Testing the Ratio of Two Poisson Rates

Kangxia Gu[1], **Hon Keung Tony Ng**[*1], **Man Lai Tang**[2], and **William R. Schucany**[1]

[1] Department of Statistical Science, Southern Methodist University, Dallas, Texas 75275-0332, USA

[2] Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong

Summary

In this paper we compare the properties of four different general approaches for testing the ratio of two Poisson rates. Asymptotically normal tests, tests based on approximate p-values, exact conditional tests, and a likelihood ratio test are considered. The properties and power performance of these tests are studied by a Monte Carlo simulation experiment. Sample size calculation formulae are given for each of the test procedures and their validities are studied. Some recommendations favoring the likelihood ratio and certain asymptotic tests are based on these simulation results. Finally, all of the test procedures are illustrated with two real life medical examples.

Key words: Asymptotic tests, Conditional test, Constrained maximum likelihood estimation, Level of significance, Mid-p; Monte Carlo simulation, Power.

1 Introduction

The Poisson Distribution, named after the famous French mathematician Simeon Denis Poisson who first formulated it (Poisson, 1837), is a mathematical rule that assigns probabilities to the number of occurences. It is most commonly used to model the number of random occurrences of some phenomenon in a specified space or time interval. Sometimes the Poisson rate is of interest because it pertains to a unit of time or

[*] Corresponding author: e-mail: ngh@mail.smu.edu, Phone: +1-214-768-2465, Fax: +1-214-768-4035

space. For example, the incidence rate of a disease (Rothman and Greenland, 1998) is defined as the number of events observed divided by the time at risk during the observation period.

Biological, epidemiological, and medical research can produce data that follow a Poisson distribution. It is sometimes called the law of rare events, for it describes the distribution of counts of such events. For example, the number of cases of a rare disease or the frequency of single gene mutations may be modeled by a Poisson distribution. In two-sample situations, researchers are likely to be interested in testing the difference or ratio of the Poisson means. The comparison of the Poisson rates from two independent samples is clearly of medical interest. For instance, Stampfer and Willett (1985) conducted a prospective study to examine the relationship of post-menopausal hormone use and coronary heart disease (CHD). With postmenopausal hormone use in 54308.7 person-years, there are 30 CHD cases; without postmenopausal hormone use in 51477.5 person-years, there are 60 CHD cases. The problem of interest is to test whether the post-menopausal hormone using group has less coronary heart disease. Another example in epidemiology is a breast cancer study reported in Rothman and Greenland (1998) and Graham, Mengersen, and Morton (2003). Two groups of women were compared to determine whether those who had been examined using x-ray fluoroscopy during treatment for tuberculosis had a higher rate of breast cancer than those who had not been examined using x-ray fluoroscopy. Forty-one cases of breast cancer in 28,010 person-years at risk are reported in the treatment group with women receiving x-ray fluoroscopy and 15 cases of breast cancer in 19,017 person-years at risk in the control group with women not receiving x-ray fluoroscopy.

The problem of comparing two Poisson rates has been studied for a long time, however, most of the early studies focused on the equal time frame situation. Przyborowski and Wilenski (1939) first proposed a conditional test for the unity of the ratio of two Poisson rates based on a conditional binomial distribution. Subsequently, Chapman (1952), Birnbaum (1953), Brownlee (1967) and Gail (1974) studied the properties and some alternatives for testing the hypothesis or constructing confidence intervals based on the conditional approach. On the other hand, Hald (1960), Ractliffe (1964), Cox and Lewis (1966), Haight (1967)

Detre and White (1970) and Sichel (1973) investigated asymptotic tests based on normal approximations for the equality of the two Poisson rates.

Development for unequal sampling frames has received more attention recently. Shiue and Bain (1982) derived a uniformly most powerful unbiased (UMPU) test for equality of two Poisson rates and showed that a test based on the normal approximation of the binomial distribution is nearly as powerful as the UMPU test. Huffman (1984) proposed an improved asymptotic test statistic, which accelerated the rate of convergence to normality by a variance stabilizing transformation. Along the same line, Thode (1997) considered an alternative test statistic that is more powerful than Shiue and Bain statistic for large Poisson rates. Recently, Ng and Tang (2005) provided systematic comparisons among these tests and presented the associated sample size formulae. All these studies are devoted to the problem of testing the equality of the two Poisson rates (or the unity of the ratio) only. Krishnamoorthy and Thomson (2004) considered tests based on estimated p-values and showed that these have uniformly better power than the conditional test.

Although many methods have been proposed to test the ratio of two Poisson rates, it is not clear when techniques are more appropriate, or how their performances might vary. Moreover, most of the works focus on testing the unity of the ratio. In the present article, we present an extensive and systematic comparative study of different test procedures for testing the non-unity of the rate ratio of two Poisson processes under unequal sampling frames. In Section 2, we review the problem of testing the general non-unity ratio of two Poisson rates over unequal-size sampling frames. Asymptotic tests based on normal approximations, tests based on approximate p-value methods, an exact conditional test, a mid-p conditional test, and a likelihood ratio test are considered. In Section 3, we present the formulae for sample size calculation for each of the test procedures. In Section 4, Monte Carlo simulations are used to study the properties of the different test procedures, as well as the validity of the sample size calculation formulae. Two real life examples are used to illustrate the test procedures in Section 5. Finally, a discussion of the results with some recommendations are presented in Section 6.

2 Test Procedures

2.1 Asymptotic tests based on normal approximations

For fixed sampling frames t_0 and t_1, two independent Poisson processes (with parameters λ_0 and λ_1) are observed. Let X_0 and X_1 be the corresponding number of outcomes, i.e. $X_i \sim$ Poisson(λ_i) with $\lambda_i = t_i\gamma_i$ for $i = 0, 1$. Here, γ_i, $i = 0, 1$ are the Poisson rates. We denote the observed values of X_0 and X_1 by x_0 and x_1, respectively. We are interested in testing that the ratio of two Poisson rates is equal to a pre-specified positive number R versus greater than R, i.e. we are testing the following one-sided hypotheses

$$H_0 : \gamma_0/\gamma_1 = R \text{ against } H_1 : \gamma_0/\gamma_1 > R. \tag{1}$$

When $R = 1$, it is equivalent to testing the equality of the two Poisson rates, i.e. $\gamma_0 = \gamma_1$. The properties of different test procedures for $R = 1$ under unequal sampling frames, i.e. $t_0 \neq t_1$, were investigated in Ng and Tang (2005).

Since γ_i, $i = 0, 1$ are the unknown parameters, two sample estimates for γ_i are considered as follows.

- Unconstrained Maximum Likelihood Estimate (MLE)

  The maximum likelihood estimate of γ_i is given by

  $$\hat{\gamma}_i = \frac{X_i}{t_i}, i = 0, 1. \tag{2}$$

- Constrained Maximum Likelihood Estimate (CMLE)

  Under the null hypothesis H_0: $\gamma_0/\gamma_1 = R$, the CMLEs of γ_0 and γ_1 can be shown to be (see Appendix A)

  $$\tilde{\gamma}_0 = \frac{X_0 + X_1}{t_0(1 + 1/\rho)} \text{ and } \tilde{\gamma}_1 = \frac{X_0 + X_1}{t_1(1 + \rho)}, \tag{3}$$

  where $d = t_1/t_0$ and $\rho = R/d$. Note that the CMLEs of γ_0 and γ_1 are the same as the method of moment estimator under the null hypothesis $H_0 : \gamma_0/\gamma_1 = R$. Specifically, $E(X_0+X_1) = t_0\gamma_0+t_1\gamma_1$ with $\gamma_0/\gamma_1 = R$ yields the estimators in (3).

Because we can re-express the null hypothesis in (1) as $H_0 : \gamma_0 - R\gamma_1 = 0$, we can develop the test statistics based on the statistic $W = \hat{\gamma}_0 - R\hat{\gamma}_1$. The variance of the statistic W is given by

$$\sigma_W^2 = \frac{\gamma_0}{t_0} + \frac{R^2\gamma_1}{t_1}$$

and it can be estimated by

$$s_W^2 = \frac{\gamma_0^*}{t_0} + \frac{R^2\gamma_1^*}{t_1},$$

where γ_i^* is any reasonable estimate of γ_i, $i = 0, 1$. Hence, we consider the test statistic W/s_W for H_0.

With MLEs $\gamma_i^* = \hat{\gamma}_i$, $i = 0, 1$, we get

$$W_1(X_0, X_1) = \frac{X_0/t_0 - RX_1/t_1}{\sqrt{X_0/{t_0}^2 + X_1R^2/{t_1}^2}} = \frac{X_0 - X_1\rho}{\sqrt{X_0 + X_1\rho^2}}.$$

Similarly, with CMLEs $\gamma_i^* = \tilde{\gamma}_i$, $i = 0, 1$, we obtain

$$W_2(X_0, X_1) = \frac{X_0/t_0 - RX_1/t_1}{\sqrt{(X_0 + X_1)(R^2/d + R)/(R + d)}} = \frac{X_0 - X_1\rho}{\sqrt{(X_0 + X_1)\rho}}.$$

Note that when $\rho = 1$, W_1 is equivalent to W_2 and is the test statistic studied by Shiue and Bain (1982) and Thode (1997).

On the other hand, testing the null hypothesis in (1) is equivalent to testing $\ln(\gamma_0/\gamma_1) - \ln(R) = 0$. Hence, we consider the statistic $U = \ln(\hat{\gamma}_0/\hat{\gamma}_1) - \ln(R) = \ln(X_0/t_0) - \ln(X_1/t_1) - \ln(R)$. By the delta method, the variance of U can be approximated by $\sigma_U^2 = 1/(t_0\gamma_0) + 1/(t_1\gamma_1)$ and it can be estimated by $s_U^2 = 1/(t_0\gamma_0^*) + 1/(t_1\gamma_1^*)$, where γ_i^* is any reasonable estimate of γ_i, $i = 0, 1$. Hence, we can consider the test statistic U/s_U to test (1).

For MLEs $\gamma_i^* = \hat{\gamma}_i$, $i = 0, 1$, we have the test statistic

$$W_3(X_0, X_1) = \frac{\ln(X_0/X_1) + \ln(d) - \ln(R)}{\sqrt{1/X_0 + 1/X_1}} = \frac{\ln(X_0/X_1) - \ln(\rho)}{\sqrt{1/X_0 + 1/X_1}}.$$

Similarly, for CMLEs $\gamma_i^* = \tilde{\gamma}_i$, $i = 0, 1$, we obtain

$$W_4(X_0, X_1) = \frac{\ln(X_0/X_1) + \ln(d) - \ln(R)}{\sqrt{(2 + d/R + R/d)/(X_0 + X_1)}} = \frac{\ln(X_0/X_1) - \ln(\rho)}{\sqrt{(2 + 1/\rho + \rho)/(X_0 + X_1)}}.$$

As the test statistics W_3 and W_4 are not defined when $X_0 = 0$ and/or $X_1 = 0$, we set $X_i = 0.5$ whenever $X_i = 0$, $i = 0, 1$.

We also construct a test statistic by considering a variance stabilizing transformation suggested by Huffman (1984) which can accelerate the rate of convergence to normality as $\lambda = \min(\lambda_0, \lambda_1)$ goes to infinity. The test statistic is given by

$$W_5(X_0, X_1) = \frac{2\left[\sqrt{X_0 + 3/8} - \sqrt{\rho(X_1 + 3/8)}\right]}{\sqrt{1+\rho}}.$$

Huffman (1984) proposed to add the terms 3/8, motivated by the results in Anscombe (1984), in order to reduce the variance of the test statistic W_5 from 1 + $O(\lambda^{-1})$ to 1 + $O(\lambda^{-2})$.

By observing the forms of the test statistics, we notice that their performances depend on the value of $\rho = R/d$ but not on R or d individually. Let us denote the observed value of the test statistics $W_j(X_0, X_1)$ as $w_j(x_0, x_1)$ for $j = 1, \ldots, 5$. Under the null hypothesis, W_j can be shown to be asymptotically distributed as a standard normal. Hence, the p-values of the asymptotic tests are given by

$$p_j^{(A)} = 1 - \Phi(w_j(x_0, x_1)),\ j = 1, \ldots, 5,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal. We reject H_0 when $p_j^{(A)} < \alpha$, where α is the prefixed level of significance of the test.

2.2 Tests based on numerical approximations to exact p-value

The right-tailed significance probabilities for testing (1) based on test statistics W_j under H_0 are

$$\begin{aligned}&\Pr(W_j \geq w_j(x_0, x_1)|H_0)\\ &\quad = \sum_{y_0=0}^{\infty}\sum_{y_1=0}^{\infty} \frac{e^{-t_0(\gamma_1 R)}[t_0(\gamma_1 R)]^{y_0}}{y_0!}\frac{e^{-t_1\gamma_1}(t_1\gamma_1)^{y_1}}{y_1!} I[W_j(y_0, y_1) \geq w_j(x_0, x_1)], \qquad (4)\end{aligned}$$

$j = 1, \ldots, 5$, where $I[\cdot]$ denotes the indicator function. Since the right-tailed probability involves the unknown parameter γ_1, we can approximate the p-value by evaluating the right-tailed probability at a reasonable value of γ_1. If we replace γ_1 by the CMLE $\tilde{\gamma}_1$ in (3), the p-value can be estimated by

$$p_j^{(P)} = \sum_{y_0=0}^{\infty}\sum_{y_1=0}^{\infty} \frac{e^{-t_0(\tilde{\gamma}_1 R)}[t_0(\tilde{\gamma}_1 R)]^{y_0}}{y_0!}\frac{e^{-t_1\tilde{\gamma}_1}(t_1\tilde{\gamma}_1)^{y_1}}{y_1!} I[W_j(y_0, y_1) \geq w_j(x_0, x_1)],\ j = 1, \ldots, 5.$$

We reject H_0 in (1) when $p_j^{(P)} < \alpha$.

2.3 Exact conditional tests

The classical method of testing the ratio of two Poisson rates is the conditional method, which was first proposed by Przyborowski and Wilenski (1939). The conditional distribution of X_0 given $X_0 + X_1 = k$ follows a binomial distribution with success probability $q = (t_0\gamma_0)/(t_0\gamma_0 + t_1\gamma_1)$. Under $H_0 : \gamma_0/\gamma_1 = R$, X_0 given $X_0 + X_1 = k$ is binomial with the number of trials k and the success probability $q_0 = \rho/(1+\rho)$. The null hypothesis H_0 is rejected if the exact p-value

$$p^{(C)} = \Pr(X_0 \geq x_0 \mid x_0 + x_1 = k, H_0) = \sum_{i=x_0}^{k} \binom{k}{i} q_0^i (1-q_0)^{k-i} < \alpha. \tag{5}$$

It is well known that the traditional conditional test is an exact method for which the actual level of significance is always below the nominal level (i.e. conservative). To overcome the conservativeness of the traditional conditional test, the mid-p adjusted version test first suggested by Lancaster (1952, 1961) can be considered. The mid-p adjustment generally corrects the conservativeness of an exact method, while at the same time its actual significance level is close to the pre-specified nominal level, as mentioned in Tang (1998). The p-value for the mid-p correction is given by

$$p^{(M)} = \frac{p^{(C)} + p^{(C^*)}}{2},$$

where $p^{(C^*)} = \Pr(X_0 \geq x_0 + 1 | x_0 + x_1 = k, H_0) = \sum_{i=x_0+1}^{k} \binom{k}{i} q_0^i (1-q_0)^{k-i}$. We reject the null hypothesis H_0 if $p^{(M)} < \alpha$.

2.4 Likelihood ratio test

A likelihood-ratio test (LRT) relies on a test statistic that is the ratio of the maximum value of the likelihood function under the constraint of the null hypothesis to the maximum likelihood with that constraint being relaxed. Here, the likelihood function is

$$L(\gamma_0, \gamma_1) = \frac{e^{-t_0\gamma_0}(t_0\gamma_0)^{x_0}}{x_0!} \frac{e^{-t_1\gamma_1}(t_1\gamma_1)^{x_1}}{x_1!}.$$

The maximum likelihood under the whole parameter space $\Omega = \{\gamma_0, \gamma_1 | \gamma_0 > 0, \gamma_1 > 0\}$ is given by Dykstra and Robertson (1982). Under the null hypothesis, the maximum likelihood under the restricted parameter space $\Omega^* = \{\gamma_0, \gamma_1 | \gamma_0 > 0, \gamma_1 > 0, \gamma_0/\gamma_1 = R\}$ is $L(\tilde{\gamma}_0, \tilde{\gamma}_1)$. Therefore, the likelihood ratio test statistic

$$\Lambda = \frac{\sup_{(\gamma_0,\gamma_1)\in\Omega^*} L(\gamma_0, \gamma_1)}{\sup_{(\gamma_0,\gamma_1)\in\Omega} L(\gamma_0, \gamma_1)} = \begin{cases} \frac{L(\tilde{\gamma}_0,\tilde{\gamma}_1)}{L(\hat{\gamma}_0,\hat{\gamma}_1)}, & \text{if } \hat{\gamma}_0/\hat{\gamma}_1 > R \\ 1, & \text{if } \hat{\gamma}_0/\hat{\gamma}_1 \leq R. \end{cases}$$

An equivalent LRT statistic is $-2\ln\Lambda$, which has a particularly handy asymptotic distribution. If the null hypothesis is true, then asymptotically $-2\ln\Lambda$ will be zero half of the time and χ^2 distributed with one degree of freedom the other half of the time (Robertson , Wright and Dykstra, 1988). Therefore, the approximate p-value for the LRT is

$$p^{(L)} = 0.5[1 - \chi_1^2(-2\ln\Lambda)],$$

where $\chi_1^2(\cdot)$ is the cumulative distribution function of the chi-square distribution with one degree of freedom. For a given level of significance α, we reject H_0 in (1) when $p^{(L)} < \alpha$.

3 Sample Size Calculation

In experimental design one is concerned about the experimental lengths (i.e. t_0 and t_1) required to achieve a specified power. To attain a power $1-\beta$ under $H_1 : \gamma_0/\gamma_1 = R^{'} > R$ at α level, given α, β, d, R, and $R^{'}$, the criterion on λ_1 for the test procedure based on $p_1^{(A)}$ is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho + c^2)(z_{1-\alpha} + z_{1-\beta})^2}{(1-c)^2}, \tag{6}$$

where $\rho = R/d$ and $c = R/R'$. For any given value of γ_1 (usually based on prior information from previous experiments/studies), we can compute the require experimental lengths t_0 and t_1 from (6) and $t_1 = dt_0$. For the test procedure based on $p_2^{(A)}$, the criterion on λ_1 is

$$\lambda_1 = \gamma_1 t_1 = \frac{(c/\rho + c^2)\left[z_{1-\alpha}\sqrt{(c+\rho)/(1+c\rho)} + z_{1-\beta}\right]^2}{(1-c)^2}. \tag{7}$$

For the test procedure based on $p_3^{(A)}$, the criterion on λ_1 is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho+1)(z_{1-\alpha}+z_{1-\beta})^2}{(\ln c)^2}. \tag{8}$$

For the test procedure based on $p_4^{(A)}$, the criterion on λ_1 is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho+1)\left\{z_{1-\alpha}\left[\frac{\sqrt{c(\rho^2+2\rho+1)}}{c+\rho}\right]+z_{1-\beta}\right\}^2}{(\ln c)^2}. \tag{9}$$

Note that the formulae (6) - (9) will reduce to those provided in Ng and Tang (2005) when $R = 1$.

For the test procedure based on $p_5^{(A)}$, the criterion on λ_1 is

$$\lambda_1 = \gamma_1 t_1 = \left[\frac{z_{1-\alpha}\sqrt{c/\rho+c}+z_{1-\beta}\sqrt{1+c/\rho}}{2\left(1-\sqrt{c}\right)}\right]^2 - \frac{3}{8}. \tag{10}$$

When $R = 1$, (10) reduces to the formula provided in Huffman (1984). The derivations of formulae (6) - (10) are presented in Appendix B.

If tests based on numerically approximate p-values (i.e. test procedures based on $p_j^{(P)}$, $j = 1, \ldots, 5$) are used, the required experimental lengths (t_0 and t_1) can be computed by solving for t_1 in the following equation to attain the given power $1-\beta$ at α level of significance

$$\begin{aligned} 1-\beta &= \sum_{k_0=0}^{\infty}\sum_{k_1=0}^{\infty}\frac{e^{-t_0\gamma_0}(t_0\gamma_0)^{k_0}}{k_0!}\frac{e^{-t_1\gamma_1}(t_1\gamma_1)^{k_1}}{k_1!} \\ &\quad \times I\left[\sum_{x_0=0}^{\infty}\sum_{x_1=0}^{\infty}\frac{e^{-t_0\tilde{\gamma}_0}(t_0\tilde{\gamma}_0)^{x_0}}{k_0!}\frac{e^{-t_1\tilde{\gamma}_1}(t_1\tilde{\gamma}_1)^{x_1}}{k_1!}I(W_j(x_0,x_1)\le W_j(k_0,k_1))<\alpha\right]. \end{aligned} \tag{11}$$

Similarly, the required experimental lengths to attain the given power $1-\beta$ at α level of significance for the exact conditional test, mid-p conditional test, and the LRT can be computed by solving t_1 in the following equation

$$1-\beta = \sum_{k_0=0}^{\infty}\sum_{k_1=0}^{\infty}\frac{e^{-t_0\gamma_0}(t_0\gamma_0)^{k_0}}{k_0!}\frac{e^{-t_1\gamma_1}(t_1\gamma_1)^{k_1}}{k_1!}I[p^*<\alpha], \tag{12}$$

where $p^* = p^{(C)}, p^{(M)}$, and $p^{(L)}$, respectively.

4 Monte Carlo Simulation Studies

Monte Carlo simulation is used to study the performance and properties of different test procedures as well as the validity of the sample size calculation formulae. The Monte Carlo simulation was coded in Fortran. We used the International Mathematical and Statistical Libraries (IMSL) subroutine RNPOI to generate the Poisson samples X_0 and X_1. For the study of type-I error rate, we consider $d = t_1/t_0 = 0.5, 1.0, 1.5$ and 2.0, $\gamma_1 =$ from 1 to 60 in increments of 1 (denote as 1(1)60), $\alpha = 0.05$, and $R = 0.5, 1.0, 1.5$. These d-values include the small d-values (i.e., $d = 0.5$), balanced ($d = 1.0$), and large d-values (1.5 and 2.0). These settings give ρ-values = 1/4, 1/3, 1/2, 2/3, 3/4, 1, 3/2, 2 and 3. Without loss of generality, we assume $t_0 = 1$. For each X_0 and X_1, we compute the values of $p_j^{(A)}$, $p_j^{(P)}$, $j = 1, 2, 3, 4, 5$ and $p^{(C)}$, $p^{(M)}$ and $p^{(L)}$ and reject the null hypothesis if the p-value is less than α. This process is repeated 10,000 times for each combination of d, R and γ_1.

For the power calculation, X_0 and X_1 were independently generated from Poisson($R'\gamma_1$) and Poisson($d\gamma_1$), respectively. We examine the power under the alternative hypothesis $\gamma_0/\gamma_1 = R^{'}$. Here R' ranges from 1 to 15 for each combination of R and d. The estimated power for $p_j^{(A)}$, $p_j^{(P)}$, $j = 1, 2, 3, 4, 5$ and $p^{(C)}$, $p^{(M)}$ and $p^{(L)}$ is computed by the number of rejections of H_0 divided by 10,000 with standard error $SE = 0.0022$.

Moreover, we also examine the accuracy of the sample size formulae in (6) - (12). The required sample sizes (sample time frames) computed from these formulae with $\alpha = 0.05$, $\beta = 0.10$ are used to obtain the estimated significance level and power again with $10,000$ replications.

5 Results and Discussions

In this section we present the results of the Monte Carlo study and compare different test procedures in terms of the level of significance and the power. Since the pattern of the simulated significance level and power values are related to the ratio of R and d ($\rho = R/d$), we present the results in terms of ρ which ranges from 0.25 to 3.

5.1 Level of significance

First, consider the levels of the significance of various test procedures. It is a desirable property of a test procedure that its actual level of significance be close to pre-specified nominal level. To examine these levels, we plot the simulated significance levels for all the test procedures considered here. For the sake of saving space, we presented only the plots of the simulated significance levels at $\rho = 0.25$ and $\rho = 2.0$ for $p_j^{(A)}$, $j = 1, \ldots, 5$ in Figure 1. The plots of the simulated levels for other test procedures show similar patterns. From these plots, we observe that the pattern of the simulated significance levels are different for small values of γ_1 (say, $\gamma_1 < 10$) and large values of γ_1 (≥ 10). For $\gamma_1 < 10$, most of the test procedures are conservative except $p_2^{(A)}$ and $p_3^{(A)}$ are liberal for $\rho < 1$ and $p_1^{(A)}$ and $p_4^{(A)}$ are liberal for $\rho > 1$. Based on the plots, we observe that among all the test procedures, those based on $p_5^{(A)}$, $p_j^{(P)}$, $j = 1, \ldots, 5$ and $p^{(L)}$ have simulated significance levels close to the nominal levels when $\gamma_1 < 10$.

For $\gamma_1 \geq 10$, to scrutinize the levels of the test procedures, we compute the percentage of configurations (based on 50 different γ_1 = 11(1)60) that are conservative (i.e., simulated significance level $\leq$ 0.04), or liberal (i.e., simulated significance level $\geq$ 0.06) for all the test procedures and each given value of ρ. These values are presented in Tables 1 and 2. We also present the percentages of configurations in the intervals (0.04, 0.05), (0.05, 0.06) and $0.05 \pm SE = (0.0478, 0.0522)$. Among all five asymptotic tests, the one based on $p_5^{(A)}$ gives simulated significance levels closest to 0.05 for all values of ρ. For $\rho \leq 2/3$, $p_1^{(A)}$ and $p_4^{(A)}$ are conservative while $p_2^{(A)}$ and $p_3^{(A)}$ are liberal. For $2/3 < \rho \leq 1$, all asymptotic tests are robust (i.e. simulated significance level being between 0.04 and 0.06). For $\rho > 1$, $p_1^{(A)}$ and $p_4^{(A)}$ are liberal while $p_2^{(A)}$, $p_3^{(A)}$ and $p_5^{(A)}$ are robust.

The tests based on numerical approximations ($p_j^{(P)}, j = 1, \ldots, 5$), have simulated significance levels close to each other and close to the nominal level for all values of ρ. In other words, they are robust for all the settings considered here. As expected, the traditional conditional test (test based on $p^{(C)}$) is always conservative for all values of ρ. From these results, it is clear that the mid-p adjustment based on $p^{(M)}$ does overcome the conservativeness of the traditional conditional test (i.e., with the simulated significance

Table 1: Percentage of configurations (based on $\gamma_1 = 11(1)60$) whose simulated significance levels are inside the interval I when $\rho = 1/4, 1/3, 1/2, 2/3, 3/4$

ρ	I	$p_1^{(A)}$	$p_2^{(A)}$	$p_3^{(A)}$	$p_4^{(A)}$	$p_5^{(A)}$	$p_1^{(P)}$	$p_2^{(P)}$	$p_3^{(P)}$	$p_4^{(P)}$	$p_5^{(P)}$	$p^{(C)}$	$p^{(M)}$	$p^{(L)}$
1/4	≤ 0.04	100	0	0	100	0	0	0	0	0	0	94	0	0
	≥ 0.06	0	12	4	0	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	0	2	6	0	82	80	82	80	80	82	6	86	84
	$(0.05, 0.06)$	0	86	90	0	18	20	18	20	20	18	0	14	16
	$(0.0478, 0.0522)$	0	4	16	0	42	40	40	40	40	42	0	50	40
1/3	≤ 0.04	98	0	0	94	0	0	0	0	0	0	92	0	0
	≥ 0.06	0	2	2	0	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	2	2	14	6	48	52	66	66	50	58	8	80	76
	$(0.05, 0.06)$	0	96	84	0	52	48	34	34	50	42	0	20	24
	$(0.0478, 0.0522)$	0	18	24	0	62	56	56	58	56	50	0	50	44
1/2	≤ 0.04	60	0	0	4	0	0	0	0	0	0	80	0	0
	≥ 0.06	0	0	0	0	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	40	20	42	96	60	66	68	68	68	64	20	60	56
	$(0.05, 0.06)$	0	80	58	0	40	34	32	32	32	36	0	40	44
	$(0.0478, 0.0522)$	0	42	62	2	66	56	60	58	58	64	0	64	58
2/3	≤ 0.04	2	0	0	0	0	0	0	0	0	0	68	0	0
	≥ 0.06	0	0	0	0	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	98	28	64	84	74	70	70	70	70	74	32	78	68
	$(0.05, 0.06)$	0	72	36	16	26	30	30	30	30	26	0	22	32
	$(0.0478, 0.0522)$	10	60	46	28	50	44	44	44	44	50	0	56	52
3/4	≤ 0.04	0	0	0	0	0	0	0	0	0	0	74	0	0
	≥ 0.06	0	0	0	0	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	94	26	44	60	44	56	54	56	54	50	26	62	44
	$(0.05, 0.06)$	6	74	56	40	56	44	46	44	46	50	0	38	56
	$(0.0478, 0.0522)$	32	62	60	54	62	58	56	58	54	64	0	60	62

levels being much closer to the pre-specified nominal level). The LRT is robust for all values of ρ but the simulated significance levels are higher than the pre-specified nominal level a majority of the time when $\rho \geq 1.0$ (see Table 2).

Table 2: Percentage of configurations (based on $\gamma_1 = 11(1)60$) whose simulated significance levels are inside the interval I when $\rho = 1.0, 1.5, 2.0, 3.0$

ρ	I	$p_1^{(A)}$	$p_2^{(A)}$	$p_3^{(A)}$	$p_4^{(A)}$	$p_5^{(A)}$	$p_1^{(P)}$	$p_2^{(P)}$	$p_3^{(P)}$	$p_4^{(P)}$	$p_5^{(P)}$	$p^{(C)}$	$p^{(M)}$	$p^{(L)}$
1.0	≤ 0.04	0	0	0	0	0	0	0	0	0	0	66	0	0
	≥ 0.06	0	0	0	4	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	52	52	72	20	50	54	54	54	54	52	34	64	38
	$(0.05, 0.06)$	48	48	28	76	50	46	46	46	46	48	0	36	62
	$(0.0478, 0.0522)$	56	56	50	40	58	52	52	52	52	58	0	62	72
1.5	≤ 0.04	0	0	0	0	0	0	0	0	0	0	56	0	0
	≥ 0.06	12	0	0	24	0	0	0	0	0	0	0	0	4
	$(0.04, 0.05)$	4	70	84	2	56	58	58	58	56	58	44	66	18
	$(0.05, 0.06)$	84	30	16	74	44	42	42	42	44	42	0	34	78
	$(0.0478, 0.0522)$	16	48	46	8	62	52	52	52	58	50	0	58	44
2.0	≤ 0.04	0	0	2	0	0	0	0	0	0	0	54	0	0
	≥ 0.06	30	0	0	50	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	0	80	88	0	40	52	48	48	54	48	46	62	30
	$(0.05, 0.06)$	70	20	10	50	60	48	52	52	46	52	0	38	70
	$(0.0478, 0.0522)$	4	48	38	2	62	60	60	60	60	60	0	72	64
3.0	≤ 0.04	0	2	8	0	0	2	2	2	2	2	60	0	0
	≥ 0.06	74	0	0	86	0	0	0	0	0	0	0	0	0
	$(0.04, 0.05)$	2	92	92	2	28	48	44	46	52	44	40	74	18
	$(0.05, 0.06)$	24	6	0	12	72	50	54	52	46	54	0	26	82
	$(0.0478, 0.0522)$	0	36	22	0	56	70	66	66	62	64	0	70	40

5.2 Power

In terms of power of the tests based on numerical approximations of p-values ($p_j^{(P)}$, $j = 1, \ldots, 5$), the tests based on $p_3^{(P)}$ and $p_5^{(P)}$ outperform $p_1^{(P)}$, $p_2^{(P)}$, $p_4^{(P)}$ most of the time.

One should view a rejection of the null hypothesis for a liberal test with caution. In this regard, conservative tests are of less concern, for the type-I error rate is controlled below the nominal level. Therefore, we consider only the robust and conservative test procedures in the power comparison. We present the

power curves for test procedures based on $p_5^{(A)}$, $p_3^{(P)}$, $p_5^{(P)}$, $p^{(M)}$, and $p^{(L)}$ in Figure 2 for values of ρ = $1/4$, $1/2$, 1.5 and 2.0.

From the figures, we can see that the LRT is the most powerful among all the procedures. Although the LRT is robust in most situations and it is the most powerful for all the values of ρ, it has a higher percentage of simulated significance levels lying in (0.05, 0.06). We advise the use of LRT with caution. Among the test procedures based on $p_5^{(A)}$, $p_3^{(P)}$ and $p_5^{(P)}$, the latter two appear to generally have type-I error rates closer to the desired level compared to $p_5^{(A)}$. If one always wants the significance level of a test well controlled below the nominal level, then $p_3^{(P)}$ and $p_5^{(P)}$ are the desirable candidates. Otherwise, $p_5^{(A)}$ provides a reliable alternative with the significance levels being greater than the nominal level occasionally. It is noteworthy that $p_5^{(A)}$ can be carried out easily using a pocket calculator while $p_3^{(P)}$ and $p_5^{(P)}$ require the evaluation of an infinite sum of probabilities, requiring a computer program. As the test procedure based on $p_5^{(A)}$ has better power performance for $\rho > 1$, we recommend the use of the asymptotic tests $p_5^{(A)}$ for $\rho > 1$ and the tests based on $p_3^{(P)}$ and $p_5^{(P)}$ for $\rho \leq 1$ in practice.

5.3 Accuracy of sample size calculation formulae

In this subsection, we examine the accuracy of the sample size formulae in (6) - (12). For fifteen combinations of $\rho = (1/4, 1/2, 1, 1.5, 2)$ and $c = (0.25, 0.50, 0.75)$ the required sample sizes (sample time frames λ_1) computed from formulae (6) - (10) with $\alpha = 0.05$, $\beta = 0.10$ for the asymptotic test procedures are presented in Table 3 and (11) and (12) in Table 4. The simulated significance levels and the power values based on these sample sizes are also presented. One should be concerned about the cases for which the simulated significance level is significantly higher than the pre-specified level ($\alpha = 0.05$) or the simulated power is significantly lower than the pre-specified value ($1 - \beta = 0.90$).

From Table 3, our only concerns are that the simulated significance levels of tests based on $p_1^{(A)}$ and $p_4^{(A)}$ are higher than the pre-specified $\alpha = 0.05$ when $\rho = 2.0$ and $c = 0.25$, while the simulated levels based on $p_2^{(A)}$ and $p_3^{(A)}$ are lower than that level when $\rho = 2.0$, $c = 0.25$. These observations agree with

Table 3: Sample size for asymptotic test procedures using formulae $(6) - (10)$

		$p_1^{(A)}$			$p_2^{(A)}$			$p_3^{(A)}$			$p_4^{(A)}$			$p_5^{(A)}$		
ρ	c	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$
0.25	0.25	16	.0213	.9512	11	.0621	.9178	9	.0547	.8576	12	.0221	.9159	13	.0466	.9389
0.5	0.25	9	.0297	.9391	7	.0543	.9207	7	.0487	.9096	7	.0376	.8889	9	.0505	.9586
1.0	0.25	5	.0485	.9197	5	.0485	.9197	6	.0373	.9463	4	.0581	.9146	6	.0463	.9710
1.5	0.25	3	.0602	.9021	4	.0422	.9180	5	.0301	.9632	4	.0802	.9252	6	.0452	.9762
2.0	0.25	3	.0886	.8906	4	.0373	.9150	5	.0245	.9718	3	.0988	.9313	5	.0467	.9822
0.25	0.5	77	.0350	.9264	62	.0529	.9040	53	.0523	.8673	64	.0347	.8837	68	.0479	.9206
0.5	0.5	43	.0393	.9182	38	.0501	.9055	35	.0486	.8855	38	.0425	.8922	42	.0466	.9260
1.0	0.5	26	.0457	.9054	26	.0457	.9054	27	.0460	.9115	25	.0501	.9021	29	.0462	.9374
1.5	0.5	20	.0490	.8929	21	.0439	.9023	24	.0437	.9253	21	.0533	.9070	26	.0457	.9418
2.0	0.5	17	.0572	.8878	19	.0423	.9006	22	.0416	.9311	19	.0576	.9102	24	.0446	.9477
0.25	0.75	486	.0449	.9079	442	.0479	.8988	412	.0503	.8795	451	.0520	.8880	461	.0507	.9097
0.5	0.75	281	.0485	.9096	267	.0501	.9068	258	.0489	.8929	269	.0420	.8951	282	.0515	.9122
1.0	0.75	179	.0485	.9016	179	.0494	.9001	180	.0509	.8992	178	.0481	.9014	192	.0439	.9157
1.5	0.75	145	.0511	.8924	150	.0480	.8978	155	.0493	.9098	148	.0528	.9004	162	.0485	.9192
2.0	0.75	128	.0549	.8918	135	.0472	.8981	142	.0491	.9090	133	.0555	.9002	147	.0519	.9247

the findings in Section 5.1. When $\rho = 1.0$, the simulated significance levels of all asymptotic tests are close to the nominal level. From Table 4, we can see that the simulated significance values are much lower than the desired nominal level when $c = 0.25$ for $p^{(C)}$ and $p^{(M)}$ and somewhat low for $p^{(C)}$ when $c = 0.5$. This may be due to the small sample sizes at $c = 0.25$. Otherwise, most of the sample size formulae provide satisfactory results. The concerns about power < 0.90 are confined to $p_1^{(A)}$, $p_3^{(A)}$, and $p_4^{(A)}$.

6 Illustrative Examples

In this section, we revisit the two examples described in Section 1 to illustrate the test procedures considered in this paper.

Table 4: Sample size for test procedures based on $p_5^{(P)}$, $p^{(C)}$, $p^{(M)}$ and $p^{(L)}$

		$p_5^{(P)}$			$p^{(C)}$			$p^{(M)}$			$p^{(L)}$		
ρ	c	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$	λ_1	$\hat{\alpha}$	1 -$\hat{\beta}$
0.25	0.25	11	.0469	.9040	12	.0269	.8970	11	.0448	.9044	11	.0468	.9040
0.5	0.25	7	.0477	.9142	8	.0289	.9250	7	.0450	.9105	6	.0592	.8828
1.0	0.25	5	.0403	.9309	5	.0226	.9096	4	.0381	.8745	4	.0580	.8767
1.5	0.25	4	.0452	.9184	4	.0236	.8896	3	.0386	.8364	4	.0569	.9304
2.0	0.25	3	.0462	.8579	4	.0194	.9224	3	.0419	.8673	3	.0547	.8797
0.25	0.5	62	.0476	.8974	65	.0385	.8932	62	.0476	.8974	76	.0506	.8984
0.5	0.5	38	.0479	.9015	40	.0369	.8963	37	.0492	.8956	46	.0503	.8981
1.0	0.5	25	.0452	.8990	27	.0400	.8991	25	.0452	.8972	31	.0513	.9021
1.5	0.5	21	.0459	.8974	22	.0351	.8977	21	.0500	.8995	21	.0477	.9020
2.0	0.5	19	.0436	.8994	20	.0347	.8949	19	.0436	.8977	23	.0490	.8977
0.25	0.75	445	.0546	.8923	460	.0458	.9039	448	.0519	.9017	450	.0478	.9041
0.5	0.75	267	.0516	.8977	277	.0470	.9025	268	.0552	.9012	269	.0465	.9049
1.0	0.75	178	.0487	.8937	185	.0431	.9040	179	.0476	.8957	179	.0507	.9033
1.5	0.75	149	.0498	.9004	154	.0471	.8995	150	.0476	.9027	149	.0550	.8970
2.0	0.75	135	.0474	.9035	139	.0461	.9021	135	.0522	.8991	134	.0482	.9010

6.1 Example 1: Coronary Heart Disease

Referring to the first example in Section 1, $x_0 = 60$ CHD cases with $t_0 = 51477.5$ for the group not using postmenopausal hormone and $x_1 = 30$ CHD cases with $t_1 = 54308.7$ for the group using postmenopausal hormone. This implies $d = 1.055$, $R = 1$ and $\rho = 0.94787$. We obtain that $W_1 = 3.3849$, $W_2 = 3.4174$, $W_3 = 3.3393$, $W_4 = 3.5406$, and $W_5 = 3.4455$ and the corresponding p-values are presented in Table 5. There is strong evidence to support the conclusion that the incidence rate of CHD in the non-hormone-use group is higher than that in post-menopausal hormone-use group.

Next suppose that an epidemiologist wants to plan another study of the research question that the incidence rate of CHD for those using the hormone is no different for those not using the hormone. The epidemiologist would like to know the required sample size for the two groups when the ratio of sampling frames is 2 ($d = 2$). Suppose also that the observation time domain is 2 years, from the previous study

Table 5: Summary of the results for Example 1: Testing $H_0 : \gamma_0/\gamma_1 = 1$ vs. $H_1 : \gamma_0/\gamma_1 > 1$.

Test Procedure	p-value	Test Procedure	p-value
$p_1^{(A)}$	0.000356	$p_1^{(P)}$	0.000298
$p_2^{(A)}$	0.000316	$p_2^{(P)}$	0.000298
$p_3^{(A)}$	0.000420	$p_3^{(P)}$	0.000307
$p_4^{(A)}$	0.000200	$p_4^{(P)}$	0.000306
$p_5^{(A)}$	0.000285	$p_5^{(P)}$	0.000298
$p^{(C)}$	0.000310	$p^{(L)}$	0.000286
$p^{(M)}$	0.000428		

that the incidence rate of CHD for those using the hormone is 0.0005 ($\gamma_1 = 0.0005$), and $R^{'}$ is 4. From formulae (6) - (12), the values of $\lambda_1 = t_1\gamma_1$ and the required sample sizes for the hormone using group to achieve 90% power at $\alpha = 0.05$ for different test procedures are presented in Table 6. The smallest are for $p_3^{(A)}$ and $p_4^{(A)}$.

6.2 Example 2: Breast Cancer

There were $x_0 = 41$ cases of breast cancer with $t_0 = 28010$ in the treatment group of women receiving x-ray fluoroscopy and $x_1 = 15$ cases of breast cancer with $t_1 = 19017$ in the control group of women not receiving x-ray fluoroscopy. Hence, $d = 0.679$ and one might be interested in testing the hypotheses that $H_0 : \gamma_0/\gamma_1 = 1.5$ against $H_1 : \gamma_0/\gamma_1 > 1.5$. We obtain the test statistics $W_1 = 0.7358$, $W_2 = 0.7069$, $W_3 = 0.7056$, $W_4 = 0.7380$, and $W_5 = 0.6747$ and their p-values based on different procedures are presented in Table 7. There is not enough evidence that the incidence rate of breast cancer in the X-ray fluoroscopy group is 1.5 times to the incidence rate of breast cancer in control group.

Table 6: The values of $\lambda_1 = t_1\gamma_1$ and the required sample sizes for the example with $1-\beta = 0.9$, $\alpha = 0.05$, $R = 1$, $R' = 4$, $\gamma_1 = 0.0005$.

Test Procedure	λ_1	Required sample size
$p_1^{(A)}$	8.53	8527
$p_2^{(A)}$	6.86	6860
$p_3^{(A)}$	6.66	6655
$p_4^{(A)}$	6.66	6655
$p_5^{(A)}$	8.63	8627
$p_j^{(P)}$, $j = 1, \ldots, 5$	6.59	6590
$p^{(C)}$	7.26	7260
$p^{(M)}$	6.58	6580
$p^{(L)}$	6.37	6370

Table 7: Summary of the results for Example 2: Testing $H_0 : \gamma_0/\gamma_1 = 1.5$ vs. $H_1 : \gamma_0/\gamma_1 > 1.5$.

Test Procedure	p-value	Test Procedure	p-value
$p_1^{(A)}$	0.2309	$p_1^{(P)}$	0.2453
$p_2^{(A)}$	0.2398	$p_2^{(P)}$	0.2453
$p_3^{(A)}$	0.2402	$p_3^{(P)}$	0.2453
$p_4^{(A)}$	0.2303	$p_4^{(P)}$	0.2453
$p_5^{(A)}$	0.2499	$p_5^{(P)}$	0.2453
$p^{(C)}$	0.2913	$p^{(L)}$	0.2367
$p^{(M)}$	0.2450		

7 Conclusion

In this paper, we study four different approaches for testing the ratio of two Poisson rates and derive their sample size formulae. Based on our Monte Carlo simulation studies of the significance levels and powers,

we find that the asymptotic test derived from variance stabilizing transformation (W_5) is the most reliable asymptotic test (i.e., conservative but high power), and the test statistics derived from log-transformation with unconstrained MLE (W_3) and variance stabilizing transformation (W_5) are the best among all the five tests based on numerical approximations of exact p-values. The likelihood ratio test is the most powerful compared with other procedures; however, its simulated significance level can be liberal. The exact conditional tests are found to be conservative even with the mid-p correction for small values of γ.

Acknowledgments

This research was partially supported by a grant from the Research Grant Council of the Hong Kong Special Administration (Project no. HKBU261007). The authors are grateful to the editor, associate editor, and two referees for their valuable suggestions that greatly enhanced the manuscript.

Appendix A. Derivation of Constrained Maximum Likelihood Estimators

The log-likelihood function is

$$\ln L(\gamma_0, \gamma_1) = -\ln(X_0!X_1!) - t_0\gamma_0 - t_1\gamma_1 + X_0 \ln t_0 + X_0 \ln \gamma_0 + X_1 \ln t_1 + X_1 \ln \gamma_1. \quad \text{(A.1)}$$

Under the null hypothesis, $H_0 : \gamma_0/\gamma_1 = R$, (A.1) can be written

$$\ln L(\gamma_1) = -\ln(X_0!X_1!) - t_0(\gamma_1 R) - t_1\gamma_1 + X_0 \ln t_0 + X_0 \ln(\gamma_1 R) + X_1 \ln t_1 + X_1 \ln \gamma_1. \quad \text{(A.2)}$$

To obtain the maximum likelihood estimator of γ_1 (for any fixed R) we take the first derivative of (A.2) and set it to zero, which yields

$$\frac{d \ln L(\gamma_1)}{d\gamma_1} = -t_0 R - t_1 + \frac{X_0}{\gamma_1} + \frac{X_1}{\gamma_1} = 0$$

$$\Longrightarrow \quad \tilde{\gamma}_1 = \frac{X_0 + X_1}{Rt_0 + t_1} \text{ which implies } \tilde{\gamma}_0 = R\tilde{\gamma}_1 = \frac{X_0 + X_1}{t_0 + t_1/R}.$$

Substituting $d = t_1/t_0$ and $\rho = R/d$ yields

$$\tilde{\gamma}_0 = \frac{X_0 + X_1}{t_0(1 + 1/\rho)} \text{ and } \tilde{\gamma}_1 = \frac{X_0 + X_1}{t_1(1 + \rho)}.$$

Appendix B. Derivation of Sample Size Formulae

The critical region for W_1 at the α significance level consists of those points (X_0, X_1) that satisfy the inequality $X_0 - \rho X_1 \geq z_{1-\alpha}\sqrt{X_0 + X_1\rho^2}$. Under $H_1 : \gamma_0/\gamma_1 = R' > R$, $(X_0 - \rho X_1)$ is asymptotically normal with mean $(\rho/c - \rho)t_1\gamma_1$ and variance $(\rho/c + \rho^2)t_1\gamma_1$, where $c = R/R'$ and $\sqrt{X_0 + X_1\rho^2}$ converges in probability to $\sqrt{(\rho/c + \rho^2)t_1\gamma_1}$. Hence, the approximate power, P, can be expressed in terms of the cumulative normal distribution as

$$P = \Phi\left[\frac{z_{1-\alpha}\sqrt{(\rho/c + \rho^2)t_1\gamma_1} - (\rho/c - \rho)t_1\gamma_1}{\sqrt{(\rho/c + \rho^2)t_1\gamma_1}}\right], \tag{B.1}$$

where $\Phi(.)$ is the standard normal distribution function. Setting $P = 1 - \beta$ and solving (B.1), we can show that the sample size formula for the test procedure based on $p_1^{(A)}$ is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho + c^2)(z_{1-\alpha} + z_{1-\beta})^2}{(1-c)^2}.$$

Similarly, $\sqrt{(X_0 + X_1)\rho}$ has limit $\sqrt{t_1\gamma_1(\rho^2/c + \rho)}$ and the sample size formula for the test procedure based on $p_2^{(A)}$ is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho + c^2)\left[z_{1-\alpha}\sqrt{(c+\rho)/(1+c\rho)} + z_{1-\beta}\right]^2}{(1-c)^2}.$$

For the test procedure based on $p_3^{(A)}$ at the α significance level consists of those points (X_0, X_1) that satisfy the inequality $\ln(X_0/X_1) - \ln(\rho) \geq z_{1-\alpha}\sqrt{1/X_0 + 1/X_1}$. Under $H_1 : \rho_0/\rho_1 = R' > R$, $\ln(X_0/X_1) - \ln(\rho)$ is asymptotically normal with mean $\ln(1/c)$ and variance $(c/\rho + 1)/(t_1\gamma_1)$. Here $\sqrt{1/X_0 + 1/X_1}$ converges in probability to $\sqrt{(c/\rho + 1)/(t_1\gamma_1)}$. Hence the approximate power may be expressed in terms of the cumulative normal as

$$P = \Phi\left[\frac{z_{1-\alpha}\sqrt{(c/\rho + 1)/(t_1\gamma_1)} - \ln(1/c)}{\sqrt{(c/\rho + 1)/(t_1\gamma_1)}}\right]. \tag{B.2}$$

Setting $P = 1 - \beta$ and solving (B.2), the sample size formula for the test procedure $p_3^{(A)}$ is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho + 1)(z_{1-\alpha} + z_{1-\beta})^2}{(\ln c)^2}.$$

Following the same procedure, $\sqrt{(2+1/\rho+\rho)/(X_0+X_1)}$ has limit $\sqrt{(2+1/\rho+\rho)/((\rho/c+1)(t_1\gamma_1))}$ and the sample size for the test procedure $p_4^{(A)}$ is

$$\lambda_1 = t_1\gamma_1 = \frac{(c/\rho+1)\left\{z_{1-\alpha}\left[\frac{\sqrt{c(\rho^2+2\rho+1)}}{c+\rho}\right]+z_{1-\beta}\right\}^2}{(\ln c)^2}.$$

For the test procedure based on $p_5^{(A)}$, the critical region of the test is

$$\frac{2\left(\sqrt{X_0+3/8}-\sqrt{\rho(X_1+3/8)}\right)}{\sqrt{1+\rho}} \geq z_{1-\alpha}.$$

After some simple algebra, we can show that

$$\frac{2\left(\sqrt{X_0+3/8}-\sqrt{\rho/c(X_1+3/8)}\right)}{\sqrt{1+\rho/c}} \geq \frac{z_{1-\alpha}\sqrt{1+\rho}-2(\sqrt{\rho/c}-\sqrt{\rho})\sqrt{X_1+3/8}}{\sqrt{1+\rho/c}}.$$

Under the alternative hypothesis, the left-hand side of the equation is asymptotically normal distributed. By setting the type-II error rate to β, we can get the sample size for the test procedure based on $p_5^{(A)}$ as

$$\lambda_1 = \gamma_1 t_1 = \left[\frac{z_{1-\alpha}\sqrt{c/\rho+c}+z_{1-\beta}\sqrt{1+c/\rho}}{2\left(1-\sqrt{c}\right)}\right]^2 - \frac{3}{8}.$$

References

Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, **35**, 246-254.

Birnbaum, A. (1953). Some procedures for comparing Poisson processes or populations. *Biometrika*, **40**, 447-449.

Brownlee, K. A. (1967). *Statistical Theory and Methodology in Science and Engineering.* Wiley, New York.

Chapman, D. G. (1952). On tests and estimates for the ratio of Poisson means. *Annals of the Institute of Statistical Mathematics*, **4**, 45-49.

Cox, D. R. and Lewis, P. A. W. (1966). *The statistical analysis of series of events.* Methuen, London

Detre, K. and White, C. (1970). The comparison for two Poisson-distributed observations. *Biometrics*, **26**, 851-854.

Dykstra, R. L. and Robertson, T. (1982). Order restricted statistical tests on multinomial and Poisson parameters: The starshaped restriction. *The Annals of Statistics*, **10**, 1246-1252.

Gail, M. (1974). Power computations for designing comparative Poisson trials. *Biometrics*, **30**, 231-237.

Graham, P. L., Mengersen, K. and Morton, A. P. (2003). Confidence limits for the ratio of two rates based on likelihood scores: non-iterative method. *Statistics in Medicine*, **22**, 2071-2083.

Haight, F. A. (1967). *Handbook of the Poisson Distribution.* Wiley, New York.

Hald, F. A.(1960). *Statistical Theory with Engineering Applications.* Wiley, New York.

Huffman, M. D. (1984). An Improved Approximate Two-sample Poisson Test. *Applied Statistics*, **33**, 224-226.

Krishnamoorthy, K. and Thomson, J. (2004). A more powerful test for comparing two Poisson means. *Journal of Statistical Planning and Inference* , **119**, 23-35.

Lancaster, H. O. (1952). Statistical Control of Counting Experiments. *Biometrika*, **39**, 419-422.

Lancaster, H. O. (1961). Significance Tests in Discrete Distributions. *Journal of American Statistical Association*, **56**, 223-234.

Ng, H. K. T. and Tang, M. L. (2005). Testing the equality of two Poisson means using the rate ratio. *Statistics in Medicine*, **24**, 955-965.

Poisson, S. D. (1837). *Recherches sur la Probabilité des Jugements en Matiére Criminelle et en Mati´ere Civile.* Paris: Bachelier.

Przyborowski, J. and Wilenski, H. (1939). Homogeneity of results in testing samples from Poisson series. *Biometrika*, **31**, 313-323

Ractliffe, J. F. (1964). The significance of the difference between two Poisson variables: an experimental investigation. *Applied Statistics*, **13**, 84-86.

Rothman, K. J. and Greenland, S. (1998). *Modern Epidemiology.* 2nd edition. Lippincott-Raven, Philadelphia.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Inference.* Wiley: New York.

Shiue, W. K. and Bain, L. J. (1982). Experiment size and power comparisons for two-sample Poisson tests. *Applied Statistics*, **31**, 130-134.

Sichel, H. S. (1973). On a significance test for two Poisson variables. *Applied Statistics*, **22**, 50-58.

Stampfer, M. J. and Willett W. C. (1985). A prospective study of postmenopausal estrogen therapy and coronary heart disease. *New England Journal of Medicine*, **313**, 1044-1049.

Tang, M. L. (1998). Small-sample study of the use of mid-p power divergence goodness-of-fit tests. *Journal of Statistical Computation and Simulation* , **62**, 137-180.

Thode, H. C. (1997). Power and sample size requirements for tests of differences between two Poisson rates. *The Statistician*, **46**, 227-230.