Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 3070–3084
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Statistical inference for the risk ratio in 2×2 binomial trials with structural zero

Ramesh C. Gupta, Suzhong Tian*

Department of Mathematics and Statistics, 5752 Neville Hall, University of Maine, Orono, ME 04469-5752, USA

Received 3 February 2006; accepted 3 February 2006
Available online 28 February 2006

Abstract

In some statistical analyses, researchers may encounter the problem of analyzing a correlated 2×2 table with a structural zero in one of the off diagonal cells. Structural zeros arise in situation where it is theoretically impossible for a particular cell to be observed. For instance, Agresti [2002. Categorical Data Analysis, 2nd ed. Wiley, Hoboken, New Jersey] provided an example involving a sample of 156 calves born in Okeechobee County, Florida. Calves are first classified according to whether they get a pneumonia infection within certain time. They are then classified again according to whether they get a secondary infection within a period after the first infection clears up. Because subjects cannot, by definition, have a secondary infection without first having a primary infection, a structural void in the cell of the summary table that corresponds with no primary infection and has secondary infection is introduced. For discussion of this phenomenon, see Tang and Tang [2002. Exact unconditional inference for risk ratio in a correlated 2×2 table with structural zero. Biometrics 58, 972–980], and Lui [1998. Interval estimation of the risk ratio between a secondary infection, given a primary infection, and the primary infection. Biometrics 54, 706–711].

The risk ratio (RR) between the secondary infection, given the primary infection, and the primary infection may be a useful measure of change in the pneumonia infection rates of the primary infection and the secondary infection. In this paper, we first develop and evaluate the large sample confidence intervals of RR. In addition to the three confidence intervals in the literatures, we propose a confidence interval based on Rao's score test. The performance of these confidence intervals is studied by means of extensive simulation studies. We also investigate the tests of hypothesis for the RR and the power of these tests. Simulation studies are carried out to examine the performance of these tests in terms of their power. An example, from the literature, is also provided to illustrate these procedures. Finally, the four confidence intervals were compared with those obtained by corrected version of Tang et al. [2004. Confidence interval for rate ratio in a 2×2 table with structural zero: an application in assessing false-negative rate ratio when combining two diagnostic tests. Biometrics 60, 550–555; 2006. Sample size determination for 2-step studies with dichotomous response. Journal of Statistical Planning and Inference 136, 1166–1180].

Keywords: Confidence interval; Test of hypothesis; Score test; Power of tests; Simulation studies

1. Introduction

In order to compare two groups, statistical inference of the risk ratio, under independent binomial sampling, has been extensively discussed in the literature (Gart and Nam, 1988). However, there are situations in which the assumption of

* Corresponding author. Tel.: +1 2075813915; fax: +1 2075813902.
E-mail address: suzhong_tian@umit.maine.edu (S. Tian).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2006.02.003

Table 1
Probability of each cell

		Secondary infection		Total
		Yes	No	
Primary infection	Yes	p_{11}	p_{12}	$p_{1\bullet}$
	No	0	p_{22}	p_{22}
Total		p_{11}	$p_{12}+p_{22}$	1

independent binomial sampling is not valid. Agresti (2002) has given an example in which calves were first classified according to whether they got primary infection and then reclassified according to whether they developed a secondary infection within a certain time period after the first infection cleared up. In this case, when assessing the risk ratio between a secondary infection, given a primary infection and the primary infection, the responses are taken from the same group of subjects and are not independent. Therefore, the statistical procedures, under independent binomial sampling are not appropriate. This kind of data can be summarized in Table 1.

Notice that calves having no primary infection cannot have secondary infection and hence the frequency of such event is zero in the above table. This is known as structural zero as opposite to sampling zero. SeeAgresti (2002, p. 392) for discussion and explanation.

In order to analyze such bivariate tables, Lui (2000) discussed the interval estimation of the simple difference between the proportion of the primary infection and the secondary infection, given the primary infection. He developed three asymptotic interval estimators using Wald's test statistic, the likelihood ratio test and the basic principle of Fieller's theorem. The simulation studies concluded that the asymptotic confidence interval using likelihood ratio test consistently perform well in all the situations.

On the other hand, Lui (1998) discussed the estimation of the risk ratio (RR) between a secondary infection, given a primary infection, and the primary infection. He developed three asymptotic interval estimators using Wald's test statistic, the logarithmic transformation, and Fieller's theorem. On the basis of his simulation studies, he concluded that when the underlying probability of primary infection is large, all three estimators perform reasonably well. When the probability of primary infection is small or moderate, the interval estimator using the logarithmic transformation outperforms the other two estimators when the sample size does not exceed 100. In addition, the coverage probability of this estimator consistently exceeds the nominal value in all situations.

In addition to the references cited above, Tang and Tang (2002) studied small sample statistical inference for RR in a correlated 2×2 table with a structural zero in one of the off diagonal cells.

The purpose of the present investigation is to further study the statistical inference in the case of 2×2 correlated table with a structural zero. In Section 2, we review the three confidence intervals of RR studied by Lui (1998) and derive a fourth confidence interval based on Rao's score test. An example is provided to compare the results. Simulation studies are carried out in Section 3 to compare the performance of these four confidence intervals in terms of the coverage probability and the length of the confidence interval. The coverage probabilities of confidence interval by Rao's score and logarithmic transformation methods are both close to the confidence level. Comparing the lengths of score based and log method, the score based confidence intervals are shortest. In addition to these, Section 3 also contains the comparison of these four confidence intervals with the corrected version of Tang et al. (2004, 2006).

In Sections 4 and 5, we derive the Rao's score test for testing H_0: RR $= 1$ and compare the four tests with respect to the power by means of extensive simulation studies. The simulation studies suggest the Rao's score method is more consistent than the other three methods although it is not the most powerful test. Actually, there is no consistent most powerful test in this study.

Finally, in Section 6, we present some conclusion and comments.

2. Derivation of confidence intervals

In this section, we will illustrate the detailed steps to derive four asymptotic estimators of RR. Three of these were proposed by Lui (1998), and are based on Wald's test statistic, the logarithmic transformation, and Fieller's theorem.