

Tests for cured proportion for recurrent event count data – Uncensored case with covariates

Sumathi K¹ and Aruna Rao K²

Department of Mathematics, Manipal Institute of Technology, Manipal University, Manipal, Karnataka, India¹

Department of Statistics, Mangalore University, Mangalagangothri, Konaje 574199, Karnataka, India²

Abstract: Sumathi and Rao (2008) proposed a cure model for the recurrent event count data. The proposed model was based on the zero inflated Poisson (ZIP) distribution. Several tests were proposed for testing the cured proportion for recurrent event count data (Sumathi and Rao (2010)). In any health related studies, whether epidemiological or long term follow up studies, it is very essential for an investigator to record the information of patients regarding their demographic and socio-economic status as well as their medical history, at the time of clinical examination. These factors, commonly known as covariates, do play a vital role in influencing the health status of the individuals. The present paper is an extension of the work of Sumathi and Rao (2010). It proposes tests for testing the cured proportion in the presence of covariates when the data are uncensored. The covariates are related to the mean parameter of the proposed model, using the log link function. Although testing for $p = 0$ has been done in the past (Broek (1995)), testing for $p = p_0$ has not been studied for the ZIP model. The small sample performances of the proposed tests are studied using simulations.

Key words: cure model, recurrent event count data, inflated Poisson distribution, covariates, size and power of the test.

1. Introduction

The widely used model in survival analysis is the Cox (1972) proportional hazards (PH) model. This model is based on the assumption that every individual in the population under study is susceptible to an adverse event such as a disease, and will eventually experience this event if the follow-up is sufficiently long. But the advancement in recent medical research has shown that a large proportion of the diseased population is being cured of various chronic diseases such as epilepsy and cancer. Hence, the commonly used Cox (1972) PH model is not applicable here because of its assumption, which in turn has led to the development of cure models.

The cure models are survival models basically developed to estimate the proportion of patients cured from cancer of mouth and throat, cervix, uteri and breast. Using these models, the investigator can also study the causes for the failure of the treatment in the uncured group of patients and obtain the trends in the survival of patients suffering from cancer. The two cure models that are extensively used are the mixture model developed by Boag (1949) and the bounded cumulative hazard (BCH) model proposed by Yakovlev, et.al. (1993).

The mixture model and the BCH model are applicable only when the study variable is the time to occurrence of an adverse event. The models do not consider the possibilities of the recurrences of the events. For example, diseases like epilepsy and malaria could recur during the lifetime of an individual. Sumathi and Rao (2008) proposed a cure model for recurrent event count data. The proposed model turned out to be the zero inflated Poisson (ZIP) model.

Some recent references on zero inflated models include that of Bhattacharya, et. al. (2008), Williamson, et. al. (2007), Czado and Min (2005) and Gupta, et. al. (2004). While Bhattacharya, et. al. (2008) have considered a Bayesian test for testing excess zeroes in a zero inflated power series distribution, Williamson, et. al. (2007) have presented power calculations for ZIP models based on the conventional likelihood ratio and Wald tests. The study based on the performance of the score test on the zero inflated models can be found in Czado and Min (2005) and Gupta, et. al. (2004).

Several tests were proposed by Sumathi and Rao (2010) for testing the cured proportion in recurrent event count data when no covariates were considered. Kannan, et. al. (2010) have proposed a cure model based on generalized exponential distribution that incorporates the effects of covariates on the survival time of individuals.

In any health related studies, whether epidemiological or long term follow up studies, it is very essential for an investigator to record the information of patients regarding their demographic and socio-economic status as well as their medical history, at the time of clinical examination. These factors, commonly known as covariates, do play a vital role in influencing the health status of the individuals.

The present paper is an extension of the work of Sumathi and Rao (2010). While the work by Kannan, et. al. (2010) is applicable when the variable of interest is continuous, viz., the time to occurrence of an event, the present paper is applicable in situations where the variable of interest is discrete, i.e., the number of recurrences of an event. The paper proposes tests for testing the cured proportion where the covariates in the study are related to the mean parameter of the proposed cure model using a log link function. Although testing for $p = 0$ has been done in the past (Broek (1995)), testing for $p = p_0 (\neq 0)$ has not been studied for the ZIP model.

The rest of this paper is organized as follows. Section 2 discusses the estimation of the parameters of the model. In section 3, tests are proposed for testing the cured proportion, i.e., for testing $p = p_0$, in the presence of covariates. Section 4 gives the small sample performances of the tests with respect to their type I error rates and their powers. Section 5 concludes with a discussion.

The results of the simulations indicate that a variant of the Wald test (see section 3 for details) emerges as the best test in terms of maintaining type I error rates and having comparatively more power than the remaining tests. The small sample comparison of the powers of the proposed tests, indicate that the powers of the tests are more when a continuous covariate is used rather than when a discrete covariate is used. This conclusion strengthens the findings arrived at by Bhatta (2003) and does not encourage the categorization of a continuous covariate such that it becomes discrete, as in the case of tumor size being categorized as 1, 2, 3, etc., when the actual sizes of the tumors are known.

2. Notations and Estimation of parameters

As mentioned earlier, the model proposed by Sumathi and Rao (2008) turned out to be a ZIP model. The following notations have been used in the sequel with reference to the ZIP model.

A random variable Y is said to follow a zero inflated Poisson distribution with parameters p and λ , if its probability mass function (p. m. f.) is given by

$$P[Y = y] = \begin{cases} p + (1 - p)\exp(-\lambda), & \text{when } y = 0 \\ (1 - p)\exp(-\lambda)\frac{\lambda^y}{y!}, & \text{when } y = 1, 2, \dots \end{cases} \quad (1)$$

Here, p denotes the inflate parameter and λ denotes the mean parameter. Let Y_1, Y_2, \dots, Y_n denote the number of recurrences of the event of interest, for the n individuals who are on follow up for a sufficiently long period. Let $X_i' = (x_{i0}, x_{i1}, x_{i2}, \dots, x_{ik})$, with $x_{i0} = 1$, denote the values of the baseline covariates for the i^{th} individual. These covariates are related to the mean parameter λ_i of the random variable Y_i through the log link function $\log \lambda_i = X_i'\beta$, where $\beta' = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)$ denotes the vector of regression coefficients. Let the inflate parameter p of the p. m. f. given by (1) denote the cured proportion. The likelihood function based on the n observations is given by

$$L(p, \beta; y) = \prod_{i=1}^n \left(p + (1 - p)\exp(-\lambda_i(\beta)) \right)^{\delta_i} \left(\frac{(1 - p)\exp(-\lambda_i(\beta))\lambda_i(\beta)^{y_i}}{y_i!} \right)^{1 - \delta_i} \quad (2)$$

where $\lambda_i(\beta) = \exp(X_i'\beta)$ and $\delta_i = \begin{cases} 1, & y_i = 0 \\ 0, & y_i \neq 0 \end{cases}$

The maximum likelihood (ML) equation for p is given by

$$\sum_{i=1}^n \left(\frac{\delta_i(1 - \exp(-\lambda_i(\beta)))}{p + (1 - p)\exp(-\lambda_i(\beta))} - \frac{1 - \delta_i}{1 - p} \right) = 0 \quad (3)$$

The ML equations for the regression coefficients $\beta_j, j = 0, 1, 2, \dots, k$, are given by

$$\sum_{i=1}^n \left(\frac{-\delta_i x_{ij}(1 - p)\lambda_i(\beta)\exp(-\lambda_i(\beta))}{p + (1 - p)\exp(-\lambda_i(\beta))} + (1 - \delta_i)x_{ij}(y_i - \lambda_i(\beta)) \right) = 0 \quad (4)$$

The ML estimators of the parameters p and β are obtained by solving equation (3) and the $k+1$ equations of the form (4), simultaneously. In the present paper, we have used the Newton-Raphson’s iterative technique to obtain the ML estimators of the parameters.

Let the ML estimators of p and β be denoted by \hat{p} and $\hat{\beta}$ respectively. For carrying out inference

on p , the determination of the restricted ML estimator (MLE) of β (denoted by $\hat{\beta}$) when $p = p_0$ (specified)

is necessary. The vector $\hat{\beta}$ is the solution of the restricted ML equations given by

$$\sum_{i=1}^n \left(\frac{-\delta_i x_{ij} (1-p_0) \lambda_i(\beta) \exp(-\lambda_i(\beta))}{p_0 + (1-p_0) \exp(-\lambda_i(\beta))} + (1-\delta_i) x_{ij} (y_i - \lambda_i(\beta)) \right) = 0, \text{ for } j = 0, 1, 2, \dots, k \quad (5)$$

The expected Fisher information matrix is given by

$$F(p, \beta) = \begin{bmatrix} F_{(p,p)} & F_{(p,\beta_0)} & F_{(p,\beta_1)} & \dots & F_{(p,\beta_k)} \\ F_{(\beta_0,p)} & F_{(\beta_0,\beta_0)} & F_{(\beta_0,\beta_1)} & \dots & F_{(\beta_0,\beta_k)} \\ F_{(\beta_1,p)} & F_{(\beta_1,\beta_0)} & F_{(\beta_1,\beta_1)} & \dots & F_{(\beta_1,\beta_k)} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ F_{(\beta_k,p)} & F_{(\beta_k,\beta_0)} & F_{(\beta_k,\beta_1)} & \dots & F_{(\beta_k,\beta_k)} \end{bmatrix}$$

where the entries in $F(p, \beta)$ are given by

$$F_{(p,p)} = \sum_{i=1}^n \frac{(1 - \exp(-\lambda_i(\beta)))}{(1-p)(p + (1-p)\exp(-\lambda_i(\beta)))};$$

$$F_{(p,\beta_r)} = F_{(\beta_r,p)} = \sum_{i=1}^n \frac{-x_{i,r} \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p)\exp(-\lambda_i(\beta)))}, \text{ for } r = 0, 1, 2, \dots, k;$$

$$F_{(\beta_r,\beta_r)} = \sum_{i=1}^n \left(x_{i,r}^2 (1-p) \lambda_i(\beta) \left(1 - \frac{p \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p)\exp(-\lambda_i(\beta)))} \right) \right), \text{ for } r = 0, 1, 2, \dots, k;$$

$$F_{(\beta_r,\beta_s)} = F_{(\beta_s,\beta_r)} = \sum_{i=1}^n \left(x_{i,r} x_{i,s} (1-p) \lambda_i(\beta) \left(1 - \frac{p \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p)\exp(-\lambda_i(\beta)))} \right) \right),$$

for $r, s = 0, 1, 2, \dots, k, r \neq s$.

The following theorem establishes the asymptotic normality of the ML estimators of the proposed model.

Theorem: Under suitable regularity conditions,

$\left(F^{1/2} (\hat{p} - p, \hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2, \dots, \hat{\beta}_k - \beta_k) \right)'$ is asymptotically distributed as multivariate

$k+2$ component normal i.e., $N_{k+2}(O, I_{k+2})$. O is the mean vector and I_{k+2} is the variance-covariance matrix which is an identity matrix of dimension $k+2$.

$\left(F^{1/2} (\hat{p} - p, \hat{\beta}_0 - \beta_0, \hat{\beta}_1 - \beta_1, \hat{\beta}_2 - \beta_2, \dots, \hat{\beta}_k - \beta_k) \right)'$ is the Cholesky square root matrix of the Fisher

information matrix $F(p, \beta)$.

Proof: Follows from Theorem 1 of Czado and Min (2005).

3. Tests for cured proportion

Just as has been done in Sumathi and Rao (2010), we estimate the cured proportion p in a recurrent event count data set when covariates are considered and develop a confidence interval for the same. Since, the confidence interval can be obtained by the acceptance region of a test (Lehmann and Romano (2005)), tests are proposed for the cured proportion. To be specific, we shall test for the null hypothesis $H_0 : p = p_0$ against the two sided alternative $H_1 : p \neq p_0$. Thirteen tests are proposed, of which the first three are the conventional likelihood ratio, Wald and the score tests, while the remaining are the variants of the Wald and the score tests. The test statistics are as follows.

3.1 Likelihood Ratio Test (LRT)

The LRT statistic for testing the null hypothesis $H_0 : p = p_0$ is given by

$$W_1 = 2 \left(\log L(\hat{p}, \hat{\beta}; y) - \log L(p_0, \hat{\beta}; y) \right), \text{ which simplifies to}$$

$$W_1 = 2 \sum_{i=1}^n \left[\delta_i \log \frac{\hat{p} + (1 - \hat{p}) e^{-\lambda_i(\hat{\beta})}}{p_0 + (1 - p_0) e^{-\lambda_i(\hat{\beta})}} + (1 - \delta_i) \left\{ \log \frac{(1 - \hat{p})}{(1 - p_0)} - \left(\lambda_i(\hat{\beta}) - \lambda_i(\hat{\beta}) \right) + y_i \log \frac{\lambda_i(\hat{\beta})}{\lambda_i(\hat{\beta})} \right\} \right] \quad (6)$$

The asymptotic null distribution of W_1 is central chi-squared with one degree of freedom. The signed LRT statistic for testing one sided alternatives is given by

$$W_2 = \text{sgn}((\hat{p} - p_0) \sqrt{W_1}) \quad (7)$$

3.2 Wald Test

The Wald test statistic for testing the null hypothesis $H_0 : p = p_0$ is given by

$$W_3 = \frac{(\hat{p} - p_0)}{SE(\hat{p})} \quad (8)$$

where the standard error of \hat{p} , denoted as $SE(\hat{p})$, can be obtained from the inverse of the Fisher information matrix $F(p, \beta)$. The asymptotic null distribution of W_3 is standard normal.

3.3 Score Test

Rao (1947) proposed the score test. The advantage of this test is that when the hypothesis is simple, the computation of the MLEs is not required. Details of the score test are available in Rao (1973), Cox and Hinkley (1974) and Severini (2000). The score test statistic is given by

$$W_4 = U^1 F^{-1} U \quad (9)$$

where $U = \left[\frac{\partial}{\partial p} \log L; \frac{\partial}{\partial \beta_0} \log L; \frac{\partial}{\partial \beta_1} \log L; \dots; \frac{\partial}{\partial \beta_k} \log L \right]$ is the score vector, L is the likelihood

function given by (2), and F denotes the Fisher information matrix, the notation which we shall follow in the sequel, instead of $F(p, \beta)$. The parameters p and β_j , $j = 0, 1, 2, \dots, k$, in U and F are replaced by p_0

and $\hat{\beta}_j$, $j = 0, 1, 2, \dots, k$. The score functions in U are given by

$$\frac{\partial \log L \left(p, \beta; \underset{\sim}{y} \right)}{\partial p} = \sum_{i=1}^n \left(\frac{\delta_i (1 - \exp(-\lambda_i(\beta)))}{p + (1-p)\exp(-\lambda_i(\beta))} - \frac{(1-\delta_i)}{(1-p)} \right) \quad (10)$$

$$\frac{\partial \log L \left(p, \beta; \underset{\sim}{y} \right)}{\partial \beta_j} = \sum_{i=1}^n \left(\frac{-\delta_i x_{ij} (1-p) \lambda_i(\beta) \exp(-\lambda_i(\beta))}{p + (1-p)\exp(-\lambda_i(\beta))} + (1-\delta_i) x_{ij} (y_i - \lambda_i(\beta)) \right) \quad (11)$$

for $j = 0, 1, 2, \dots, k$. The asymptotic null distribution of W_4 is central chi-squared with one degree of freedom. For the one sided alternative hypothesis, the signed score test statistic, which has the same form as the signed LRT statistic given by (7), can be proposed. Asymptotically, the LRT, Wald and the score tests are equivalent under the null as well as the alternative hypotheses. The results follow from the standard asymptotics for the LRT, Wald and the score tests. For details, see Cox and Hinkley (1974) and Severini (2000). The small sample performances of these tests differ. The results of the simulations discussed in section 4.2 indicate the same.

3.4 Variants of the Wald and the Score Tests

The standard Wald test uses the estimates \hat{p} and $\hat{\beta}$ in the Fisher information matrix F to obtain the $SE(\hat{p})$, while the standard score test uses p_0 and $\hat{\beta}$ in F . Cox and Hinkley (1974), and Kale (1999) suggest that p_0 and $\hat{\beta}$, as well as p_0 and $\hat{\lambda}$ may be used in the Fisher information matrix F for the computation of the Wald test statistic. Similarly, p_0 and $\hat{\beta}$, and \hat{p} and $\hat{\beta}$ may be used in F for the score test. The three versions of the score tests which include the standard score test, are examined by Morgan, et. al. (2007). For the present hypothesis, let the variants of the Wald test statistic be denoted as W_5 and W_6 , and that of the score test statistic as W_7 and W_8 , respectively.

The test statistics W_5 and W_6 have the same forms as that of the conventional Wald test statistic W_3 given by (8) except that in the expression for $SE(\hat{p})$ of W_3 , p and β in F are replaced by p_0 and $\hat{\beta}$, respectively to obtain W_5 , and by p_0 and $\hat{\lambda}$, respectively to obtain W_6 . Similarly, W_7 is obtained by replacing p and β in F by p_0 and $\hat{\lambda}$, and W_8 on replacement with \hat{p} and $\hat{\beta}$, respectively. Again, these test statistics W_7 and W_8 have the same forms as that of the conventional score test statistic W_4 given by (9).

3.5 New Variants of the Score test

Motivated by the work of Sumathi and Rao (2010), six new variants of the score test are proposed in addition to the tests proposed above. The test statistics are as follows.

$$W_9 = \left(\frac{\partial}{\partial p_0} \log L \left(p_0; \hat{\beta}, \underset{\sim}{y} \right) \right) F_{p,p}^{-1/2} \left(p_0; \hat{\beta} \right) \quad (12)$$

$$W_{10} = \left(\frac{\partial}{\partial p_0} \log L \left(p_0; \hat{\beta}, \underset{\sim}{y} \right) \right) F_{p,p}^{-1/2} \left(p_0; \hat{\beta} \right) \quad (13)$$

$$W_{11} = \left(\frac{\partial}{\partial p_0} \log L \left(p_0; \hat{\beta}, \underset{\sim}{y} \right) \right) F_{p,p}^{-1/2} \left(p_0; \hat{\beta} \right) \quad (14)$$

$$W_{12} = \left(\frac{\partial}{\partial p_0} \log L \left(p_0; \hat{\beta}, y \right) \right) F_{p,p}^{-1/2} \left(p_0; \hat{\beta} \right) \tag{15}$$

$$W_{13} = \left(\frac{\partial}{\partial p_0} \log L \left(p_0; \hat{\beta}, y \right) \right) F_{p,p}^{-1/2} \left(\hat{p}; \hat{\beta} \right) \tag{16}$$

$$W_{14} = \left(\frac{\partial}{\partial p_0} \log L \left(p_0; \hat{\beta}, y \right) \right) F_{p,p}^{-1/2} \left(\hat{p}; \hat{\beta} \right) \tag{17}$$

4. Small sample performances of the test statistics

As has been done in Sumathi and Rao (2010), the performances of only thirteen test statistics $W_1, W_3, W_4, W_5, W_6, W_7, W_8, W_9, W_{10}, W_{11}, W_{12}, W_{13}$, and W_{14} were assessed based on a simulation study by including a covariate. The test W_2 has not been dealt with since it is the test statistic for one sided alternative, while in this paper, only the two sided alternative $H_1 : p \neq p_0$ is considered. The simulations were performed as follows.

4.1 Description of the simulation experiment

In the present study, only one covariate has been considered. However, this does not restrict the generality of the test statistics when there are $k = 1, 2, 3, \dots$ covariates. The single covariate X , has been linked to the mean parameter λ_i using the log link function $\log \lambda_i = \beta_0 + \beta_1 X_i$.

The performances of the proposed tests have been studied when the covariate was continuous as well as discrete. A binary variable was taken for the discrete case. For the continuous case, n observations were generated from standard normal distribution. The observations were then generated from an inflated Poisson distribution with parameters p and $\lambda_i = \exp(\beta_0 + \beta_1 X_i)$, $i = 1, 2, \dots, n$. For a given sample of size n , the covariates once generated were fixed and used throughout the 10000 simulations. For the discrete covariate, the first $n/2$ observations were taken as zero and the remaining were taken as one.

The simulation configurations were as follows. The level of significance was $\alpha = 0.05$. The sample sizes considered were $n = 20, 40, 80$. The regression coefficients were $\beta_0 = 0.25, \beta_1 = 0.2$. The values of the inflate parameter under the null hypothesis were $p_0 = 0.3, 0.5$. The probability distribution of the covariate was standard normal. The number of simulation was 10000. The values of β_0 and β_1 were determined such that the average value of λ would be 3. The observed Fisher information matrix was used for the simulations since the practitioners are more likely to use the observed Fisher information matrix rather than the expected Fisher information matrix to carry out the tests. For the zero inflated Poisson distribution considered in the present paper, the observed conditional Fisher information matrix is given by

$$O(p, \beta) = \begin{bmatrix} o_{(p,p)} & o_{(p,\beta_0)} & o_{(p,\beta_1)} & \dots & o_{(p,\beta_k)} \\ o_{(\beta_0,p)} & o_{(\beta_0,\beta_0)} & o_{(\beta_0,\beta_1)} & \dots & o_{(\beta_0,\beta_k)} \\ o_{(\beta_1,p)} & o_{(\beta_1,\beta_0)} & o_{(\beta_1,\beta_1)} & \dots & o_{(\beta_1,\beta_k)} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ o_{(\beta_k,p)} & o_{(\beta_k,\beta_0)} & o_{(\beta_k,\beta_1)} & \dots & o_{(\beta_k,\beta_k)} \end{bmatrix}$$

where
$$o_{(p,p)} = \sum_{i=1}^n \left(\frac{\delta_i (1 - \exp(-\lambda_i(\beta)))^2}{(p + (1-p)\exp(-\lambda_i(\beta)))^2} + \frac{1 - \delta_i}{1-p} \right);$$

$$o_{(p,\beta_r)} = o_{(\beta_r,p)} = \sum_{i=1}^n \left(\frac{-x_{i,r} \delta_i (1-p) (1 - \exp(-\lambda_i(\beta))) \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p) \exp(-\lambda_i(\beta)))^2} + \frac{x_{i,r} \delta_i \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p) \exp(-\lambda_i(\beta)))} \right)$$

for $r = 0, 1, 2, 3, \dots, k$;

$$o_{(\beta_r,\beta_r)} = \sum_{i=1}^n \left(\frac{\delta_i x_{i,r}^2 (1-p)^2 (\lambda_i(\beta))^2}{(p + (1-p) \exp(-\lambda_i(\beta)))^2} + \frac{\delta_i x_{i,r}^2 (1-p) (1 - \lambda_i(\beta)) \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p) \exp(-\lambda_i(\beta)))} + (1 - \delta_i) x_{i,r}^2 \lambda_i(\beta) \right)$$

for $r = 0, 1, 2, 3, \dots, k$;

$$o_{(\beta_r,\beta_s)} = o_{(\beta_s,\beta_r)}$$

$$= \sum_{i=1}^n \left(\frac{\delta_i x_{i,r} x_{i,s} (1-p)^2 (\lambda_i(\beta))^2}{(p + (1-p) \exp(-\lambda_i(\beta)))^2} + \frac{\delta_i x_{i,r} x_{i,s} (1-p) (1 - \lambda_i(\beta)) \lambda_i(\beta) \exp(-\lambda_i(\beta))}{(p + (1-p) \exp(-\lambda_i(\beta)))} + (1 - \delta_i) x_{i,r} x_{i,s} \lambda_i(\beta) \right)$$

for $r, s = 0, 1, 2, 3, \dots, k$, $r \neq s$.

The value of k was 1 since only one covariate has been used.

4.2 Small sample type I error rates

The performances of the tests with reference to the attained level of significance ($\hat{\alpha}$) are evaluated as follows. A test is said to maintain type I error rate if $\hat{\alpha}$ is in the interval $[0.04, 0.06]$ i.e., $\alpha \pm 0.01$. The test is liberal if $\hat{\alpha} > 0.06$ and conservative, if $\hat{\alpha} < 0.04$. The test becomes too conservative if $\hat{\alpha} < 0.02$ (i.e., 60% below the nominal level α). When two tests under consideration fail to maintain type I error rates, the test which is conservative is preferred to a test which is liberal.

In the following paragraphs, comparisons between the tests are made in the order of their performances with reference to their ability to maintain type I error rates. Table 1 gives the estimated values of type I error rates of the tests for sample sizes 20, 40 and 80 when the covariate was continuous. The following are the observations based on the table 1 reported.

When the covariate was continuous, out of the six combinations of n and p_0 considered, the Wald test W_3 and the score test W_4 maintain type I error rates at 4 configurations. For a sample of size $n=20$, the Wald test W_3 is slightly conservative at $p_0 = 0.3$ and is liberal when testing for $p_0 = 0.5$, while the score test W_4 is conservative when testing for both values of p_0 considered. When the sample size was increased to 40 and 80, both the tests maintained type I error rates when testing for both the values of p_0 that were considered. The score test W_4 is conservative at 2 configurations whereas the Wald test W_3 is conservative at one configuration and liberal at the other. Since a conservative test is preferred to a liberal test, the score test stands first followed by the Wald test.

The LRT W_1 and the variants of the Wald test viz., W_5 and W_6 maintain type I error rates at 3 configurations. The LRT W_1 maintains type I error rate for sample size $n=20$ and $p_0 = 0.3$. The test is liberal when testing for $p_0 = 0.5$ when the sample size is 20. For a sample of size 40, the test is conservative when testing for $p_0 = 0.3$ and is liberal when testing for $p_0 = 0.5$. For the sample size $n=80$, the test maintains type I error rate at both the values of p_0 that were considered. The variants of the Wald test viz., W_5 and W_6 maintain type I error rates when testing for $p_0 = 0.3$ for all the sample sizes considered but are liberal when testing for $p_0 = 0.5$. Since the LRT W_1 is liberal at only two configurations while the tests W_5 and W_6 are liberal at three configurations out of the six considered, the LRT is preferred to the tests W_5 and W_6 . Therefore, the LRT stands third in line based on the simulation experiment with respect to maintaining type I error rates. It has already been mentioned in the earlier paragraph that the score test W_4 stands first followed by the Wald test W_3 .

The performances of the variants of the score test *viz.*, W_7 and W_8 , and the new variant W_{13} , with respect to type I error rates, are alike. Each of them maintain type I error rates at two configurations out of six. The test W_7 maintains type I error rate when the sample size $n = 80$ for both the values of p_0 that were considered and is conservative at the remaining 4 configurations. The test W_8 maintains type I error rate when testing for $p_0 = 0.3$ when the sample sizes are 20 and 40 but is liberal at the remaining four configurations. The test W_{13} is conservative at two configurations and liberal at the remaining two. The new variants W_{10} , W_{12} and W_{14} maintain type I error rates at one configuration each and are liberal at the remaining five configurations, while the tests W_9 and W_{11} are conservative at all the six configurations.

Table 2 gives the estimated values of the type I error rates of the thirteen tests for sample sizes 20, 40 and 80 when the covariate was discrete. The traditional LRT W_1 , the Wald test W_3 and the score test W_4 maintained type I error rate at five configurations out of six, thereby being first in line. These tests were conservative when testing for $p_0 = 0.3$ for a sample of size 20.

A variant of the score test *viz.*, W_7 maintained type I error rate at four configurations. It maintained type I error rate for sample sizes 40 and 80 when testing for both the values of p_0 that were considered. The test was conservative at the remaining 2 configurations. It was conservative when the sample size was 20 while testing for both the values of p_0 . The variants W_8 and W_{13} of the score test and the variants of the Wald test *viz.*, W_5 and W_6 maintained type I error rates at 3 configurations each. The test W_{13} was conservative at the remaining three, while the tests W_8 , W_5 and W_6 were conservative at 1 configuration and liberal at the remaining 2. The new variants of the score test namely, W_{10} , W_{12} and W_{14} maintained type I error rates at 1 configuration, became conservative at 1 configuration and were liberal at the remaining 4, while the tests W_9 and W_{11} were conservative at all the 6 configurations.

Table 3 summarizes the performance of the thirteen tests with respect to the number of times the tests maintain type I error rates, the number of times they become conservative and the number of times they are liberal, for the various simulation configurations considered. Based on the tables 1, 2 and 3, it is clear that the score test W_4 stands first followed by the Wald test W_3 and the likelihood ratio test (LRT) W_1 , irrespective of the covariate being continuous or discrete, when evaluating the performances of the proposed tests based on type I error rates.

4.3 Small sample power comparisons

This section presents the small sample power comparisons of the tests. The simulation configurations are similar to those presented in section 4.1 except that the observations are generated from the values of p and $\lambda_i = \exp(\beta_0 + \beta_1 X_i)$ under the alternative hypothesis, the covariate X_i remaining unchanged. During the estimation of the type I error rates, the lower and the upper $\alpha/2$ th percentiles of the null distribution of the test statistics were also recorded. The criteria for assessment of the tests based on the powers remain the same as that stated in section 5 of Sumathi and Rao (2010). The criteria are as follows.

- (a) Power of the test in the neighborhood of the null hypothesis $H_0 : p = p_0$ and
- (b) The rate at which the power function attains the value 1.

The power functions of the tests are obtained for various sample sizes by fixing the values of β_0 and β_1 . Figures 1 to 6 show the power functions of the tests for $n=20, 40$ and 80 , when the covariate was continuous. The performances of the tests are assessed only in those cases where they maintain type I error rates.

Figures 1 and 4 show that the powers of some of the tests are very low (< 0.1), when the sample size was $n=20$. As seen in the absence of covariates (Sumathi and Rao (2010)), the score test and its variants exhibited fluctuations in the power functions (decreasing, or randomly decreasing and increasing after a certain stage) rather than increasing on either sides of the specified value of the inflate parameter. For details, see Sumathi and Rao (2010), and also Sumathi and Rao (2011). But, as the sample size increased to 40 and 80, the

power of all the tests increased considerably on either sides of p_0 . The powers of all the tests were considerably more on the left of p_0 than at an equidistant position on its right.

The rate of convergence of the power functions to the value 1 also acts as an indicator for evaluating the performance of the tests. From the figures 1 to 6, it is evident that the Wald test W_3 has more power for all alternatives in the neighborhood of p_0 compared to the remaining tests. Moreover, the power function of the test approaches 1 faster when compared to the remaining tests. Therefore, the Wald test W_3 emerges as the best test followed by its variant W_5 .

Figures 7 to 12 show the power functions of the tests for $n=20, 40$ and 80 , when the covariate was made discrete. From the figures, it is evident that for the discrete covariate, the conclusions remain the same as that of continuous covariate, for all the sample sizes that have been considered here. The Wald test W_3 has again more power for all the alternatives in the neighborhood of p_0 when compared to the remaining tests. The rate of convergence of the power function of the test to 1 is also faster when compared to the remaining tests and thus emerges as the best test. Its variant W_5 stands second in line.

When the covariate was made discrete, the numerical values of the power of the tests were less compared to that when a continuous covariate was used. It is a common practice among the applied researchers to convert a discrete covariate into a continuous covariate. The results of the simulations suggest that it is not advisable to do so. The fluctuations in the values of the power functions of the score test and its variants were observed again as has been seen when a continuous covariate was used. The power functions of the LRT W_1 , the Wald test and its variant W_6 also exhibit inconsistency when the proportion of zeroes was increased to $p_0 = 0.5$.

The simulations are restricted to the sample sizes $n=20, 40$ and 80 because as the sample size increases, the simulations were taking considerably more time. It was observed that the increase in the duration was mainly for the estimation of the parameters because the number of iterations required for the convergence of the Newton-Raphson method was considerably more.

4.4 Small sample performances of the tests when testing for $p_0 = 0$

In this section, we present the small sample performances of the tests with respect to type I error rates and powers, when testing for $p_0 = 0$. Broek (1995) proposed score test for testing the hypothesis $H_0 : p_0 = 0$. Earlier, tests for the same hypothesis were derived by Cochran (1954) and Rao and Chakravarti (1956). However, these two tests are applicable only when there are no covariates. When the parameter space for the inflate parameter p is the closed interval $[0,1]$, 0 is a boundary point of the parameter space. The asymptotic null distributions of the likelihood ratio test, the Wald test and any other test which uses the unrestricted maximum likelihood estimators of the parameters when testing for hypothesis $H_0 : p_0 = 0$, are not chi-squared with one degree of freedom. (For details, see Czado and Min (2005).) We have also conducted simulations to study the performances of the various tests proposed in this paper, for testing the hypothesis $H_0 : p_0 = 0$ under the assumption that the parameter space for the inflate parameter p is the interval $(-\varepsilon, 1]$, the value of ε is that value of p which satisfies the relation $p + (1-p)e^{-\lambda} > 0$. A similar parameter space is used by Bhattacharya, et. al. (2008) while proposing the Bayesian test for testing the hypothesis $H_0 : p_0 = 0$.

Table 4 gives the estimated values of type I error rates of the tests for sample sizes $20, 40$ and 80 while testing for $p_0 = 0$ when a continuous covariate was used. For a sample of size $n=20$, score test W_4 and a variant of the Wald test viz., W_5 maintain type I error rate. When the sample of size was increased to 40 , the Wald test W_3 , its variant W_5 and two variants of the score test viz., W_7 and W_{12} maintain the same. When the sample size was 80 , the LRT W_1 and the Wald test W_3 maintain type I error rate. The variant W_6 of the Wald test and the variants W_8, W_9, W_{11} and W_{13} of the score test are conservative, whereas the variants W_{10} and W_{14} are liberal for all sample sizes considered. From the above results, it is clear that the variant W_5 of the

Wald test performs well for small sample sizes such as $n=20, 40$ and for moderate sample sizes, the LRT and the Wald test may be used.

Table 5 gives the estimated values of type I error rates of the tests for sample sizes 20, 40 and 80 while testing for $p_0 = 0$ when a discrete covariate was used. The table shows that the score test W_4 maintains type I error rate for all the sample sizes considered while the LRT maintains the same for $n=40$ and 80. The test is conservative for $n=20$. The Wald test W_3 and the variants W_7, W_9 and W_{11} of the score test are conservative for all sample sizes, while a variant W_5 of the Wald test and the variants W_8, W_{10} and W_{14} of the score test are liberal for all. The variant W_{12} of the score test is conservative for $n=20$ and 40, whereas the variant W_{13} is liberal for the same. Both these tests maintain type I error rate at $n=80$. The variant W_6 of the Wald test maintains type I error rate at $n=20$ and is liberal for $n=40$ and 80. From these observations, it is clear that when a discrete covariate is used in a ZIP model, the score test W_4 performs well based on type I error rate.

Figures 13, 14 and 15 show the power functions of the tests for $n=20, 40$ and 80, when testing for $p_0 = 0$ in the presence of a continuous covariate while figures 16, 17 and 18 show the same when testing for $p_0 = 0$ in the presence of a discrete covariate. The graphs show that the powers of the tests are very low. The fluctuations are again seen in the power functions of some of the tests as seen in the case of testing for $p_0 = 0.3$ and 0.5. The power of a variant W_7 of the score test is more compared that of the remaining tests. The variants W_{12} and W_{10} follow second in line. But as mentioned earlier, the power comparisons of the tests become meaningful only when they maintain type I error rates. Since these tests are either conservative or liberal at many configurations that have been considered, these tests cannot be recommended. Among the tests which maintain type I error rate, the variant W_5 of the Wald test has more power and thereby emerges as the best. The conclusion here remains the same as that of testing for $p_0 = 0.3$ and 0.5.

5. Discussion and Conclusion

In this paper, we have compared the small sample performance of the tests for testing the cured proportion in the presence of a single covariate. The performances were studied based on a continuous and a discrete covariate. The variant of the Wald test *viz.*, W_5 , where the specified value of the inflate parameter p and the unrestricted MLEs $\hat{\beta}_0$ and $\hat{\beta}_1$ of the regression coefficients are used, emerges as the best test. Although the Wald test has maximum power, as in the case of score test and its variants, it also exhibits fluctuations in the powers for a small sample when a discrete covariate is considered. One of the variants of the score test *viz.*, W_{12} also has maximum power next to the Wald test but since it maintains type I error rates at lesser number of places compared to all the other tests, it cannot be recommended to the practitioners.

Our conclusion regarding the performance of the Wald test as the best for the inflated Poisson distribution, agrees well with the conclusion of Czado and Min (2005). It is also observed that a minimum sample of size 80 is required so that the tests maintain type I error rates. The power of the tests is considerably more on the left of the specified value p_0 , than at a point which is equidistant on the right of p_0 . The numerical values of the powers of the tests were less when the covariate was discrete when compared to that when a continuous covariate was used. It is a common practice among the applied researchers to make a continuous covariate as discrete, as in the case of tumor size being categorized as 1, 2, 3, etc., when the actual sizes are known. But, the powers of the tests obtained from the simulations in the present study, indicates that it is not advisable to do so.

Testing for $p_0 = 0$ has also been carried out when covariates are included in the recurrent event count data cure model. In the past, the performances of the likelihood ratio, Wald and score tests for testing $p_0 = 0$ have been examined by Broek (1995), Gupta, et. al. (2004), Czado and Min (2005), Williamson, et. al. (2007) and Bhattacharya, et. al. (2008). The asymptotics for the likelihood ratio and Wald tests is valid only when the parameter space for the inflate parameter p is $[-\varepsilon, 1]$. This parameter space is meaningful for a ZIP model but lacks interpretation for a cure model. Although the present paper is regarding a cure model, the investigation regarding the performances of the various tests is carried out when the parameter space is $(-\varepsilon, 1]$ as the tests are valid for a ZIP model (not necessarily a cure model).

Williamson, et. al. (2007) have derived the asymptotic expressions of the power functions of the likelihood ratio, Wald and score tests for a two sample ZIP model. The small sample power comparisons of the three tests were also studied. The sample size used for simulations in Broek (1995), Gupta, et. al. (2004), Czado and Min (2005), Williamson, et. al. (2007) and Bhattacharya, et. al. (2008) was moderate to large, whereas the sample size used in the present study was small. Thus the results obtained from the present investigation complements the conclusions arrived at in the earlier studies.

Bhattacharya, et. al. (2008) have derived a Bayesian test for testing $p = 0$. The Bayesian tests can be extended for testing $p = p_0 (\neq 0)$. However, the finite sample computations of these tests being intensive, calls for a separate investigation and is beyond the scope of the present paper.

Table 1: Type I error rates of the tests when a continuous covariate is used

n	p ₀	W ₁	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
20	0.3	0.0471	0.0339	0.0276	0.0497	0.0477	0.0059	0.0492	0.0015	0.0523	0.0008	0.0382	0.0470	0.0508
40	0.3	0.0297	0.0401	0.0406	0.0499	0.0476	0.0202	0.0526	0.0061	0.0748	0.0054	0.0699	0.0136	0.0695
80	0.3	0.0433	0.0489	0.0461	0.0538	0.0570	0.0423	0.0621	0.0124	0.1024	0.0137	0.1016	0.0131	0.1014
20	0.5	0.0964	0.0833	0.0306	0.1544	0.1594	0.0175	0.1771	0.0082	0.0780	0.0090	0.0719	0.0977	0.1530
40	0.5	0.0675	0.0503	0.0496	0.1236	0.1297	0.0283	0.1355	0.0173	0.0936	0.0171	0.0957	0.0729	0.1355
80	0.5	0.0536	0.0457	0.0549	0.0810	0.0854	0.0452	0.0903	0.0196	0.1022	0.0185	0.1003	0.0401	0.1135

Table 2: Type I error rates of the tests when a discrete covariate is used

n	p ₀	W ₁	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
20	0.3	0.0132	0.0224	0.0205	0.0162	0.0140	0.0049	0.0151	0.0021	0.0293	0.0020	0.0231	0.0111	0.0228
40	0.3	0.0441	0.0469	0.0507	0.0452	0.0441	0.0428	0.0453	0.0208	0.0789	0.0234	0.0767	0.0101	0.0595
80	0.3	0.0519	0.0511	0.0531	0.0521	0.0506	0.0501	0.0529	0.0298	0.0848	0.0319	0.0832	0.0228	0.0765
20	0.5	0.0600	0.0491	0.0410	0.1053	0.1131	0.0298	0.1101	0.0190	0.0533	0.0198	0.0569	0.0550	0.0952
40	0.5	0.0561	0.0513	0.0520	0.0749	0.0817	0.0409	0.0802	0.0270	0.0755	0.0250	0.0737	0.0469	0.0825
80	0.5	0.0506	0.0462	0.0490	0.0591	0.0588	0.0461	0.0600	0.0282	0.0729	0.0287	0.0731	0.0458	0.0749

Table 3. Summary of the simulations with respect to type I error rates based on tables 1 and 2 (Values given within brackets are the percentages)

Feature of Type I error rate	Covariate Type	W ₁	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
		Maintains	Continuous	3 (50)	4 (67)	4(67)	3 (50)	3(50)	2(33)	2(33)	0 (00)	1(17)	0(00)	1(17)
	Discrete	5 (83)	5(83)	5(83)	3(50)	3(50)	4(67)	3(50)	0 (00)	1(17)	0(00)	1(17)	3(50)	1(17)
Conservative	Continuous	1(17)	1(17)	2(33)	0(00)	0(00)	4(67)	0(00)	6(100)	0(00)	6(100)	0(00)	2(33)	0(00)
	Discrete	1(17)	1(17)	1(17)	1(17)	1(17)	2(33)	1(17)	6(100)	1(17)	6(100)	1(17)	3(50)	1(17)
Liberal	Continuous	2(33)	1(17)	0(00)	3(50)	3(50)	0(00)	4(67)	0(00)	5(83)	0(00)	5(83)	2(33)	5(83)
	Discrete	0(00)	0(00)	0(00)	2(33)	1(17)	0(00)	2(33)	0(00)	4(67)	0(00)	4(67)	0(00)	4(67)

Table 4: Type I error rates of the tests while testing for $p_0 = 0$ when a continuous covariate is used

n	W ₁	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
20	0.0272	0.0292	0.0423	0.0514	0.0246	0.0380	0.0316	0.0010	0.1565	0.0006	0.0052	0.0273	0.2338
40	0.0391	0.0415	0.0671	0.0532	0.0144	0.0413	0.0327	0.0000	0.1987	0.0000	0.0528	0.0206	0.2628
80	0.0501	0.0542	0.0712	0.0651	0.0209	0.0329	0.0382	0.0009	0.187	0.0000	0.0868	0.0154	0.2363

Table 5: Type I error rates of the tests while testing for $p_0 = 0$ when a discrete covariate is used

n	W ₁	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
20	0.0326	0.0116	0.0459	0.1340	0.0590	0.0260	0.0965	0.0004	0.1506	0.0000	0.0019	0.0775	0.3232
40	0.0511	0.0314	0.0533	0.1438	0.0688	0.0140	0.0947	0.0011	0.1405	0.0000	0.0285	0.0727	0.2787
80	0.0512	0.0348	0.0455	0.1237	0.0717	0.0108	0.0823	0.0026	0.1207	0.0000	0.0457	0.0581	0.2246

Power functions of the tests when a continuous covariate was used

Figure 1: sample size $n = 20, p_0 = 0.3$

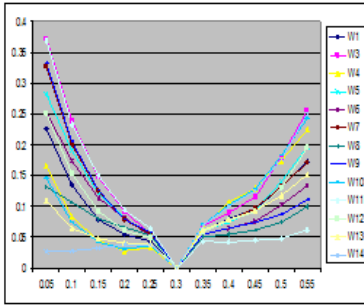


Figure 2: sample size $n = 40, p_0 = 0.3$

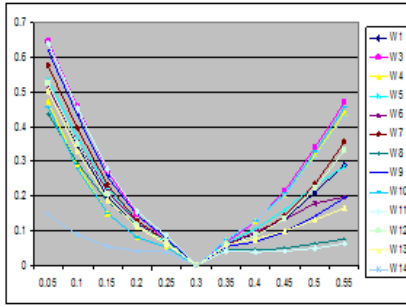


Figure 3: sample size $n = 80, p_0 = 0.3$

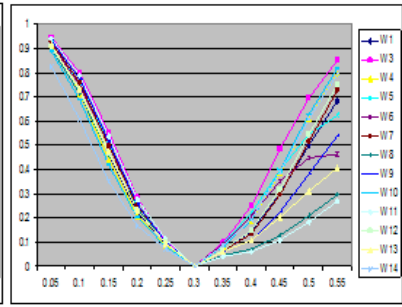


Figure 4: sample size $n = 20, p_0 = 0.5$

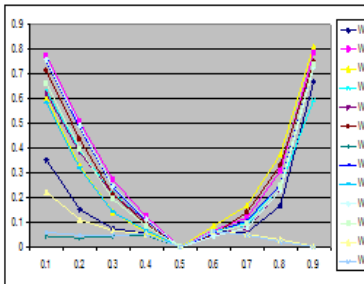


Figure 5: sample size $n = 40, p_0 = 0.5$

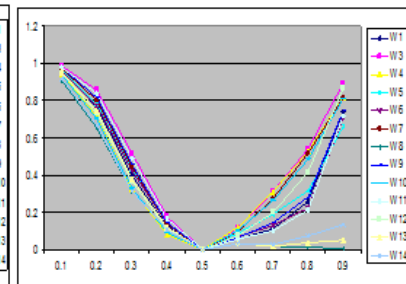
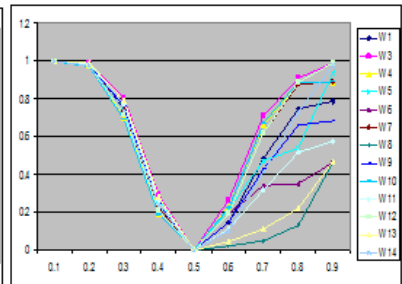


Figure 6: sample size $n = 80, p_0 = 0.5$



Power functions of the tests when a discrete covariate was used

Figure 7: sample size $n = 20, p_0 = 0.3$

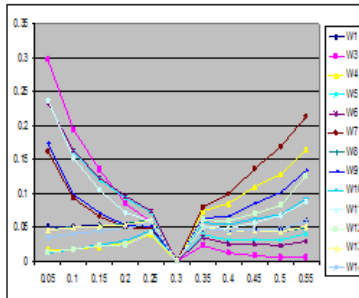


Figure 8: sample size $n = 40, p_0 = 0.3$

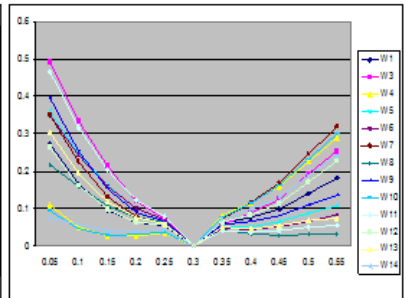


Figure 9: sample size $n = 80, p_0 = 0.3$

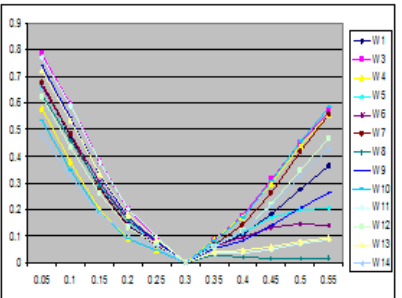


Figure 10: sample size $n = 20, p_0 = 0.5$

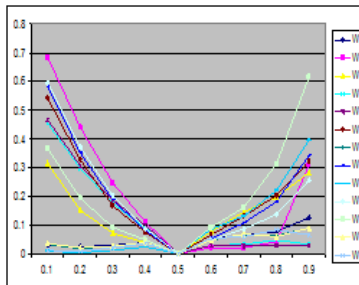


Figure 11: sample size $n = 40, p_0 = 0.5$

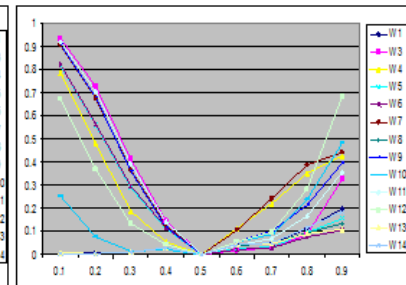
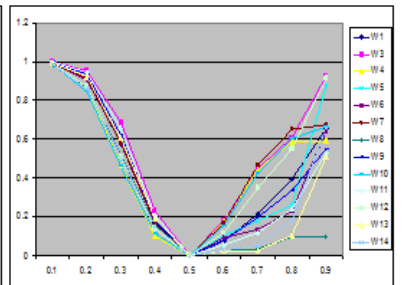


Figure 12: sample size $n = 80, p_0 = 0.5$



Power functions of the tests while testing for $p_0 = 0.0$ when a continuous covariate was used

Figure 13: sample size $n = 20, p_0 = 0.0$

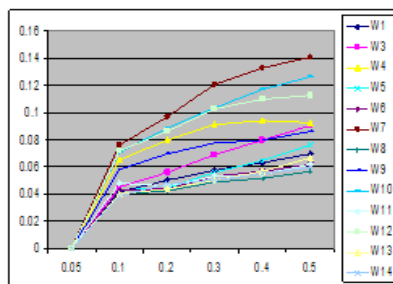


Figure 14: sample size $n = 40, p_0 = 0.0$

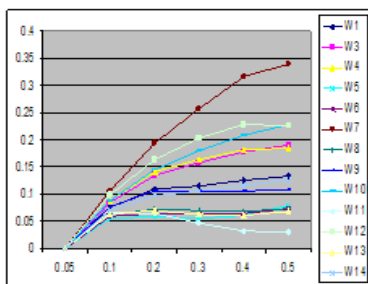
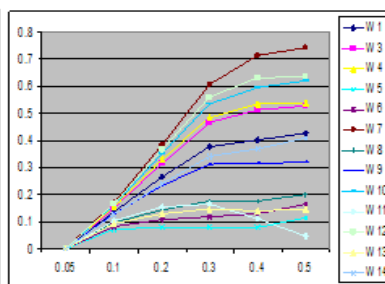


Figure 15: sample size $n = 80, p_0 = 0.0$



Power functions of the tests while testing for $p_0 = 0.0$ when a discrete covariate was used

Figure 16: sample size $n = 20, p_0 = 0.0$

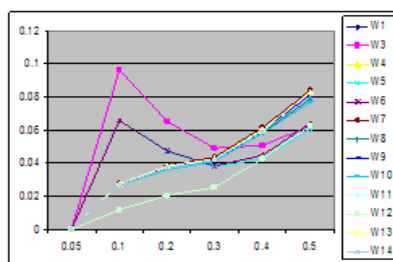


Figure 17: sample size $n = 40, p_0 = 0.0$

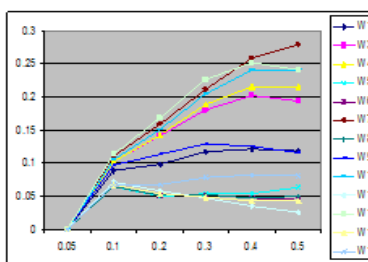
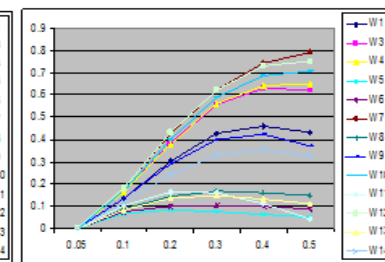


Figure 18: sample size $n = 80, p_0 = 0.0$



References

- [1]. Bhatta, A. R. S. (2003). Some Contributions to Inference in Generalized Linear Models and their Applications in Agriculture. Unpublished Ph. D. thesis, Department of Statistics, Mangalore University, Karnataka, India.
- [2]. Bhattacharya, A., Clarke, B.S. and Datta, G.S. (2008). A Bayesian test for excess zeros in a zero-inflated power series distribution. IMS collections. Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K Sen. Vol. 1, 89-104.
- [3]. Boag, J. W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society, Series B 11, 15-44.
- [4]. Broek, J. (1995). A score test for zero inflation in a Poisson distribution. Biometrics, 51, 738-743.
- [6]. Cochran, W. G. (1954). Some methods of strengthening χ^2 tests. Biometrics, 10, 417-451. MR0067428
- [7]. Cox, D.R (1972). Regression models and life tables (with discussion). Journal of the Royal Statistical Society, Series B 34, 187-220.
- [8]. Cox, D.R. and Hinkley, D.V. (1974). Theoretical Statistics. Chapman and Hall : London.
- [9]. Czado, C and Min, A. (2005). Consistency and asymptotic normality of the maximum likelihood estimator in a zero-inflated generalized Poisson regression. Sonderforschungsbereich 386, Paper 423. (<http://epub.ub.uni-muenchen.de/>)
- [10]. Gupta, P. L., Gupta, R. C. and Tripathi, R. C. (2004). Score test for Zero Inflated Generalized Poisson Regression Model. Communications in Statistics-Theory and Methods, Vol. 33, No. 1, pp. 47-64.
- [11]. Kale, B. K. (1998). Optimal estimating equations for discrete data with higher frequencies at a point. Journal of the Indian Statistical Association, 36, 125-136.
- [12]. Kale, B. K. (1999). A First Course on Parametric Inference. Narosa Publishing House, New Delhi.
- [13]. Kannan, N., Kundu, D., Nair, P. and Tripathi, R. C. (2010). The Generalized exponential cure rate model with covariates. Journal of Applied Statistics, Vol 37, No. 9-10, 12, 1625-1636.
- [14]. Lehmann, E.L. (1986). Testing Statistical Hypothesis. Second Edition. John Wiley & Sons, New York, Inc.
- [15]. Morgan, B. J. T., Palmer, K. J. and Ridout, M. S. (2007). Negative Score Test Statistic. The American Statistician, Vol. 61, No. 4, 285-288.
- [16]. Rao, C. R. (1947). Large Sample tests of Statistical Hypotheses concerning several parameters with applications to problems of estimation. Proceedings of the Cambridge Philosophical Society, 44, 50-57.
- [17]. Rao, C. R. and Chakravarti, I. M. (1956). Some small sample tests of significance for a Poisson distribution. Biometrics, 12, 264-282. MR0081596
- [18]. Rao, C. R. (1973). Linear Statistical Inference and its Applications. Second Edition. John Wiley & Sons, New York, Inc.
- [19]. Severini, T. A. (2000). Likelihood Methods in Statistics. Oxford University Press Inc, New York.
- [20]. Sumathi, K. and Rao, A. K. (2008). Cure rate models – A Partial Review with an application to recurrent event or count data. INTERSTAT.
- [21]. Sumathi, K. and Rao, A.K (2010). Tests for Cured Proportion for Recurrent Event Data. Optimisation and Statistics, 338-349.
- [22]. Sumathi, K. and Rao, A.K. (2011). A Note on Inconsistency of the Score test. Pakistan Journal of Statistics and Operation Research. Vol. VII, No. 1, pp 55-66.
- [23]. Williamson, J. M., Lin, H. M., Lyles, R. H. and Hightower, A. W. (2007). Power Calculations for ZIP and ZINB Models. Journal of Data Science, 5, 519-534.
- [24]. Yakovlev A.Y., Asselain, B., Bardou, V.J., Fourquet, A., Hoang, T., Rochefediere, A., and Tsodikov, A.D (1993) . A Simple Stochastic Model of Tumor Recurrence and Its Applications to Data on pre-menopausal Breast Cancer. In Biometrie et Analyse de Dormees Spatio – Temporelles 12 (Eds. B. Asselain, M. Boniface, C. Duby, C. Lopez, J.P.Masson, and J.Tranchefort). Société Francaise de Biométrie, ENSA Rennes, France, pp. 66-82.