Invited Paper

Sample Sizes to Compare Two Poisson Rates

Edsel A. Peña
Department of Statistics
University of South Carolina[1]

In this note, procedures for determining the sample sizes needed to compare the rates of two Poisson populations to achieve a pre-specified power at a given ratio of the rates are proposed. The first method relies on a conditional uniformly most powerful test (CUMPT) which leads to sample sizes that will guarantee the desired power, but at the cost of using more units than necessary. The second method relies on a normal approximation and may not always guarantee that the desired power will be achieved, but generally yields a power close to the pre-specified value and prescribes smaller sample sizes than the CUMPT-based method. Properties of the procedures are examined using simulation studies. The particular applicability and motivating situations leading to these procedures are in colon cancer research. Illustrations of the applicability of the procedures in studies dealing with tumor counts in mice are presented.

Keywords and phrases: Conditional uniformly most powerful test, normal approximation test, power function, test function

1. Introduction

In colon cancer studies a response variable of interest is the number of tumors that develop in experimental mice. This is a discrete random variable and the typical distributional model for this variable, denoted below by X, is the Poisson probability function (POI(λ)) given by

1 E. Peña [e-mail: pena@stat.sc.edu] is Professor in the Department of Statistics and Director of the Biometry Core of the Center for Colon Cancer Research (CCCR) at the University of South Carolina. He acknowledges research support provided by US NIH COBRE Grant RR17698 and RØ1 CA154731 and US NSF Grant DMS 1106435. He thanks Prof. F. Berger and Dr. M. Peña for discussions about practical situations in colon cancer research which motivated this research. He also thanks Professor Erniel Barrios for inviting him to contribute to an issue of *The Philippine Statistician.*

$$p(x \mid \lambda) = \Pr\{X = x \mid \lambda\} = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0,1,2,\ldots, \tag{1}$$

where λ is the mean rate of occurrence of tumors over a specified period of time. Other distributional models may be postulated for this variable such as over-dispersed models. However, we focus on the Poisson model, which is appropriate when modeling rare events such as tumor occurrence, and since it also arises in many other settings such as in reliability, engineering, biology, agriculture, ecology, etc. See for example (Przyborowski and Wilenski, 1940; Birnhaum, 1954; Bickel and Doksum, 1977; Roussas, 1997, Krishnamoorthy and Thompson, 2004) and references therein for discussions of the appropriateness of the Poisson model in applied work. Focusing on the Poisson model also enables a clear description of the problem and a more vivid demonstration of the solution. The main idea behind the proposed approach is extendable to other distributional settings.

A common problem encountered in colon cancer research is the comparison of a control group and a treatment group in terms of the mean number of tumors that develop in mice for a specified period of time. For instance, one may be interested in testing whether the average number of tumors that develop over a span of six weeks for the control group is greater than the average number of tumors that develop over the same period of time for the treatment group. A significant test would imply that the treatment being administered is more effective than the control. A perennial problem however that the experimenter encounters before performing the study is determining the appropriate sample sizes for the two groups in order that an appropriate statistical test with a given significance level will achieve a pre-specified detection power at a pre-determined ratio of occurrence rates that is judged by the experimenter to be of practical importance. For example, the experimenter may wish to determine the sample sizes so a 5% level test will detect a ratio in occurrence rates of at least 1.5 with power of at least 80%. An existing approach to dealing with this problem is to simply use sample size determination methods derived under the normal distribution which are typically available in existing statistical packages. However, one problem with such an approach is that the resulting testing procedure may not achieve the desired level of significance and the pre-specified power requirements. This sample size problem also arises in other settings. For instance, see Lemeshow et al. (1981) for the problem of comparing binomial proportions with very small success probabilities, and Heilbrun and McGee (1985) for problems dealing with other distributions.

In this note we provide two possible solutions to this sample size determination problem. In section 2 we describe the technical problem and provide preliminaries. Section 3 provides a solution based on conditionally uniformly most powerful tests. Section 4 provides a solution that relies on a normal approximation. The

properties of these procedures will be examined in Section 5. An illustration on how the procedures could be used in applied work is provided in Section 6. Section 7 will provide some concluding thoughts.

2. Problem and Preliminaries

Let $X_m = (X_1, X_2, \ldots, X_m)$ be independent and identically distributed (IID) random variables from a POI(θ) distribution and $\mathrm{Y}_n = (Y_1, Y_2, \ldots, Y_n)$ be IID random variables from a POI(λ) distribution, with X_m and Y_n being independent random vectors. The X_is represent the sample observations from the control group (say), whereas the Y_js represent the sample observations from the treatment group. Denote the ratio of the mean rates by $\rho = \theta/\lambda$. As pointed out by Birnbaum (1954) as well as by Lehmann (1997), this ratio is an important parametric function as it represents the relative rate of event occurrence of the first population relative to the second population. Of interest is to test the hypotheses

$$H_0: \rho \leq \rho_0\,,\ \lambda > 0 \quad \text{versus} \quad H_1: \rho > \rho_0, \lambda > 0, \tag{2}$$

where ρ_0 is some pre-specified ratio. The sample size problem is to determine, for a given λ, the sample sizes m and n satisfying $r = m/(m+n)$, with r pre-specified, in order for an α-level test $\delta(\mathrm{X}_m, \mathrm{Y}_n)$ of H_0 versus H_1 to have power at least equal to some pre-specified value $1 - \beta$ for every $\rho \geq \rho_1$, with $\rho_1 > \rho_0$ and pre-specified. Recall that a test $\delta(\mathrm{X}_m, \mathrm{Y}_n)$ is a statistic taking values in [0,1] with the interpretation that if $(\mathrm{X}_m, \mathrm{Y}_n) = (\mathrm{x}_m, \mathrm{y}_n)$, then $\delta(\mathrm{X}_m, \mathrm{Y}_n)$ is the probability of rejecting H_0 (see Lehmann, 1997). The specification of ρ_1 and β is decided by the experimenter based on certain scientific or practical considerations.

In the sequel, we denote by $I(A)$ the indicator function of an event A, so $I(A) = 1$ if and only if event A has occurred, otherwise it takes the value of zero. For X_m, denote by $S_m = \sum_{i=1}^{m} X_i$, and analogously for Y_n, denote by $T_n = \sum_{j=1}^{n} Y_j$. We recall the following basic results: (i) if X_m consists of IID POI(θ) random variables, then S_m is complete and sufficient for θ (see Bickel and Doksum, 1977; Roussas, 1997); (ii) if $V \sim POI(\theta)$ and $W \sim POI(\lambda)$ with V and W independent, then $Z = V + W \sim POI\ (\theta+\lambda)$ (see Bickel and Doksum, 1977; Roussas, 1997); and (iii) under the same conditions as in (ii), given $Z = z$, the distribution of V is binomial with parameters z and $\eta = \theta/(\theta+\lambda) = \rho/(\rho+1)$ (see Roussas, 1997), that is,

$$\Pr\{V = v \mid Z = z\} = \binom{z}{v} \eta^{v} (1-\eta)^{z-v} I\{v \in \{0,1,2,\ldots,z\}\}.$$

For the relevant situation in this paper where X_m IID POI(θ) and Y_n IID POI(λ), by virtue of the Sufficiency Principle (Bickel and Doksum, 1977), it suffices to consider procedures based only on the jointly sufficient statistic (S_m, T_n). Furthermore, using the preliminary results above,

$$Z_{mn} = S_m + T_n \sim \text{POI}(m\theta + n\lambda) \text{ and } S_m \mid Z_{mn} = k) \sim BIN(k, m = m(p,r)) \qquad (3)$$

where $\eta(\rho,r) = r\rho / [r\rho + (1-r)]$ and $r = m / (m+n)$.

3. CUMPT-Based Procedure

Consider first the situation where it is given that $Z_{mn} = S_m + T_n = k$ with $k \geq 0$. First, note that because $\eta(\rho, r)$ is increasing in ρ for fixed r, the set of hypotheses $H_0: \rho \leq \rho_0$ versus $H_1: \rho > \rho_0$ is equivalent to

$$H_0 : \eta \leq \eta_0 = \frac{r\rho_0}{r\rho_0 + (1-r)} \quad versus \quad H_1 : \eta > \eta_0.$$

By the Neyman-Pearson theory of testing hypothesis, and the fact that the family of binomial distributions possesses the monotone likelihood ratio (MLR) property (Lehmann, 1997), a conditionally uniformly most powerful test (CUMPT) α-level test for testing H_0 versus H_1 is the randomized test

$$\delta_k(S_m, T_n) = I\{S_m > C_k\} + \gamma_k I\{S_m = C_k\}. \qquad (4)$$

With B_{k,η_0} representing a binomial random variable with parameters (k, η_0) and Z denoting the set of integers, the constants C_k and γ_k in (4) are determined according to

$$C_k = \begin{cases} \min\{j \in Z : P(B_{k,\eta_0} > j) \leq \alpha\} & if \quad k > 0 \\ 0 & if \quad k = 0 \end{cases};$$

$$\gamma_k = \begin{cases} \{\alpha - P(B_{k,\eta_0} > C_k)\} / P(B_{k,\eta_0} = C_k) & if \quad k > 0 \\ \alpha & if \quad k = 0 \end{cases}.$$

These choices of C_k and γ_k guarantee that the test has size of α , i.e.,

$$E_{\eta_0}\{\delta_k(S_m, T_n) \mid S_m + T_n = k\} = \alpha$$

For relevant theory concerning the Neyman-Pearson theory, UMP test, and MLR property, we refer the reader to Lehmann (1997). See also Przyborowski and Wilenski (1940), Shiue and Bain (1982), Sahai and Misra (1992), and Krishnamoorthy and Thompson (2004) for related conditional and other tests pertaining to the comparison of Poisson rates.

The conditional power of the test in (4) is given by

$$\pi_k(\eta) = Pr\,(B_{k,\eta} > C_k) + \gamma_k Pr\,(B_{k,\eta} = C_k) \tag{5}$$

We now demonstrate in the following lemma monotonocity properties of the power function $\pi_k(\eta)$ in both k and η. These properties will be crucial in the development of the procedure for the sample size determination.

Lemma 1 *For fixed $\eta > \eta_0$, $\pi_k(\eta) \leq \pi_{k+1}(\eta)$ for every k = 0,1,2,..., and for fixed k = 0,1,2,..., $\pi_k(\eta)$ is nondecreasing in η.*

Proof: Let $V_1, V_2, ..., V_{k+1}$ be IID Bernoulli random variables with success probability of η. Consider the problem of testing $H_0: \eta \leq \eta_0$ versus $H_1: \eta > \eta_0$. Two possible exact α-level randomized tests for this hypothesis are

$$\delta_1^*(V_1, V_2, \ldots, V_{k+1}) = I\left\{\sum_{i=1}^{k+1} V_i > C_{k+1}\right\} + \gamma_{k+1} I\left\{\sum_{i=1}^{k+1} V_i = C_{k+1}\right\};$$

$$\delta_2^*(V_1, V_2, \ldots, V_{k+1}) = I\left\{\sum_{i=1}^{k} V_i > C_k\right\} + \gamma_k I\left\{\sum_{i=1}^{k} V_i = C_k\right\}.$$

The respective power functions of these two tests are $\pi_{k+1}(\eta)$ and $\pi_k(\eta)$. For each k, and because of the monotone likelihood ratio property of the binomial class of distributions, $\pi_k(\eta_1) \leq \pi_k(\eta_2)$ for every $\eta_1 < \eta_2$ (Lehmann, 1997), proving the second assertion in the lemma. Also, because test δ^*_1 is the UMP-α level test of H_0 versus H_1 based on $V_1, V_2, ..., V_{k+1}$, and because δ^*_2 is also an α -level test of the same hypotheses based on $V_1, V_2, ..., V_{k+1}$, then it follows that $\pi^*_{k+1}(\eta) \geq \pi^*_k(\eta)$ for every $\eta > \eta_0$.

Let $\beta \in (0,1)$ with $\alpha < 1 - \beta$. Choose $\beta_1, \beta_2 \in (0,1)$ such that $(1-\beta) = (1-\beta_1)(1-\beta_2)$. In particular, we may take $\beta_1 = \beta_2 = 1 - \sqrt{1-\beta}$. Also, for the ρ_1 with $\rho_1 > \rho_0$ for which we want the test's power to be at least $(1-\beta)$, denote by η_1 the value of η at this ρ -value and for the fixed value of r . Define

$$k^* \equiv k^*(\alpha, \rho_0, \beta_1, \rho_1, r) = \min\{k \in Z : \pi_k(\eta_1) \geq 1 - \beta_1\}. \quad (6)$$

Note that k^* will be finite because $\pi_k(\eta_1) \to 1$ as k$\to\infty$. For this value of k^*, and with the notation that P_v denotes a Poisson random variable with mean rate of v, define for a given λ and r the quantity

$$M^*_\lambda = \min\{m \in Z_+ : Pr\{P_{m\lambda(\rho_1+(1-r)/r)} \geq k^*\} \geq 1 - \beta_2\}. \quad (7)$$

Then, define $N^*_\lambda = M^*_\lambda(1-r)/r$. That M^*_λ (and also N^*_λ) is finite follows from the fact that if $V \sim POI(v)$, then for every $v_1 < v_2$, $P_{v1}(V \geq k) \leq P_{v2}(V \geq k)$ for all k = 0,1,2,.... A simple proof of this result follows from the gamma representation (see [9]) of the Poisson distribution function.

For this M^*_λ, N^*_λ, let $X_{M^*_\lambda} = (X_1, X_2, ..., X_{M^*_\lambda})$ IID POI(θ) and $Y_{M^*_\lambda} = (Y_1, Y_2, ..., Y_{N^*_\lambda})$ IID POI(λ). For these samples, define the unconditional randomized test

$$\delta^*(X_{M^*_\lambda}, Y_{N^*_\lambda}) = I\left\{S_{M^*_\lambda} > C_{Z_{M^*_\lambda N^*_\lambda}}\right\} + \gamma_{Z_{M^*_\lambda N^*_\lambda}} I\left\{S_{M^*_\lambda} = C_{Z_{M^*_\lambda N^*_\lambda}}\right\}. \quad (8)$$

Denote the power function of this test by

$$\pi^*(\eta, \lambda) = E_{(\eta, \lambda)}\left\{\delta^*(X_{M^*_\lambda}, Y_{N^*_\lambda})\right\}. \quad (9)$$

Theorem 1 *For a fixed $\lambda > 0$, the test δ^* in* (8) *is of level α for testing $H_0: \eta \leq \eta_0$* versus *$H_1: \eta > \eta_0$ and for every $\eta \geq \eta_1$, $\pi^*(\eta, \lambda) \geq 1 - \beta$.*

Proof: First, note by the iterated expectation rule that

$$\pi^*(\eta, \lambda) = E_{(\eta, \lambda)}\{E_{(\eta, \lambda)}[\delta^*(X_{M^*_\lambda}, Y_{N^*_\lambda}) \mid Z_{M^*_\lambda N^*_\lambda}]\}. \quad (10)$$

Since, by construction of the test δ^*, we have for every k = 0,1,..., $M^*_\lambda + N^*_\lambda$,

$$E_{(\eta_0, \lambda)}\{\delta^*(X_{M^*_\lambda}, Y_{N^*_\lambda}) \mid Z_{M^*_\lambda N^*_\lambda} = k\} = \pi_k(\eta_0) = \alpha,$$

then it follows that δ^* is of level α, where we also used the monotonicity of the conditional power function in η as provided by Lemma 1. From (10), we also have that for $\eta > \eta_1$,

$$\pi^*(\eta,\lambda) = \sum_{k=0}^{\infty} \pi_k(\eta) P_{(\eta,\lambda)} \left\{ Z_{M_\lambda^* N_\lambda^*} = k \right\} \geq \sum_{k=k^*}^{\infty} \pi_k(\eta) P_{(\eta,\lambda)} \left\{ Z_{M_\lambda^* N_\lambda^*} = k \right\}$$

$$\geq \pi_{k^*}(\eta) P_{(\eta,\lambda)} \left\{ Z_{M_\lambda^* N_\lambda^*} \geq k^* \right\} \geq \pi_{k^*}(\eta_1) P_{(\eta_1,\lambda)} \left\{ Z_{M_\lambda^* N_\lambda^*} \geq k^* \right\}$$

$$\geq (1-\beta_1)(1-\beta_2) = 1-\beta,$$

where we used the monotonicity results in Lemma 1 to obtain the intermediate inequalities. Thus, the test δ^* has power at least $1-\beta$ whenever $\eta \geq \eta_1$, or equivalently, whenever $\rho \geq \rho_1$.

From these results, a solution of the sample size determination problem is contained in the following corollary. Observe that from the proof of Theorem 1, the procedure for determining the sample sizes is conservative in that it leads to possibly larger samples sizes than are necessary.

Corollary 1 *Given a λ , a sample size pair (m,n) with $r = m/(m+n)$ which will guarantee the existence of an α-level test for $H_0: \eta \leq \eta_0$ versus $H_1: \eta > \eta_0$ with power of at least $1-\beta$ for $\eta \geq \eta_1$ is given by $(m,n) = M_\lambda^*, N_\lambda^*$ where M_λ^* and N_λ^* are as defined above. One may take $\beta_1 = \beta_2 = 1-\sqrt{1-\beta}$, and the test is as provided by δ^* defined above.*

4. Normal Approximation-Based Method

Perhaps, the usual approach to this problem is to utilize a normal approximation. Given X_m IID POI(θ) and Y_n IID POI(λ), with $\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i$ and $\bar{Y}_n = \frac{1}{n}\sum_{j=1}^{n} Y_j$ being the respective sample means, a possible test statistic is

$$W_{m,n}(\rho_0) = \frac{(\bar{X}_m - \rho_0 \bar{Y}_n)}{\sqrt{\bar{X}_m / m + \rho_0^2 \bar{Y}_n / n}}. \tag{11}$$

Under $\rho=\rho_0$, and provided that m and n are large, $W_{m,n}(\rho_0)$ will be approximately distributed as standard normal. Consequently, an approximate α -level test for $H_0: \rho \leq \rho_0$, $\lambda > 0$ versus $H_1: \rho \leq \rho_0$, $\lambda > 0$ is

$$\delta^{**}(X_m, Y_n) = I\{W_{m,n}(\rho_0) \geq z_\alpha\} \tag{12}$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$ is the $100(1-\alpha)$th percentile of Φ (.), the standard normal distribution function. The approximate power function of this test at (ρ, λ) is

$$\pi^{**}(\rho,\lambda) \approx 1-\Phi\left(z_\alpha + \frac{\sqrt{m\lambda}(\rho_0-\rho)}{\sqrt{\rho+\rho_0^2 r/(1-r)}} \right). \tag{13}$$

If at $\rho = \rho_1$ we desire an approximate power of $1-\beta$, using this approximate power function, we must have

$$\frac{\sqrt{m\lambda}(\rho_0-\rho_1)}{\sqrt{\rho_1+\rho_0^2 r/(1-r)}} + z_\alpha = -z_\beta .$$

Solving for *m* yields the expression

$$M_\lambda^{**} = \frac{\left(z_\alpha + z_\beta\right)^2 \left(\rho_1 + \rho_0^2 \frac{r}{1-r} \right)}{\lambda(\rho_1-\rho_0)^2} \tag{14}$$

and we also obtain $N_\lambda^{**} = M_\lambda^{**}(1-r)/r$. We may therefore employ the procedure where, for a given λ, the sample sizes are determined by $(M_\lambda^{**}, N_\lambda^{**})$ and to test $H_0: \rho \le \rho_0$ versus $H_1: \rho > \rho_0$ we use

$$\delta^{**}(X_{M_\lambda^{**}}, Y_{N_\lambda^{**}}) = I\left\{ W_{M_\lambda^{**}, N_\lambda^{**}}(\rho_0) \ge z_\alpha \right\}. \tag{15}$$

Since this procedure is based on the large-sample normal approximation to the distribution of the statistic in (11), there is a need to examine whether the desired significance level and the desired power at $\rho = \rho_1$ are actually achieved, and to compare its performance with the CUMPT-based procedure in the preceding section. The examination of the level and power of this test and its comparison with the CUMPT-based procedure will be undertaken in the next section.

It should be mentioned in passing that one may develop and employ a sequential approach analogous to Stein's two-stage sampling scheme (Bickel and Doksum, 1977), or possibly a group sequential method. However, a sequential approach may not be feasible in practice because of the length of time needed to make a complete observation on an experimental unit. There are also other approaches that rely on asymptotic approximations such as those based on generalized linear models discussed by Self and Mauritsen (1988), Signorini (1991), and Self et al. (1992).

5. Simulated Properties of Procedures

5.1 Comparison of CUMPT and Normal-Based test

We first examined the level and power properties of the CUMPT and the normal approximation based test. An R object (Ihaka and Gentleman, 1996) named CondTest, which implements the CUMPT, as well as another R object, called NormApproxTest, which performs the normal-based test, are available upon request from the author. In performing the comparison, for fixed sample sizes m and n, and for specified values of ρ_0, λ, and α, at a given value of ρ we generated 2000 sets of samples X_m IID POI($\rho\lambda$) and Y_n IID POI(λ). For each set we performed the test of $H_0: \rho \leq \rho_0$ versus $H_1: \rho > \rho_0$ and determined the percentage out of the 2000 replications in which H_0 was rejected. This is the estimate of the power function at the given ρ value. This was done using the CUMPT and the normal-based test. Some of the results of these simulations are presented in Figure 1.

Aside from comparing the powers of these two test procedures, another goal of this simulation study is to determine the achieved level of the normal-based test as it is anchored on an approximation. Based on the results of the simulation runs, when λ is not small, the normal-based test achieves the desired level even for small sample sizes. An explanation for this excellent behavior is that when λ is not small, then the Poisson distribution is well approximated by a normal distribution. On the other hand, when λ is small, then the achieved level of the normal-based test maybe higher than the nominal level as for instance depicted by the third plot in Figure 1. In other runs where the λ is set to be quite small, the achieved level could be a lot smaller than the nominal level, thus making this normal-based test quite conservative. By virtue of its development, the CUMPT always achieves the desired nominal level.

5.2 Comparison of the sample size procedures

Next, we compared the power functions of the two-step procedure wherein the sample sizes are obtained according to either the CUMPT-based approach or the normal-based approach, and then determining through simulation the achieved power functions when each test is then applied to the resulting sample data. The simulation study is described as follows. We specify λ, ρ_0, α, ρ_1, and β, as well as $r = .05$. Given these values, we use the two procedures to determine the appropriate sample sizes. For instance, the CUMPT-based method will yield an (m_1,n_1), whereas the normal-based approach will yield (m_2,n_2). For each ρ-value, 2000 replications of the experiment described in subsection 5.1 were performed where for the CUMPT the sample sizes (m_1,n_1) are utilized, whereas for the normal-based test the sample sizes (m_2,n_2) are used. This is repeated for several ρ-values. The resulting estimates of the power functions are depicted in Figure 2.

Examining these figures, we note that both procedures achieve very closely the nominal level of significance of α= 5% for all the cases considered in the eight combinations of λ, ρ_0, and ρ_1. As is to be expected from the theoretical developments, the CUMPT-based procedure leads to a power at ρ_1 that is always larger than the desired pre-specified power, which means that it specifies a larger sample size than necessary. This is the price paid to guarantee that the desired power is achieved. On the other hand, the normal-based test, which dictates a smaller sample size also performs quite well, at least for the combinations of parameter values depicted in the figures. There is some slight degradation in the achieved powers at ρ_1 in that it need not guarantee that the achieved power is at least what is desired, such as in the first panel in Figure 2. Thus, in essence, if one truly desires a certain minimum power for a given ρ_1, the normal-based procedure is not guaranteed to achieve it, whereas the CUMPT-based procedure will definitely guarantee that the achieved power at that ρ_1 value will be at least what is desired, albeit at a larger sample size than necessary, hence at a higher cost.

5.3 Desired sample sizes

In this subsection we provide tables and curves for determining the appropriate sample sizes for two values of λ : 1 and 20. Tables 1 and 2 provide the needed sample sizes to achieve a power of $1-\beta$ at ρ when the test is of level α =.05. The first table uses the CUMPT-based procedure, while the second table is based on the normal approximation. Notice that the CUMPT-based sample sizes are larger than the normal-based procedure, which is the reason that the powers associated with the CUMPT approach in the preceding subsection are larger than those for the normal-based procedure. Tables 3 and 4 are the sample sizes needed for the case of λ = 20. Pictorial representations of Table 1 and Table 3 are provided by Figure 3. To obtain the sample sizes for other combinations of α, ρ_0, λ, β, ρ_1, and r, one may use the R object named SampSize which is also available upon request from the author.

6. Illustration of Procedures

In this section we demonstrate how to use the procedures in applied work, in particular, in work at the Center for Colon Cancer Research (CCCR) at the University of South Carolina.

A colon cancer researcher is studying the occurrence of intestinal tumors in APC $^{Min^{+/-}}$ mice and wishes to compare a control group and a treatment group. Mice are to be observed over a period of 6 weeks, and during this period a treatment protocol will be applied on the treatment group. After 6 weeks, the mice will be

sacrificed and the tumors counted. It is of interest to determine the number of mice that need to be included in the study to guarantee that in testing the null hypothesis that the rates of the control and the treatment groups are equal versus the research hypothesis that the treatment protocol has reduced the tumor occurrence rate by a factor of 1.5, that is, that $\rho > 1.5$, the power of the test is at least 80%. The test of hypothesis is to be conducted at a level of significance of 5%. Furthermore, it is known from prior experience that the tumor occurrence rate for the control group during 6 weeks is about 35. If the alternative hypothesis is true, then this means that the rate of occurrence for the treatment group will be about 20 tumors during the 6 weeks. To determine the required sample sizes, we set $r = .5$ so that $m = n$. We first apply the CUMPT-based method. By referring to Table 3 or Figure 3, taking $\beta = .2$ and $\rho = 1.5$, we find that the desired sample size is $m = 5$ mice per group. On the other hand, if one is to use the normal-based test, by referring to Table 4, we find that the appropriate sample size is $m = 3$. However, with this latter method, it is not guaranteed that the resulting power at $\rho = 1.5$ will be at least 80%, but it will be close to this desired value, whereas if one uses the sample size determined using the CUMPT-based method, then it is guaranteed that the power at $\rho = 1.5$ will be at least 80%.

Another researcher at the CCCR is studying the occurrence of tumors in the colon, and it is known that the rates of occurrence of colon tumors over the desired study period is quite low. Suppose that this researcher wishes to compare a treatment group and a control group, and using a 5% level hypothesis test for the equality of the tumor occurrence rates for the two groups, would like to be assured that the power of the test when $\rho = 2$ is at least 90%. From prior knowledge, the rate of tumor occurrence for the control group is about $\theta = 2$, hence at the desired ρ-value the approximate rate for the treated group, if the research hypothesis is correct, will be $\lambda = 1$. Referring to Table 1, we therefore find that the researcher should use about 37 experimental units per group using the CUMPT-based method. If one is to use the normal approximation method, then the desired sample size per group is 25 obtained from Table 2. However, if this latter sample size is used, then the researcher could not guarantee that the power at $\rho = 2$ will in fact be at least 90% .

7. Concluding Remarks

There are other issues that arise in considering this practical problem of determining the appropriate sample sizes in comparing two Poisson rates. First, the procedures proposed have the limitation that the researcher should have an idea of the value of λ, one of the parameters. As this is usually obtained from knowledge of the rate of event occurrence for the control group, then in many situations the researcher may have this knowledge. However, in other settings,

such knowledge may not be available. Without recourse to a sequential approach, as pointed out by Lehmann (1997), if we truly allow λ to be a nuisance parameter taking values in $(0, \infty)$, it is impossible to have a sample size formula that could guarantee a specified power larger than the desired significance level because any (good) test will have power that converges to the level of significance as $\lambda \downarrow 0$. This leads to the interesting question of whether there is a two-stage approach, similar to Stein's two-stage method (see Bickel and Doksum, 1977)) under the normal setting, which solves this sample size problem. A related proposal by Birnbaum (1954) and Lehmann (1997) is to keep sampling until the value of k^* is achieved, and then to perform the test. However, as pointed out earlier, such two-stage or sequential approaches may not be feasible in practical colon cancer studies because of the length of time it takes to observe an experimental unit such as a mouse.

Second, there is the question on whether the CUMPT-based approach proposed could be generalized to settings where the response variable is discrete, but may have a distribution other than the Poisson distribution. For example, how would the sample sizes be determined if the response distribution is over-dispersed compared to the Poisson distribution, or if the distribution is Bernoulli or negative binomial? Based on preliminary studies, it appears that the idea is indeed extendable to other more general settings and this will be pursued in forthcoming papers. Furthermore, would the proposed method be extendable to the case where there are several (say Poisson) populations to compare?

Finally, it was observed in the simulation studies that the CUMPT-based approach provides a conservative estimate of the desired sample sizes in that more experimental units, hence higher costs and resource requirements, are dictated by this procedure; whereas the normal approximation method provides a smaller sample size requirement, but does not guarantee that the pre-specified power requirement is achieved. Consequently, there is the question of whether one may be able to devise an intermediate approach combining these two approaches so the resulting sample sizes will be lower than that dictated by the CUMPT-based method, but guarantee the pre-specified power.

References

BICKEL, P. and K. DOKSUM, 1977, *Mathematical Statistics,* Holden-Day, Inc.

BIRNBAUM, A., 1954, Statistical methods for Poisson processes and exponential populations, *Journal of the American Statistical Association*, 49, 254-256.

HEILBRUN, L. and D. MCGEE, 1985, Sample size determination for the comparison of normal means when one sample size is fixed, *Computational Statistics and Data Analysis,* 3, 99-102.

IHAKA, R. and R. GENTLEMAN, 1996, R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, 5, 299-314.

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

LEHMANN, E., 1997, *Testing Statistical Hypothesis*, 2nd edition, Springer Verlag.

LEMESHOW, S., D. HOSMER, and J. STEWART, 1981, A comparison of sample size determination methods in the two group trial where the underlying disease is rare, *Communications in Statististics: Simulation and Computation*, 5, 437-449.

PRZYBOROWSKI, J. and H. WILENSKI, 1940, Homogeneity of results in testing samples from Poisson series, *Biometrika*, 31, 313-323.

ROUSSAS, G., 1997, *A Course in Mathematical Statistics*, Academic Press.

SAHAI, H. and S. MISRA, 1992, Comparing Means of Two Poisson Distribution, *Math Scientist*, 17, 60-67.

SELF, S. and R. MAURITSEN, 1988, Power/Sample Size Calculations for Generalized Linear Models, *Biometrics*, 44, 79-86.

SELF, S., R. MAURITZEN and J. OHARA, 1992, Power Calculations for Likelihood Ratio Tests in Generalized Linear Models, *Biometrics*, 48, 31-39.

SHIUE, W. and L. BAIN, 1982, Experiment size and power comparisons for two-sample Poisson tests, *Applied Statistics,* 31, 130-134.

SIGNORINI, D., 1991, "Sample Size for Poisson Regression," *Biometrika*, 78, 446-450.

Table 1: Required sample size (m) for sample 1 to achieve a power of $1-\beta$ at ρ when testing $H_0: \rho \leq \rho_0 = 1$ versus $H_1: \rho > 1$ for $\lambda = 1$ and when the test is to have level $\alpha = .05$ and when $n = m$. These sample sizes are based on the CUMPT procedure.

	ρ	β = 1 - Power							
1		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
2	1.10	2308.0	1796.0	1473.0	1226.0	1019.0	833.0	657.0	479.0
3	1.15	1061.0	825.0	676.0	562.0	466.0	381.0	299.0	218.0
4	1.20	617.0	479.0	392.0	325.0	270.0	220.0	173.0	125.0
5	1.25	407.0	315.0	258.0	214.0	178.0	144.0	114.0	82.0
6	1.30	292.0	226.0	185.0	153.0	126.0	103.0	81.0	58.0
7	1.35	221.0	170.0	140.0	115.0	95.0	78.0	60.0	44.0
8	1.40	174.0	135.0	110.0	91.0	75.0	61.0	47.0	34.0
9	1.45	141.0	109.0	89.0	74.0	61.0	50.0	39.0	28.0
10	1.50	117.0	90.0	74.0	61.0	50.0	41.0	32.0	23.0
11	1.55	100.0	76.0	63.0	52.0	42.0	34.0	27.0	20.0
12	1.60	86.0	66.0	54.0	44.0	37.0	29.0	23.0	17.0
13	1.65	75.0	58.0	47.0	39.0	32.0	26.0	20.0	15.0
14	1.70	66.0	51.0	41.0	34.0	28.0	23.0	18.0	13.0
15	1.75	59.0	45.0	37.0	30.0	25.0	20.0	16.0	11.0
16	1.80	53.0	41.0	33.0	27.0	22.0	18.0	14.0	10.0
17	1.85	48.0	37.0	30.0	25.0	20.0	16.0	13.0	9.0
18	1.90	44.0	33.0	27.0	23.0	18.0	15.0	12.0	8.0
19	1.95	40.0	31.0	25.0	21.0	17.0	13.0	11.0	8.0
20	2.00	37.0	28.0	23.0	19.0	15.0	13.0	10.0	7.0

Table 2: Required sample size (m) for sample 1 to achieve a power of $1-\beta$ at ρ when testing $H_0: \rho \leq \rho_0 = 1$ versus $H_1: \rho > 1$ for $\lambda = 1$ and when the test is to have level $\alpha = .05$ and when $n = m$. These sample sizes are based on the normal approximation procedure.

	r	β = 1 - Power							
1		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
2	1.10	1798.0	1298.0	988.0	756.0	568.0	406.0	263.0	135.0
3	1.15	818.0	590.0	449.0	344.0	258.0	185.0	119.0	61.0
4	1.20	471.0	340.0	258.0	198.0	148.0	106.0	69.0	35.0
5	1.25	308.0	222.0	169.0	129.0	97.0	69.0	45.0	23.0
6	1.30	218.0	157.0	120.0	92.0	69.0	49.0	32.0	16.0
7	1.35	164.0	118.0	90.0	69.0	51.0	37.0	24.0	12.0
8	1.40	128.0	92.0	70.0	54.0	40.0	29.0	18.0	9.0
9	1.45	103.0	74.0	56.0	43.0	32.0	23.0	15.0	7.0
10	1.50	85.0	61.0	47.0	36.0	27.0	19.0	12.0	6.0
11	1.55	72.0	52.0	39.0	30.0	22.0	16.0	10.0	5.0
12	1.60	61.0	44.0	33.0	26.0	19.0	13.0	9.0	4.0
13	1.65	53.0	38.0	29.0	22.0	16.0	12.0	7.0	4.0
14	1.70	47.0	34.0	25.0	19.0	14.0	10.0	6.0	3.0
15	1.75	41.0	30.0	23.0	17.0	13.0	9.0	6.0	3.0
16	1.80	37.0	27.0	20.0	15.0	11.0	8.0	5.0	2.0
17	1.85	33.0	24.0	18.0	14.0	10.0	7.0	4.0	2.0
18	1.90	30.0	22.0	16.0	12.0	9.0	6.0	4.0	2.0
19	1.95	27.0	20.0	15.0	11.0	8.0	6.0	4.0	2.0
20	2.00	25.0	18.0	14.0	10.0	8.0	5.0	3.0	1.0

Table 3: Required sample size (m) for sample 1 to achieve a power of $1-\beta$ at ρ when testing $H_0: \rho \leq \rho_0 = 1$ versus $H_1: \rho > 1$ for $\lambda = 20$ and when the test is to have level $\alpha = .05$ and when $n = m$. These sample sizes are based on the CUMPT procedure.

	r	β = 1 - Power							
1		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
2	1.10	115.0	90.0	74.0	61.0	51.0	42.0	33.0	24.0
3	1.15	53.0	41.0	34.0	28.0	23.0	19.0	15.0	11.0
4	1.20	31.0	24.0	20.0	16.0	13.0	11.0	9.0	6.0
5	1.25	20.0	16.0	13.0	11.0	9.0	7.0	6.0	4.0
6	1.30	15.0	11.0	9.0	8.0	6.0	5.0	4.0	3.0
7	1.35	11.0	9.0	7.0	6.0	5.0	4.0	3.0	2.0
8	1.40	9.0	7.0	5.0	5.0	4.0	3.0	2.0	2.0
9	1.45	7.0	5.0	4.0	4.0	3.0	2.0	2.0	1.0
10	1.50	6.0	5.0	4.0	3.0	3.0	2.0	2.0	1.0
11	1.55	5.0	4.0	3.0	2.0	2.0	2.0	1.0	1.0
12	1.60	4.0	3.0	3.0	2.0	2.0	1.0	1.0	1.0
13	1.65	4.0	3.0	2.0	2.0	2.0	1.0	1.0	1.0
14	1.70	3.0	3.0	2.0	2.0	1.0	1.0	1.0	1.0
15	1.75	3.0	2.0	2.0	2.0	1.0	1.0	1.0	1.0
16	1.80	3.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0
17	1.85	2.0	2.0	2.0	1.0	1.0	1.0	1.0	1.0
18	1.90	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
19	1.95	2.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
20	2.00	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Table 4: Required sample size (m) for sample 1 to achieve a power of $1-\beta$ at ρ when testing $H_0: \rho \leq \rho_0 = 1$ versus $H_1: \rho > 1$ for $\lambda = 20$ and when the test is to have level $\alpha = .05$ and when $n = m$. These sample sizes are based on the normal approximation.

	r	β = 1 - Power							
1		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
2	1.10	89.0	64.0	49.0	37.0	28.0	20.0	13.0	6.0
3	1.15	4.0	29.0	22.0	17.0	12.0	9.0	5.0	3.0
4	1.20	23.0	17.0	12.0	9.0	7.0	5.0	3.0	1.0
5	1.25	15.0	11.0	8.0	6.0	4.0	3.0	2.0	1.0
6	1.30	10.0	7.0	6.0	4.0	3.0	2.0	1.0	1.0
7	1.35	8.0	5.0	4.0	3.0	2.0	1.0	1.0	1.0
8	1.40	6.0	4.0	3.0	2.0	2.0	1.0	1.0	1.0
9	1.45	5.0	3.0	2.0	2.0	1.0	1.0	1.0	1.0
10	1.50	4.0	3.0	2.0	1.0	1.0	1.0	1.0	1.0
11	1.55	3.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
12	1.60	3.0	2.0	1.0	1.0	1.0	1.0	1.0	1.0
13	1.65	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
14	1.70	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
15	1.75	2.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
16	1.80	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
17	1.85	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
18	1.90	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
19	1.95	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
20	2.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Figure 1: Simulated power functions of the CUMPT and the normal approximation test procedures for fixed sample sizes. The plots are for $m = n = 2$, and for varied values of ρ_0 and λ. The level of significance is $\alpha = 5\%$. There were 2000 replications per ρ-value.

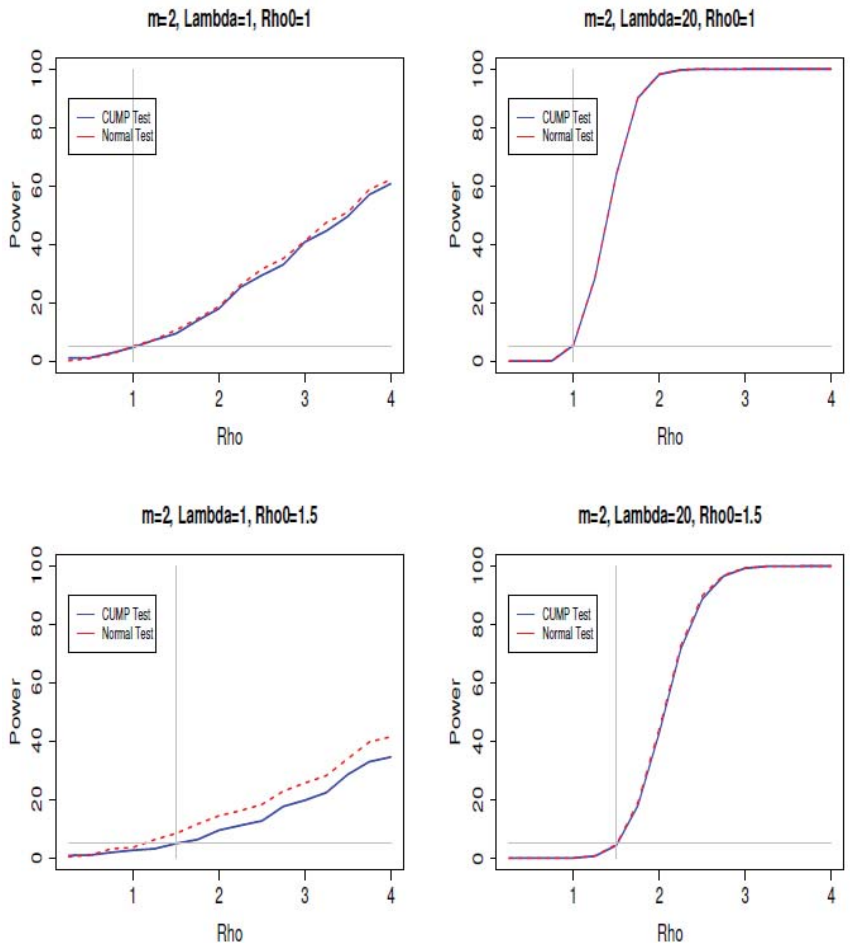

Figure 2: Achieved power functions of the procedures for determining sample size and then performing test based on the CUMPT and the normal approximation test for varied values of ρ_0, λ, and $\rho_1 = 2$. The pre-specified power at ρ_1 is $(1-\beta)100=80\%$, the level of significance is 5%, and $r = .5$. After the sample sizes have been determined, for each ρ value, 2000 replications were performed.

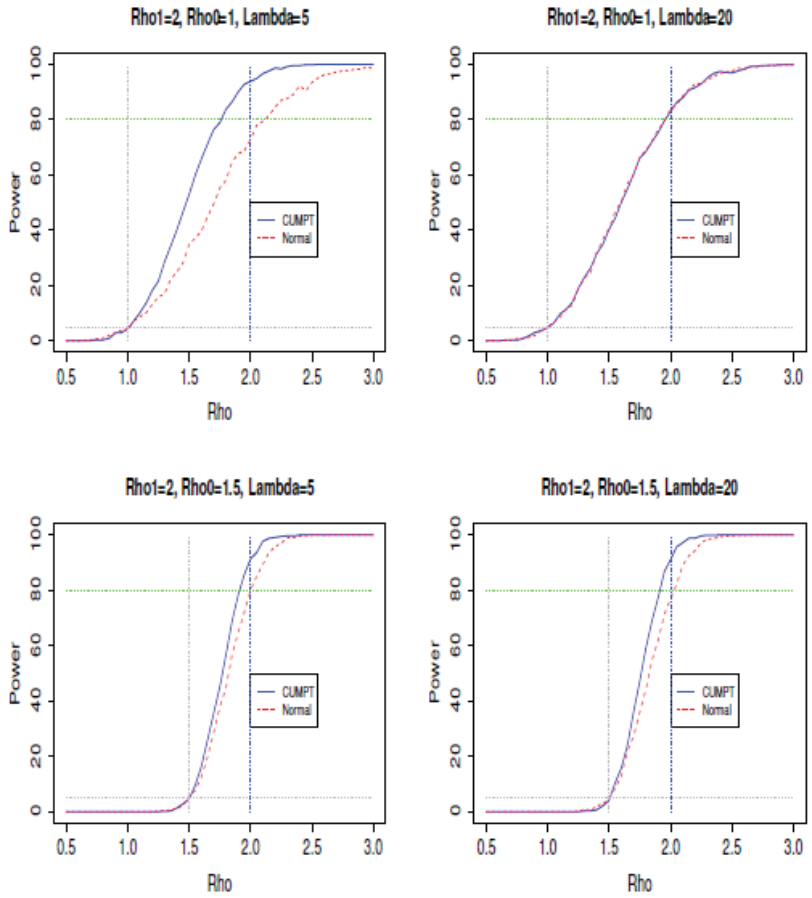

Figure 3: Curves for determining the size of sample 1 for $\lambda \in \{1,20\}$ in the case where $r = .5$ for varied values of β [power = 1–β] and ρ_1. These curves are for the case where $\rho_0 = 1$ and $\alpha = .05$.