ProbStat Forum, Volume 05, April 2012, Pages 54–61
ISSN 0974-3235

ProbStat Forum is an e-journal.
For details please visit www.probstat.org.in

On the use of marginal and conditional likelihood method for estimating the inflated parameter in zero inflated poisson regression model

Vasudeva Guddattu[a], K. Aruna Rao[b]

[a]*Department of Statistics, Manipal University, Manipal, Karnataka, India*
[b]*Department of Statistics, Mangalore University, Mangalore, Karnataka, India*

Abstract. In this paper we investigate utility of conditional and marginal likelihood approach for estimating inflation parameter form zero inflated Poisson regression. We consider a hospital based prospective study in which factors associated with number of adverse drug reactions is of interest. Further number of Adverse Drug Reactions (ADR) depends on length of stay in hospital. In first method we consider length of stay as an offset variable and estimate estimates of inflation parameter and factors associated with Number of ADR. In second method we consider length of stay as a random effect which follows exponential distribution and estimated estimate of inflation parameter, factors associated with mean response by considering marginal distribution of mean response. The marginal distribution turns out to be zero inflated geometric distribution. Further simulation study was performed to investigate the validity of the exponential distribution assumptions. Results from simulation study suggest that estimation of coverage probability and average length of confidence interval for inflation parameter by condition likelihood method performs better than marginal likelihood methods.

1. Introduction

In current days count regression is a wide tool among practitioners to analyze count data. We encounter count data in various disciplines. Logistic, Poisson, negative binomial, log binomial are commonly used models. These models do not fit if number of zero is more than determined by the corresponding probability distribution. Zero inflated models are alternative to these models where additional probability was assigned for occurrence of zero. The commonly used models are zero inflated Poisson (ZIP), zero inflated negative binomial models (ZIGP) and zero inflated generalized Poisson (ZIGP). Lambert (1992) used zero inflated Poisson regression for modeling number of defects in manufacturing. This model has received much attention in the literature in recent times. Van den Broek (1995) demonstrated its usefulness in modeling the frequency of urinary tract infection of HIV-infected men, where the majority of patients have zero number of episodes. Miaou (1994) analyzed the relationship between truck accidents and the geometric design of roads based on the ZIP structure. Bohning et al. (1999) further applied the ZIP model to study caries prevention in dental epidemiology.

In follow up studies such as cohort studies or prospective study number of times a condition recurred or number of adverse drug reactions (ADR) are of common interest. We consider hospital based prospective

Keywords. Zero inflated Poisson regression, conditional & marginal likelihood, coverage probability, average length of confidence interval

Received: 14 May 2011; Revised: 27 March 2012; Accepted: 17 April 2012

Email address: vasudev.guddattu@gmail.com (Vasudeva Guddattu)

study by Thiyagu et al. (2011) where information of number of adverse drug reaction(ADR), length of stay in hospital was collected. The objective of study is to find the factors associated with ADR. Further 59.2% of subjects are not having ADR. Since ZIP model fits well to the data, We used this model to find factors associated with ADR and to predict the proportion of people who are not having ADR. Further, Patel VK (2011) showed that number of days of hospital stay is a factor associated with ADR. In this paper we discuss conditional and marginal inferences for variables associated with mean response and estimation of inflation parameter W in zero inflated Poisson regression model by both approaches. In first approach we took length of stay as a offset variable and variables associated with mean response (λ), estimate of inflation parameter was estimated. In second approach we assumed length of stay in hospital as a random effect which follows exponential distribution with mean θ. Further marginal distribution of mean response with length of stay as exponentially distributed turn out to be geometric distribution with parameter p which depend on λ and θ. We had considered zero inflated geometric distribution to model inflation parameter W and mean response P. Inferences obtained from conditional and marginal likelihood methods for inflation parameter and mean response was compared. To validate the assumption of exponential distribution for length of stay in hospital, we empirically verified exponential distribution assumption by using a chi-square goodness of fit. Simulation study were performed to check the performance of both methods in estimating inflation parameter W. We estimated coverage probability for inflation parameter through simulations by generating random observation from zero inflated Poisson model with mean λ depending upon θ from exponential distribution.

The organization of paper is as follows. In Section 2 we present the notations and estimation of parameters by both approaches. In Section 3 we present case study and present results of estimates by both approaches. In Section 4 we present results of simulation study, conclusion and discussion in Section 5. Results of case study suggest that for factors associated with mean response by both methods will give same inferences in selection of variable and direction of association. Further from simulation study estimation of coverage probability and average length of confidence interval for inflation parameter through conditional likelihood method is performing better than estimation through marginal likelihood approach.

2. Notation and estimation of parameters

The probability mass function of zero inflated Poisson distribution with parameter W, λ is given by

$$f(y_i, w, \lambda_i) = \begin{cases} w + (1-w)e^{-\lambda_i} & y_i = 0, \\ \frac{(1-w)e^{-\lambda_i}\lambda_i}{yi} & y_i \neq 0, \end{cases} \quad \lambda > 0,\ 0 < w < 1.$$

Mean λ is model to set of covariates with offset variable t_i using log link function and is given by

$$\lambda_i = (1/t_i)\exp(B'X_i).$$

With $B = (B0, B_1, \dots, B_{k-1})$ a k dimensional vector of regression coefficients and $X_i = (X_{i0}, X_{i1}, \dots, X_{i,k-1})$ is a vector of covariates.

The conditional log likelihood function for zero inflated Poisson distribution with parameters w, λ with t_i as offset variable based on random sample of size n is given by

$$\log(L(Y, w, \lambda_i, t_i)) = \sum_{i=1}^{n} d_i \log(w + (1-w)\exp(-\exp(B'X_i)/t_i))$$
$$+ \sum_{i=1}^{n}(1-d_i)(\log(1-w) - \exp(B'X_i)/t_i + y_i\ \log(\exp(B'X_i)/t_i) - \log(y_i!)$$

where $d_i = 1$ if $y_i = 0$ else $d_i = 0$.

The Maximum likelihood (ML) equations for estimation of $w, B_0, \ldots, B_{k-1}$ is given by

$$\frac{\partial \log L}{\partial w} = \sum \frac{d_i\{1-\exp(-\exp(B'X_i)/t_i)\}}{[w+(1-w)\exp(-\exp(B'X_i)/t_i)]} - \sum \frac{(1-\partial_i)}{1-w}$$

$$\frac{\partial \log L}{B_j} = -\sum \frac{d_i(1-w)\exp(-\exp(B'X_i)t_i^{-1})\exp(B'X_i)t_i^{-1}X_{ij}\}}{[w+(1-w)\exp(-\exp(B'X_i)t_i^{-1})]} - \sum (1-d_i)(y_iX_{ij} - \exp(B'X_i)t_i^{-1}X_{ij}).$$

Solving above ML equation by equating it to zero yields ML estimator of regression parameters and inflation parameters. Newton Raphson iterative method was used for estimation. The estimate of standard errors of regression parameters and inflation parameters can be obtained from inverse of observed fisher information matrix O. The elements in O given by

$$\frac{\partial^2 \log L}{\partial w^2} = \sum \frac{d_i(1-\exp(t_i^{-1}\exp(-B_i'X_i)))^2}{[w+(1-w)\exp(-\exp(B'X_i)t_i^{-1})]^2} - \sum \frac{(1-d_i)}{(1-p)^2}$$

$$\frac{\partial^2 \log L}{\partial w \partial B_j} = -\sum \frac{d_i(\exp(t_i\exp(-B_iX_i))t_i^{-1}X_{ij}\exp(B_i'X_i)}{[w+(1-w)\exp(-\exp(B'X_i)t_i^{-1})]^2}$$

$$\frac{\partial^2 \log L}{\partial B_j^2} = \sum \frac{d_i\{(1-w)\exp(-\exp(B'X_i)t_i^{-1})\exp(B_i'X_i)t_i^{-1}X_{ij}}{[w+(1-w)\exp(-\exp(B'X_i)t_i^{-1})]^2}$$

$$-\sum \frac{d_i(1-w)\exp(-\exp(B'X_i)t_i^{-1})\exp(2B'X_i)t_i^{-1}X_{ij}^2\{1-t_i^{-1}\}}{[w+(1-w)\exp(-\exp(B'X_i)t_i^{-1})]} - \sum(1-d_i)(\exp(B'X_i)t_i^{-1}X_{ij}^2)$$

$$\frac{\partial \log L}{\partial B_j \partial B_k} = \sum \frac{d_i\{(1-w)\exp(-\exp(B'X_i)t_i^{-1})\exp(B'X_i)t_i^{-1}\}^2X_{ij}X_{ik}}{[w+(1-w)\exp(-exp(B'X_i)t_i^{-1})]^2}$$

$$-\sum \frac{d_i(1-w)\exp(-\exp(B'X_i)t_i^{-1})\exp(2B'X_i)t_i^{-1}X_{ij}X_{ik}\{1-t_i^{-1}\}}{[w+(1-w)\exp(-\exp(B'X_i)t_i^{-1})]}$$

$$\sum(1-d_i)(\exp(B'X_i)t_i^{-1}X_{ij}X_{ik})$$

Taking the distribution of length of stay as exponential with mean θ, the marginal distribution of Count outcome variable follow zero inflated geometric distribution with parameters W, λ, θ and is given by

$$f(y_i, w, \lambda_i, \theta) = \begin{cases} w+(1-w)P_i & y_i = 0 \\ (1-w)P_i(1-P_i)^{y_i} & y_i \neq 0 \end{cases} \qquad 0 < w,\ P_i < 1,\ P_i = \frac{\theta_i}{\lambda_i + \theta_i}$$

We will model P to set of covariates using logit link as

$$logit(Pi) = \log(Pi/(1-Pi)) = G'X_i$$

With $G = (G_0, G_1, \ldots, G_{k-1})$ a k dimensional vector of regression coefficients and $X_i = (X_{i0}, X_{i1}, \ldots, X_{ik-1})$ is a vector of covariates. We had used different notations for representing regression parameters used in conditional and marginal approach. Here we use G for regression parameters.

The log likelihood function is given by

$$\log(L) = \sum d(y_i)\log(w + \frac{(1-w)\exp(G'X_i)}{1+\exp(G'X_i)}) + \sum(1-d(y_i))(\log(1-w) + G'X_i - (y_i+1)\log(1+\exp(G'X_i))).$$

The Maximum likelihood equation for estimating regression parameter is given by

$$\frac{\partial \log L}{\partial w} = \sum \frac{d(y_i)}{w+\exp(G'X_i)} - \sum \frac{(1-d(y))}{(1-w)}$$

$$\frac{\partial \log L}{\partial G_j} = \sum \frac{d(y_i)X_{ij}\exp(G'X_i)(1-w)}{w+\exp(G'X_i)} + \sum \frac{(1-d(y_i))X_{ij}\exp(G'X_i)(1-y_i\exp(G'X_i))}{(1+\exp(G'X_i))}$$

Solving above ML equation by equating it to zero yields ML estimators of regression parameters G and inflation parameters w. Newton Raphson iterative method was used for estimation. The estimate of standard errors of regression parameters and inflation parameters can be obtained from inverse of observed fisher information matrix O1. The elements in O1 is given by

$$\frac{\partial^2 \log L}{\partial w^2} = -\sum \frac{d(y_i)}{(w + \exp(G'X_i))^2} - \sum \frac{(1 - d(y))}{(1 - w)^2}$$

$$\frac{\partial^2 \log L}{\partial w \partial G_j} = -\sum \frac{d(y_i)\exp(G'X_i)X_{ij}}{(w + \exp(G'X_i))^2}$$

$$\frac{\partial^2 \log L}{\partial G_k \partial G_j} = \sum \frac{d(y_i)(1 - w)\exp(G'X_i)X_{ij}X_{ik}}{(w + \exp(G'X_i))^2} - \sum \frac{(1 - d(y_i))\exp(G'X_i)(y_i + 1)X_{ij}X_{ik}}{(1 + \exp(G'X_i))^2}$$

$$\frac{\partial^2 \log L}{\partial G_{jj}^2} = \sum \frac{d(y_i)(1 - w)\exp(G'X_i)X_{ij}^2}{(w + \exp(G'X_i))^2} - \sum \frac{(1 - d(y_i))\exp(G'X_i)(y_i + 1)X_{ij}^2}{(1 + \exp(G'X_i))^2}$$

3. Case study

In this section we consider prospective study of Thiyagu (2011) in which factors associated with Adverse drug reaction (ADR) is studied. The outcome variable of interest is number of Adverse drug reaction and covariates used were age, gender and different types of drugs given. The identity of the drugs has been hidden in the data. For sake of simplicity we call these drugs as D1, D2, D3, and D4. Our intention is to estimate the proportion of no drug reactions. We had used chi-square goodness of fit for testing zero inflated Poisson, exponential distribution assumption for number of ADR and length of stay in hospital. Both distribution fit well with the observed data ($p = 0.901$ for Number of ADR and $P = 0.752$ for length of stay in hospital). There were 959 participants with 59.8% of people not having ADR. We used both conditional and marginal approaches for estimating inflation parameter and factors associated with number of ADR .The results obtained from both approaches are given in table 1.

Table 1: Estimate of inflation parameter and covariates associated with number of ADR by conditional and marginal likelihood

Conditional approach			Marginal Approach	
Variable	B(standard error)	P value	G(standard error)	P value
Inflation Parameter w	0.58 (0.05)	< 0.001	0.61 (0.07)	< 0.001
Age	0.036 (0.006)	$P < 0.001$	0.59 (0.23)	0.005
Gender (Females)	0.35 (0.16)	0.032	0.027 (0.003)	< 0.001
D1	-0.74 (0.29)	.012	-0.47 (0.13)	< 0.001
D2	-1.03 (0.24)	$P < 0.001$	-0.77 (0.32)	0.008
D3	0.42 (0.19)	.024	0.32 (0.065)	$P < 0.001$
D4	0.41 (0.12)	.038	0.28 (0.03)	$P < 0.001$

We can observe that in both methods estimate of inflation parameter are very close to observed one. Further the regression parameters obtained from both methods have same direction of magnitude. Also we adopted backward selection criteria for selection of variable. The probability of removal of variable at each step was fixed at 0.25 (Hosmer & Lemshow). Both methods arrive at same set of regression parameters in final step. In next section we present the results of simulation study for estimating coverage probability and average length of confidence interval when distribution of length of stay is correctly specified and misspecified.

4. Simulation study

A confidence interval for estimating a parameter of a probability distribution must show two basic properties. First, it must contain the value of the parameter with a prescribed probability (the confidence level) and second, it must be as narrow as possible in order to be useful. In practice a precise confidence interval is a one which has better coverage probability and smaller average length. In this section we estimate coverage probability (CP) and average length (AL) for the interval estimate of inflation parameter in ZIP by conditional and marginal likelihood approach by simulation study. We estimated CP and AL for inflation parameter in presence of discrete, continuous and no covariates for mean response. Further we had checked the validity of exponential distribution in estimating CP and AL by mispecifying the distribution. The simulation configuration for estimating CP and AL is as follows.

Table 2: Estimated coverage probability (CP) and average length of interval (AL) for inflation parameter with no covariates for mean response and distribution for length of stay is correctly specified

$N = 50$	$\lambda = 2, t \sim \exp(7)$		$\lambda = 4, t \sim \exp(7)$	
W	Marginal CP(AL)	Conditional CP(AL)	Marginal CP(AL)	Conditional CP(AL)
0.2	0.941 (0.22)	0.945 (0.21)	0.93 (0.23)	0.945 (0.21)
0.3	0.939 (0.25)	0.943 (0.23)	0.932 (0.26)	0.943 (0.23)
0.4	0.935 (0.29)	0.944 (0.24)	0.931 (0.27)	0.944 (0.24)
0.5	0.938 (0.28)	0.942 (0.27)	0.94 (0.29)	0.942 (0.27)
0.6	0.937 (0.27)	0.941 (0.23)	0.937 (0.23)	0.941 (0.23)
$N = 100$				
0.2	0.942 (0.18)	0.947 (0.14)	0.942 (0.17)	0.947 (0.14)
0.3	0.944 (0.2)	0.945 (0.17)	0.944 (0.18)	0.945 (0.17)
0.4	0.947 (0.21)	0.948 (0.15)	0.947 (0.19)	0.948 (0.15)
0.5	0.944 (0.18)	0.944 (0.13)	0.944 (0.19)	0.944 (0.13)
0.6	0.945 (0.19)	0.942 (0.14)	0.945 (0.17)	0.942 (0.14)

Level of Significance was fixed at 5% and sample size used for estimating AL and CP was 50, 100. The value of inflation parameter considered was 0.2, 0.3, 0.4, 0.5, 0.6. The value of mean response was set at $l = 2, 4$ in without covariates case. For discrete covariates we considered following two models namely $\log(l) = (1/t)\exp(-0.69 + .9X1)$, $\log(l) = (1/t)\exp(-0.69 - 0.28X2)$ with $X1$, $X2$ binary covariates generated from Bernoulli distribution with probability 0.6, 0.3.For continuous covariate we considered $\log(l) = (1/t)\exp(-0.6 + .5X3)$, $\log(l) = (1/t)\exp(-0.6 - 0.15X4)$, with $X3, X4$ generated from standard normal distribution. Distribution of length of stay was considered to be Exponential with mean 7 under correct specification and uniform (0,7) under misspecification. Total number of simulations carried out for estimation was 1,00,000. Matlab software was used for simulations.

Table 3: Estimated coverage probability (CP) and average length of interval (AL) for inflation parameter with no covariates for mean response and distribution for length of stay is misspecified

$N = 50$	$\lambda = 2, t \sim$ Uniform (0,7)		$\lambda = 4, t \sim$ Uniform (0,7)	
W	Marginal CP(AL)	Conditional CP(AL)	Marginal CP(AL)	Conditional CP(AL)
0.2	0.89 (0.27)	0.94 (0.21)	0.87 (0.31)	0.942 (0.21)
0.3	0.92 (0.29)	0.939 (0.22)	0.91 (0.32)	0.944 (0.23)
0.4	0.91 (0.24)	0.944 (0.25)	0.89 (0.28)	0.941 (0.24)
0.5	0.88 (0.29)	0.942 (0.28)	0.86 (0.31)	0.943 (0.27)
0.6	0.91 (0.27)	0.943 (0.24)	0.88 (0.29)	0.942 (0.23)
$N = 100$				
0.2	0.92 (0.22)	0.948 (0.17)	0.91 (0.25)	0.947 (0.14)
0.3	0.93 (0.24)	0.946 (0.16)	0.93 (0.28)	0.945 (0.12)
0.4	0.93 (0.2)	0.948 (0.17)	0.91 (0.27)	0.946 (0.14)
0.5	0.92 (0.22)	0.944 (0.18)	0.92 (0.25)	0.946 (0.15)
0.6	0.92 (0.25)	0.942 (0.16)	0.91 (0.24)	0.945 (0.12)

Table 2–3 presents the results of simulation study for estimation CP and AL for inflation parameter with no covariates for mean response with distribution of length of stay correctly specified and misspecified. Further the estimated CP and AL for estimation of inflation parameter with discrete, continuous covariate for mean response when distribution of length of stay is correctly specified and misspecified is given in Table 4–7.

Table 4: Estimated coverage probability (CP) and average length of interval (AL) for inflation parameter with binary covariates for mean response and distribution for length of stay is correctly specified

$N = 50$	$\log(\lambda) = (1/t)\exp(-0.69 + 0.9X1)$, $t \sim \exp(7)$		$\log(\lambda) = (1/t)\exp(-0.69 - 0.28X2)$, $t \sim \exp(7)$	
W	Marginal CP(AL)	Conditional CP(AL)	Marginal CP(AL)	Conditional CP(AL)
0.2	0.931 (0.26)	0.942 (0.25)	0.932 (0.25)	0.943 (0.24)
0.3	0.935 (0.25)	0.941 (0.24)	0.937 (0.26)	0.945 (0.25)
0.4	0.917 (0.29)	0.943 (0.27)	0.924 (0.29)	0.941 (0.27)
0.5	0.927 (0.25)	0.941 (0.23)	0.932 (0.24)	0.942 (0.23)
0.6	0.933 (0.27)	0.944 (0.25)	0.939 (0.27)	0.944 (0.23)
N=100				
0.2	0.942 (0.25)	0.947 (0.24)	0.941 (0.24)	0.947 (0.21)
0.3	0.944 (0.23)	0.946 (0.23)	0.943 (0.23)	0.949 (0.23)
0.4	0.937 (0.25)	0.946 (0.24)	0.945 (0.27)	0.948 (0.24)
0.5	0.939 (0.24)	0.945 (0.211)	0.943 (0.21)	0.947 (0.21)
0.6	0.942 (0.29)	0.948 (0.21)	0.943 (0.23)	0.949 (0.22)

Table 5: Estimated coverage probability (CP) and average length of interval (AL) for inflation parameter with binary covariates for mean response and distribution for length of stay is misspecified

$N = 50$	$\log(\lambda) = (1/t)\exp(-0.69 + 0.9X1)$, $t \sim$ Uniform (0,7)		$\log(\lambda) = (1/t)\exp(-0.69 - 0.28X2)$, $t \sim$ Uniform (0,7)	
W	Marginal CP(AL)	Conditional CP(AL)	Marginal CP(AL)	Conditional CP(AL)
0.2	0.88 (0.27)	0.941 (0.26)	0.87 (0.28)	0.942 (0.26)
0.3	0.89 (0.29)	0.942 (0.25)	0.89 (0.27)	0.944 (0.25)
0.4	0.91 (0.29)	0.941 (0.26)	0.88 (0.29)	0.942 (0.27)
0.5	0.91 (0.26)	0.943 (0.25)	0.91 (0.25)	0.943 (0.24)
0.6	0.92 (0.26)	0.947 (0.25)	0.90 (0.26)	0.944 (0.24)
N=100				
0.2	0.91 (0.24)	0.949 (0.24)	0.911 (0.26)	0.947 (0.24)
0.3	0.91 (0.27)	0.948 (0.23)	0.917 (0.25)	0.948 (0.23)
0.4	0.92 (0.27)	0.947 (0.23)	0.912 (0.26)	0.947 (0.24)
0.5	0.92 (0.23)	0.947 (0.21)	0.933 (0.23)	0.948 (0.21)
0.6	0.93 (0.23)	0.949 (0.21)	0.921 (0.23)	0.949 (0.21)

We can observe from above tables that the coverage probability for estimating inflation parameter is more for conditional likelihood method than marginal likelihood in all cases. Further coverage probability for marginal likelihood method reduces with distribution of length of stay is misspecified. The average length of interval is shorter for conditional likelihood approach in most of cases under different combinations of covariates specified for mean response l.

5. Discussion and conclusion

It is practice to use Poisson regression to estimate relative risk for count outcome in follow up studies. Further Guangyong and Zou suggested modified Poisson regression for estimating relative risk in prospec-

Table 6: Estimated coverage probability (CP) and average length of interval (AL) for inflation parameter with continuous covariates for mean response and distribution for length of stay is correctly specified

$N = 50$	$\log(\lambda) = (1/t)\exp(-0.6 + 0.5X3)$, $t \sim \exp(7)$		$\log(\lambda) = (1/t)\exp(-0.6 - 0.15X4)$, $t \sim \exp(7)$	
W	Marginal CP(AL)	Conditional CP(AL)	Marginal CP(AL)	Conditional CP(AL)
0.2	0.93(0.25)	0.941(0.21)	0.87(0.28)	0.941(0.25)
0.3	0.917(0.24)	0.937(0.23)	0.89(0.27)	0.942(0.21)
0.4	0.924(0.27)	0.942(0.24)	0.88(0.29)	0.943(0.27)
0.5	0.918(0.26)	0.941(0.24)	0.91(0.25)	0.942(0.24)
0.6	0.931(0.24)	0.939(0.22)	0.937(0.23)	0.941(0.22)
N=100				
0.2	0.940(0.23)	0.949(0.18)	0.939(0.23)	0.947(0.24)
0.3	0.931(0.21)	0.946(0.19)	0.942(0.21)	0.949(0.2)
0.4	0.937(0.24)	0.948(0.21)	0.937(0.27)	0.947(0.22)
0.5	0.933(0.23)	0.947(0.19)	0.935(0.22)	0.948(0.21)
0.6	0.942(0.21)	0.946(0.21)	0.941(0.23)	0.946(0.21)

Table 7: Estimated coverage probability (CP) and average length of interval (AL) for inflation parameter with continuous covariates for mean response and distribution for length of stay is correctly misspecified

$N = 50$	$\log(\lambda) = (1/t)\exp(-0.6 + 0.5X3)$, $t \sim$ Uniform(0,7)		$\log(\lambda) = (1/t)\exp(-0.6 - 0.15X4)$, $t \sim$ Uniform(0,7)	
W	Marginal CP(AL)	Conditional CP(AL)	Marginal CP(AL)	Conditional CP(AL)
0.2	0.912(0.29)	0.942(0.24)	0.891(0.27)	0.941(0.25)
0.3	0.907(0.25)	0.941(0.23)	0.912(0.24)	0.942(0.23)
0.4	0.903(0.38)	0.944(0.25)	0917(0.28)	0.945(0.27)
0.5	0.917(0.29)	0.943(0.26)	0.920(0.26)	0.947(0.23)
0.6	0.891(0.28)	0.942(0.24)	0.897(0.26)	0.943(0.23)
N=100				
0.2	0.920(0.27)	0.949(0.23)	0.922(0.26)	0.947(0.24)
0.3	0.927(0.23)	0.948(0.21)	0.929(0.24)	0.948(0.23)
0.4	0.93(0.27)	0.949(0.22)	0.931(0.27)	0.949(0.21)
0.5	0.932(0.24)	0.946(0.24)	0.937(0.23)	0.949(0.19)
0.6	0.910(0.26)	0.94(0.22)	0.932(o.23)	0.949(0.19)

tive studies. Also one can use random effect Poisson regression for estimating relative risk. It is practice to assume distribution of random effects as normal. In real situations this may not be appropriate. We considered a case study of number of hospitalized adverse drug reaction from 959 subjects where probability of zero is more than expected through Poisson regression model. Further number of adverse drug reactions depends on length of stay in hospital. We had used Zero inflated Poisson regression to model this data with length of stay as an offset variable and random effect. In later we assumed the distribution of random effect as exponential distribution. The parameters were estimated using conditional and marginal likelihood approach and results were compared. Both methods give same set of covariates and precise estimate of inflation parameter. Further to validate the assumption of exponential distribution for random effect we performed simulation study to investigate coverage probability and average length of confidence interval of inflation parameter when distribution of random effect is correctly specified and misspecified. Results indicate that estimate of coverage probability and average length of confidence interval of inflation parameter was better for conditional likelihood approach than marginal likelihood approach. Further when distribution of random effects in misspecified, conditional likelihood method performs better than marginal likelihood method. In marginal likelihood approach, the coverage probability for interval estimate of inflation parameter drops down substantially when distribution of random effects is misspecified. In practice we recommend to use conditional likelihood method to model count outcome from a prospective study.

References

[1] Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing, Technometrics 34, 1–14.

[2] van den Broek, J. (1995) A score test for zero inflation in a Poisson distribution, Biometrics 51, 738–743.

[3] Miaou, S. P. (1994) The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, Accident Analysis and Prevention 26, 471–482.

[4] Bohning, D., Dietz, E., Schlattmann, P., Mendonca, L., Kirchner, U. (1999) The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology, Journal of the Royal Statistical Society, Series A. 162, 195–209.

[5] Patel, V.K., Acharya, L.T., Mallayasamy, S.R., Guddattu, V., Ramachandran, P. (2011) Potential Drug Interactions at One Indian Teaching Hospital, AMJ 4(1), 9-14.

[6] Thiyagu, R. (2011) Modeling of predictors for adverse drug reactions and pharmacoeconomic impact in a tertiary care hospital, Ph.D. [dissertation], Manipal University, 2010.

[7] Hosmer, D. W., Lemeshow, S. (2000) Applied Logistic Regression, New York:Wiley.

[8] Guangyong, Z. (2004) A Modified Poisson Regression Approach to Prospective Studies with Binary Data, Am J Epidemiol 159, 702–706.