Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 1166–1180
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Sample size determination for 2-step studies with dichotomous response

Man-Lai Tang[a,*], Nian-Sheng Tang[b], Vincent J. Carey[c]

[a] *Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong*
[b] *Research Center of Applied Statistics, Yunnan University, Kunming 650091, PR China*
[c] *Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA*

Received 18 November 2003; accepted 14 July 2004
Available online 28 October 2004

Abstract

In this article, we consider problems with correlated data that can be summarized in a 2×2 table with structural zero in one of the off-diagonal cells. Data of this kind sometimes appear in infectious disease studies and two-step procedure studies. We propose two kinds of approximate sample size formulas, based on rate ratio, for comparison of the marginal and conditional probabilities in a correlated 2×2 table with structural zero. The first type of formula is derived to guarantee a pre-specified power of a hypothesis test at certain significance level while the second type of formula is developed to bound the width of a confidence interval with specified confidence level. Our empirical studies confirm that sample size formulas based on the log-transformation and score tests outperform that based on the Wald's test. We illustrate our methodologies with a real example from a two-phase treatment study.

MSC: Primary 62P10; Secondary 62P99

Keywords: Asymptotic inference; Correlated binary data; Rate ratio; Score test; Structural zero

1. Introduction

In studies of repeated exposures to infectious agents, researchers may be interested in assessing the effect due to the primary infection on the risk of developing secondary infection

* Corresponding author.
E-mail address: mltang@math.hkbu.edu.hk (M.-L. Tang).

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.07.016

Table 1
Observed frequencies for two stage treatment regimes for severely ill patients

Phase I	Phase II	
	No improvement	Improvement
No improvement	6	21
Improvement	—	43

(Agresti, 1990). Subjects who do not suffer a primary infection from a certain infectious disease cannot develop a secondary infection from the same disease. If the outcome of interest is binary (e.g., infected or not infected), the resultant data can usually be summarized in a correlated 2×2 table with a structural zero in one of the off-diagonal cells. Johnson and May (1995) reported another scenario in which a treatment regimen for a particular disease was administered to qualified patients in two stages. After the initial stage, researchers recorded the patient's status as either improved or not improved. Patients were initiated to the second phase and were asked to return for assessment *only if* the initial treatment provided no improvement. Since the researchers recorded the patient's response to the second phase of the treatment *only if* the initial response was negative, they introduced a structural zero in the cell of the summary table that corresponds with positive response for the first treatment and negative response for the second treatment. The corresponding data for patients with initial disease status being severe are reported in Table 1.

Let π_{11} be the probability associated with lack of improvement after both phases of the treatment and π_{1+} be the marginal probability of no improvement after the first phase. In this case, π_{11}/π_{1+} represents the conditional probability that a patient shows no improvement in the second phase of the treatment, given that the same patient shows no improvement in the first phase of the treatment. One has evidence to believe that the second phase of treatment is beneficial if the rate ratio $\pi_{11}/\pi_{1+}^2 (=[(\pi_{11}/\pi_{1+})/\pi_{1+}])$ is less than one (Johnson and May, 1995). Other examples include two-stage treatment studies (Johnson and May, 1995) and two-step tuberculosis testing studies (Tang and Tang, 2002, 2003).

Statistical tools for analyzing correlated 2×2 tables with structural zero have been developed by Lui (1998, 2000), Tang and Tang (2002, 2003), and Tang et al. (2004). Currently, comparative measures under study include the risk/rate difference (RD) and the risk/rate ratio (RR). Lui (1998) investigated the confidence interval estimators for the risk ratio based on the Wald's test, the logarithm transformation test and the test according to Fieller's theorem. Asymptotic confidence intervals obtained from the logarithm transformation test were found by simulation to consistently outperform the other two interval procedures in terms of coverage probability and expected width. Lui (2000) then studied the confidence interval estimators for the simple risk difference. Empirical study showed that the confidence interval estimator based on the likelihood ratio test consistently performed well in various situations. Tang and Tang (2003) revisited the work by Lui (2000) and derived the score statistic for testing the null hypothesis of non-zero risk difference. Their simulation study showed that the score-test-based confidence interval estimator and the likelihood-ratio-test-based confidence interval estimator perform equally well in most of the small sample settings. Most