Scott, 1992; Rosner, 1982). This would occur, for example, if the exposure of interest was a genetic characteristic of the mother of children in a cluster. This assumption is not required in other research (Rosner & Milton, 1988; Begg, 1999) but these methods require enough clustered observations to allow the nature of the correlation to be estimated from the data.

Another possible approach to analysis would be to use a logistic regression model with correlation between siblings from the same family. Standard errors that take the correlation into account can be obtained using generalized estimating equations (Diggle, Liang & Zeger, 1994). Advantages of this modeling approach are that additional covariates can be added to the model, the covariates can be specific to each cluster unit and and the exposures of interest need not be dichotomous. However, its complexity is a problem and since it requires availability of the raw data it could not ordinarily be used to evaluate published results.

Our modified procedures require knowledge of the correlation, ρ, which would be difficult to estimate unless the number of pairs is large. However, by repeating the analysis over a range of possible values for ρ, one can assess the sensitivity of conclusions to the presence of correlation.

Determining a reasonable range of plausible values for ρ is difficult in part because correlations of binary variables have unusual properties. It is known that the correlation between binary variables is constrained by the true probabilities as follows (Prentice, 1988): If $\pi_1 < \pi_2$ then

$$\max\left[-\sqrt{\frac{\pi_1\pi_2}{(1-\pi_1)(1-\pi_2)}}, -\sqrt{\frac{(1-\pi_1)(1-\pi_2)}{\pi_1\pi_2}}\right] < \rho < \sqrt{\frac{\pi_1(1-\pi_2)}{\pi_2(1-\pi_1)}} \quad .$$

Estimates of π_1 and π_2 can therefore aid in setting bounds on ρ. Published results from analyses that naively assumed independence can easily be checked in such a sensitivity analysis, provided one is given enough information about the numbers of pairs in which both pair members are exposed, both are unexposed and exposure status is discordant. Unlike the value of ρ, these numbers would ordinarily be known when analyzing one's own data but may not be known when assessing the impact of non-independence on published results, in which case a range of possible numbers can be used in a sensitivity analysis.

Although the methods here are presented as for epidemiologic risk relations, they could also apply to clinical trials in which some (but not necessarily all) subjects have more than one "outcome," for example on two eyes in ophthalmologic studies, two legs in studies of walking impairment or multiple teeth in dental studies.

The procedures in this paper are most useful when there is a small amount of clustering so that the correlation cannot be reliably estimated, and when it is desired to evaluate the robustness of

conclusions to deviations from the assumption of independence.

In conclusion, it is important to recognize that non-independent observations, such as subjects within the same family, may make conventional statistical analyses based on independence assumptions prone to be conservative or anti-conservative. Simple correction methods, such as that described here for dichotomous exposure and outcome, are of value in ensuring that appropriately valid inferences are drawn when non-independent observations are present.

References

Agresti, A. (1996). *An introduction to categorical data analysis*. New York: John Wiley and Sons.

Begg M. D. (1999). Analyzing k (2 x 2) tables under cluster sampling. *Biometrics*, *55*, 302-307.

Cox D. R., Hinkley, D. V. (1974). *Theoretical statistics*. London: Chapman and Hall.

Dickover, R. E., Garratty, E. M., Herman, S. A., & et al. (1996). Identification of levels of maternal HIV-1 RNA associated with risk of perinatal transmission. Effect of maternal zidovudine treatment on viral load. *Journal of the American Medical Association, 275*, 599-605.

Diggle, P. J., Liang, K-Y., & Zeger, S. L. (1994). *Analysis of longitudinal data.* New York: Oxford University Press.

Donald, A., & Donner, A. (1987). Adjustments to the Mantel-Haenszel chi-square statistic and odds ratio variance estimator when the data are clustered. *Statistics in Medicine, 6*, 491-499.

Donner, A. (1989). Statistical methods in ophthalmology: An adjusted chi-square approach. *Biometrics*, *45*, 605-611.

Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, *44*, 1033-1048.

Rao J. N. K., & Scott, A. J. (1992). A simple method for the analysis of clustered binary data. *Biometrics*, *48*, 577-585.

Rosner, B., & Milton, R. C. (1988). Significance testing for correlated binary outcome data. *Biometrics, 44*, 505-512.

Rosner, B. (1997). Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. *Biometrics*, *38*, 105-114.

Zhang, J., & Boos, D. D. (1997). Mantel-Haenszel test statistics for correlated binary data. *Biometrics*, *53*, 1185-1198.