Journal of Statistics Applications & Probability Letters
An International Journal

http://dx.doi.org/10.18576/jsapl/040103

Estimation of Number of Involved Lymph Nodes in Breast Cancer Patients using Bayesian Regression Approach

Prafulla Kumar Swain[1,*], *Gurprit Grover*[2], *Sangeeta Chakravorty*[3], *Komal Goel*[2] *and Vikas Singh*[4]

[1] Department of Statistics, Utkal University, Vanivihar, Bhubaneswar-751004, India
[2] Department of Statistics, Faculty of Mathematical Sciences, University of Delhi, Delhi-110007, India
[3] Institute of Economic Growth, University of Delhi, Delhi-110007, India
[4] Department of General Surgery, Institute of Postgraduate Medical Education & Research, Kolkata, India

Received: 28 Feb. 2016, Revised: 23 Nov. 2016, Accepted: 3 Dec. 2016
Published online: 1 Jan. 2017

Abstract: Prediction of number of involved lymph nodes in breast cancer patients is an important criterion to assess the severity and progression of disease. The number of involved nodes is count data which often displays over-dispersion, hence the Poisson and Negative Binomial distribution is ultimate choice for modeling. In this paper we have made an attempt to estimate the number of involved lymph nodes in breast cancer patients using Bayesian regression approach assuming multivariate normal prior for the parameters. The posterior estimates have been derived using MCMC pack and the best model has been selected based on Deviance Information Criterion (DIC) values. The Bayesian Negative Binomial regression over performed than the Poisson regression. The predictors' viz., tumor size, tumor grade, CA 15-3 marker and progesterone receptor status are significantly associated with the involved lymph nodes of the breast cancer patients.

Keywords: Lymph nodes, Bayesian Poisson regression, Negative Binomial, Over-dispersion

1 Introduction

Breast cancer is the most commonly diagnosed malignancy among women and has become a big threat to human beings globally. It has been described as an alarmingly health problem in India [1]. According to Indian Council of Medical Research (ICMR) report on the metropolitan cities *viz*. Delhi, Mumbai, Bangalore and Chennai; from 1982 to 2005; has shown that the incidences of breast cancer has doubled. Over the years, the incidence of breast cancer in India has steadily increased and as many as 100,000 new patients are being detected every year [2, 3]. As per Indian population census data, the rate of mortality due to cancer in India is high and alarming with about 806000 existing cases by the end of the last century [4]. The rising graph of breast cancer both in developed and developing countries is a great challenge for biomedical researchers, especially in India it is the first common cancer of urban women and second of rural women.

The risk factors which are associated to cancer increased manifold in the last century, It includes air pollution, smoking, diet changes, insufficient physical activity, obesity, stress and so on. It is possible to control 40 percent death caused by cancer if the risk factors are recognized and managed properly [5, 6]. Besides these other prognostic factors that are considered to be independent variables include lymph node status, tumor size, tumor grade, estrogen/progesterone receptor (ER/PR) status.

The most significant prognostic indicator for patients with early stage breast cancer is the presence or absence of auxiliary lymph node involvement. Furthermore, there is a direct relationship between the number of involved auxiliary nodes and the risk of distant recurrence [7, 8]. The accurate prediction of lymph nodes in breast cancer patients helps in grading severity of disease, according to which extensive auxiliary surgery dissections can be avoided [9, 10]. Although it is an important prognostic factor but it is not necessarily associated with stages of cancer, as the patient with same number of lymph nodes may be in different stages and the patients with more number of lymph nodes are not necessarily

* Corresponding author e-mail: prafulla86@gmail.com

in more advanced stage [11]. Many authors have tried to investigate the status of lymph nodes (present or absent) in breast cancer patients [12, 21] and also to determine the prognostic value of the number of negative lymph nodes with respect to disease free survival of breast cancer patients after mastectomy [13]. The number of involved lymph nodes which is considered to be a discrete variable, are highly variable within the population, hence Poisson Regression model [14] is found to be the most appropriate form of analysis. Dwivedi et. al. (2010) [15] explored a number of statistical models (viz., Poisson, Negative Binomial, Zero hurdle and Zero inflated Negative Binomial) to test model abilities to predict the number of involved nodes. Kendal (2007) [16] described how well Negative Binomial distribution takes care of number of involved nodes in cancer patients. Yeh et.al. (2013) [17] applied a Bayesian model for censored positive count data in evaluating breast cancer progression.

In this paper we have made an attempt to estimate the number of lymph nodes using Bayesian regression approach assuming multivariate normal prior for the parameters. Bayesian approach is widely applied for fitting several models such as zero inflated generalized Poisson model [18], zero inflated regression model [19].

Bayesian methods make it easier to estimate and analyze complicated problems, while using standard classical inference methods are quite cumbersome. Also, the Bayesian approach allows us to include any prior information that we have on the parameters in the model and hence obtain a much refined set of posterior estimates. In this work, we analyze the standard Poisson and Negative Binomial regression model in a Bayesian setting, by adding a multivariate normal prior on the regression coefficients. The reason for choosing normal prior is that the likelihood of Poisson distribution and Normal distribution belongs to exponential family, and when a family of conjugate priors exists, choosing a prior from that family simplifies calculations of the posterior distribution.

The article is organized as follows; in section 2, we have discussed the material and methods used. In section 3, results have been given, and finally the discussion and conclusion has been presented in section 4.

2 Material and Methods

Data source

The study population includes all female primary breast cancer patients treated at breast clinic. (Dept of Gen. Surgery, IPGMER, SSKM Hospital, Kolkata) from Jan 2009 to Dec 2010, and had their pre-op serum CA15-3 measured and it was reported on 7, 30 post-op day and every 6 months for 2 years. Patients were excluded if any other malignancy was known from their previous history or if staging investigations at the time of diagnosis revealed evidence of instant metastasis. A total of 85 patients fulfilled the criteria for this analysis. Patients were treated with either modified radical mastectomy (MRM) or quandrantectomy and auxiliary lymph node dissection with local radiotherapy (RT). After completion of surgery, RT and appropriate adjuvant chemotherapy or hormone therapy was not altered according to marker levels but was administered as indicated based on international guidelines.

Methods

Poisson regression: Poisson regression analysis derives its name from the Poisson distribution which is a mathematical distribution often used to describe the probability of occurrence of count data. Let Y_i denotes the number of nodes for the i^{th} breast cancer patient. Since these data are in terms of counts, therefore, we assume that Y_i follows a Poisson distribution with mean λ_i (mean number of involved nodes). Hence, the probability of observing any specific count Y_i is given by the following formula:

$$P(Y_i = y_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \qquad y_i = 0,1,2,..., \qquad \lambda_i > 0$$

We postulate that the mean value λ_i depends on a set of predictors $x_1, x_2, x_3, ..., x_p$ such that

$$log(\lambda_i) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$$

Or,

$$\lambda_i = exp(\beta_0 + \beta_1 x_1 + ... + \beta_p x_p)$$

$$\lambda_i = e^{X'\beta}$$

Bayesian Poisson regression

We have, $f(y_i/\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \qquad y_i = 0,1,2,...$

Let $\lambda_i = exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)$, be the linear combination of covariates, where $x_{ij}\,(i=1,2,...,n;\, j=1,2,...,p)$ are the covariates and $\beta_j^{'}s$ are the regression coefficients, then

$$f(y_i) = \frac{exp\left[-exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)\right]\left[exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)\right]^{y_i}}{y_i!}$$

$$= \frac{exp\left[-exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) + \sum_i y_i exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right)\right]^{i}}{y_i!}$$

$$l_y(\beta) = \prod_{i=1}^{n} f(y_i)$$

$$= \frac{exp\left[-\sum_{i=1}^{n} exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) + \sum_j \beta_j \sum_{j=1}^{p} x_{ij}\beta_j\right]}{\prod_{i=1}^{n} y_i!}$$

Let us assume the prior distribution for β as

$$\beta_j \sim N(a_j, b_j) \quad for \;\; j = 1,2,...,p$$

Then the joint density of $\beta^{'}s$ can be written as:

$$p(\beta_1,\beta_2,\beta_3,...,\beta_p) = \prod_{j=1}^{p} \frac{1}{(2\pi b_j)^{\frac{1}{2}}} exp\left[-\frac{(\beta_j - a_j)^2}{2b_j}\right] \qquad ;-\infty < a_j < \infty, b_j > 0$$

Therefore the posterior distribution for $\beta^{'}s$ can be obtained as:

$$P(\beta_j/Y_i) = L_y \cdot p(\beta)$$

$$= \frac{exp\left[-\sum_{i=1}^{n} exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) + \sum_j \beta_j \sum_{j=1}^{p} x_{ij}\beta_j\right] \cdot \prod_{j=1}^{p} \frac{1}{(2\pi b_j)^{\frac{1}{2}}} exp\left[-\frac{(\beta_j - a_j)^2}{2b_j}\right]}{\prod_{i=1}^{n} y_i!}$$

$$= \frac{exp\left[-\sum_{i=1}^{n} exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) + \sum_j \beta_j \sum_{j=1}^{p} x_{ij}\beta_j - \sum_{j=1}^{p} \frac{(\beta_j - a_j)^2}{2b_j}\right]}{\prod_{i=1}^{n} y_i! \prod_{j=1}^{p} \sqrt{2\pi b_j}}$$

$$= \frac{exp\left[-\sum_{i=1}^{n} exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) - \frac{1}{2}\sum_j \frac{\beta_j^2}{b_j} + \sum_j \beta_j \left[\sum x_{ij} y_j + \frac{a_j}{b_j}\right] - \sum \frac{a_j^2}{2b_j}\right]}{\prod_{i=1}^{n} y_i! \prod_{j=1}^{p} \sqrt{2\pi b_j}}$$

putting $d_j = \sum x_{ij} y_j + \frac{a_j}{b_j} \qquad for \; j = 1,2,...,p$
We will get,

$$P(\beta_j/Y_i) \propto exp\left[-\sum_{i=1}^{n} exp\left(\sum_{j=1}^{p} x_{ij}\beta_j\right) - \frac{1}{2}\sum_j \frac{\beta_j^2}{b_j} + \sum_j \beta_j d_j\right] \tag{1}$$

which on simplification will yield a normal distribution with mean d_j variance $2b_j$.

Bayesian Negative Binomial Regression

The Poisson regression model does not suit well to over dispersed data (i.e. variance is greater than mean), in that case the Negative Binomial model is the best alternate choice for analysis [14]. For introducing Negative Binomial regression model to breast cancer patient data Bayesian setting,

Let Y_i be the number of nodes for the i^{th} breast cancer patient, following Negative Binomial distribution with parameter r (number of negative nodes preceding the first positive node) and p (probability of having positive node). Then, the probability of observing any specific count Y_i is given by:

$$P(Y=y_i)=\frac{\Gamma(y_i+r)}{y_i!\Gamma r}p^r(1-p)^{y_i} \qquad y_i=0,1,2,3,...;\ \ 0\le p\le 1$$

with $E(y_i)=\frac{r(1-p)}{p}$

and $V(y_i)=\frac{r(1-p)}{p^2}=\mu_i^2+\frac{\mu_i^2}{r}$

$$P(y_i)=\frac{\Gamma(y_i+r)}{y_i!\Gamma r}\left[\frac{r}{(\mu_i+r)}\right]^r\left[\frac{\mu_i}{\mu_i+r}\right]^{y_i}$$

Let $\mu_i=exp\left(\sum_{j=1}^{p}x_{ij}\beta_j\right)$, the mean number of involved nodes be the linear combination of covariates, where $x_{ij}\,(i=1,2,3,...,n;\,j=1,2,3,...,p)$ are the covariates and $\beta_j^{'}s$ are the regression coefficients.

Then the likelihood of Y_i can be written as:

$$\begin{aligned}L(Y)&=\prod_{i=1}^{n}P(y_i)\\&=\left(\prod_{i=1}^{n}\frac{\Gamma(y_i+r)}{y_i!\Gamma r}\right)\left[\frac{r^n}{\prod_{i=1}^{n}(\mu_i+r)}\right]\prod_{i=1}^{n}\left[\frac{\mu_i}{\mu_i+r}\right]^{y_i}\end{aligned}$$

$$\begin{aligned}logL(Y)&=C+nlogr-\sum_i log(\mu_i+r)-y_i\sum_i[log\mu_i-log(\mu_i+r)]\\&=C_1-(1+y_i)\sum_i log(\mu_i+r)-y_i\sum_i log\mu_i\\&=C_1-(1+y_i)\sum_i log\left(e^{x_{ij}\beta_j+r}\right)-y_i\sum_i\sum_j x_{ij}\beta_j\end{aligned}$$

Then the posterior for $\beta^{'}s$ assuming the same normal prior can be obtained as:

$$\begin{aligned}P(\beta_j/Y_i)&=\left(C_1-(1+y_i)\sum_i log(\mu_i+r)-y_i\sum_i log\mu_i\right)\prod_{j=1}^{p}\frac{1}{(2\pi b_j)^{\frac{1}{2}}}exp\left[-\frac{(\beta_j-a_j)^2}{2b_j}\right]\\&=C_1-(1+y_i)\sum_i log\left(e^{x_{ij}\beta_j+r}\right)-y_i\sum_j x_{ij}\beta_j-\frac{1}{2b_j}\sum_j(\beta_j-a_j)^2\end{aligned} \tag{2}$$

Considering multivariate normal prior for regression coefficients $\beta^{'}s$, we have obtained the posterior summaries of regression coefficients under both Poisson and Negative Binomial distributions separately. The posterior summaries have been obtained using MCMCpack [25] and INLA package in R [24].

3 Model Comparison

Deviance Information Criterion(DIC

The deviance information criterion (DIC) [20] is a model assessment tool, which is a Bayesian alternative to Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). DIC is a Bayesian measure that takes account of

both the goodness of fit and the complexity of a fitted model. The DIC is defined as follows;

$$DIC = \hat{D} + 2\left(\bar{D} - \hat{D}\right)$$

Where $\bar{D}$ is the average of deviance $(-2\ln l)$ over the posterior distribution, and $\hat{D}$ is the deviance calculated at the posterior mean parameters. $P_D = \left(\bar{D} - \hat{D}\right)$ effective number of parameters as a penalty term on the goodness of fit. The smaller DIC value will be treated as better model fit. The p-values less than 0.05 were considered as significant results. All the statistical analysis have been performed in R (version 3.2.0).

4 Results

The study population includes 85 breast cancer patients, who were diagnosed from Jan, 2009 to Dec, 2010. The mean age of patients at diagnosis is 50.09 years (SD=12.82), ranging from 25 to 85 years. The descriptive characteristics of various prognostic factors are shown in Table 1 Out of 85 patients in the study, the number of involved nodes is found in 35 (41.2%) patients. The mean and standard deviation of number of involved nodes per patient are 4.4 and 4.7 respectively. The Table 2 shows the summary statistics of posterior estimates obtained from Poisson distribution and Negative Binomial distribution. The larger tumor size of the range (2-5cm) is significantly associated with increased risk of higher number of lymph nodes. Also it reveals that tumor size, tumor grade (III), and CA15-3 (preoperative value) is consistently significant across both the models. PR status is found to be statistically significant in Poisson regression model. Whereas the other predictors viz., age, ER status and HN2 status are not significant in the both models. Since the DIC value for Negative Binomial regression model (445.62) is smaller than the Poisson regression model (613.04) implying that the Negative Binomial distribution can better explain the distribution of number of involved lymph nodes. Figure 1 shows the trace plots for convergence diagnostics and marginal posterior kernel density plots. The trace plot indicates that the Markov chain has stabilized with good mixing and hence MCMC algorithm converged, and the kernel density plot estimates the posterior marginal distribution.

Table 1: Descriptive Characteristics of Breast Cancer Patients (N=85)

Factors	Categories(Code)	Frequency	Percentage
CA 15-3 U/ml	<25	10	11.8
	≥25	75	87.2
Tumor Size (cm)	<2 (0)	24	27.9
	2-5 (1)	48	55.8
	≥5 (2)	13	15.1
Lymph nodes	0-3 (0)	50	58.8
	4-9 (1)	19	22.4
	≥9 (2)	16	18.8
Tumor Grade	I (1)	23	27.1
	II (2)	42	49.4
	III (3)	20	23.5
ER Status	Negative (0)	40	47.1
	Positive (1)	45	52.9
PR Status	Negative (0)	48	56.5
	Positive (1)	37	43.5
HN2 Status	Negative (0)	53	62.4
	Positive (1)	32	37.6

Table 2: Posterior Estimates obtained by using Poisson Regression and Negative Binomial Regression Model

Parameters	Poisson regression model			Negative Binomial regression model		
	Mean	SD	95 % HPD	Mean	SD	95 % HPD
(Intercept)	0.216	0.309	(-0.403, 0.811)	0.186	0.851	(-1.841, 1.510)
Age	0.007	0.004	(-0.002 , 0.016)	0.01	0.013	(-0.014, 0.036)
Tumor Size(cm)						
2-5	0.756	0.163	(0.444 , 1.083)	0.938	0.426	(0.097, 1.775)
$\geq$5	1.181	0.19	(0.812, 1.558)	1.303	0.525	(0.293, 2.359)
Tumor Grade						
II	-0.052	0.15	(-0.345, 0.244)	-0.049	0.415	(-0.871, 0.761)
III	0.165	0.154	(0.014, 0.469)	0.21	0.448	(0.071, 1.092)
ER Status	0.036	0.136	(-0.23, 0.304)	0.149	0.416	(-0.677, 0.959)
PR Status	0.137	0.125	(0.010, 0.383)	0.108	0.399	(-0.892, 0.678)
HN2 Status	-0.057	0.116	(-0.286, 0.169)	-0.088	0.35	(-0.593, 0.786)
CA15 (Pre-op)	0.124	0.201	(-0.258, 0.533)	0.322	0.518	(0.021, 1.310)
DIC	**613.04**			**445.62**		

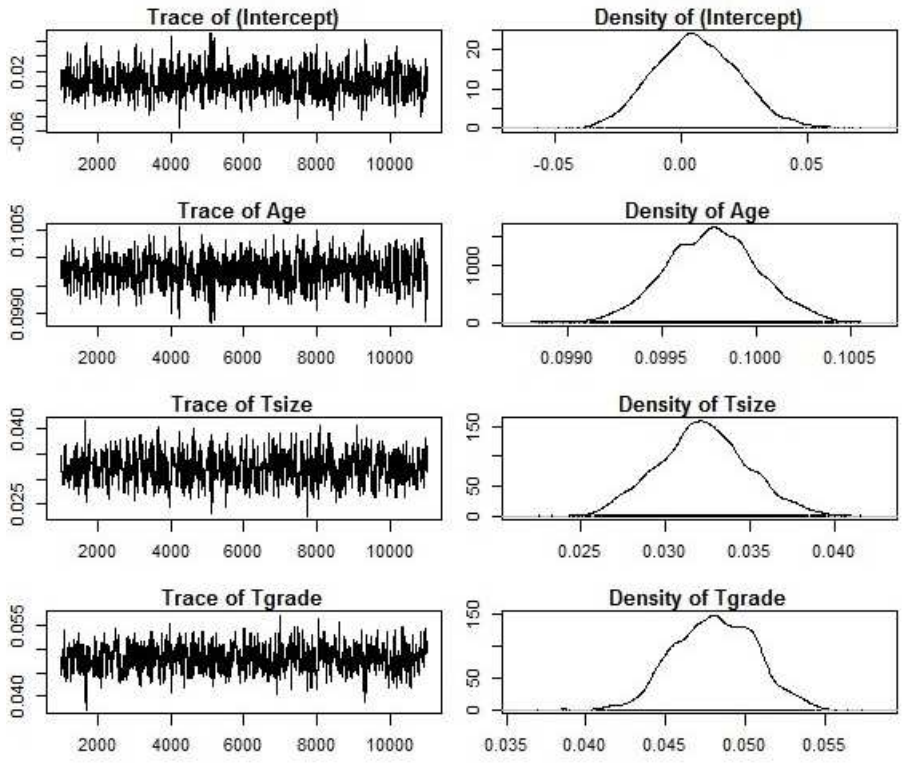

Fig. 1: Trace plots of intercept for convergence diagnostics and marginal posterior kernel density plots.

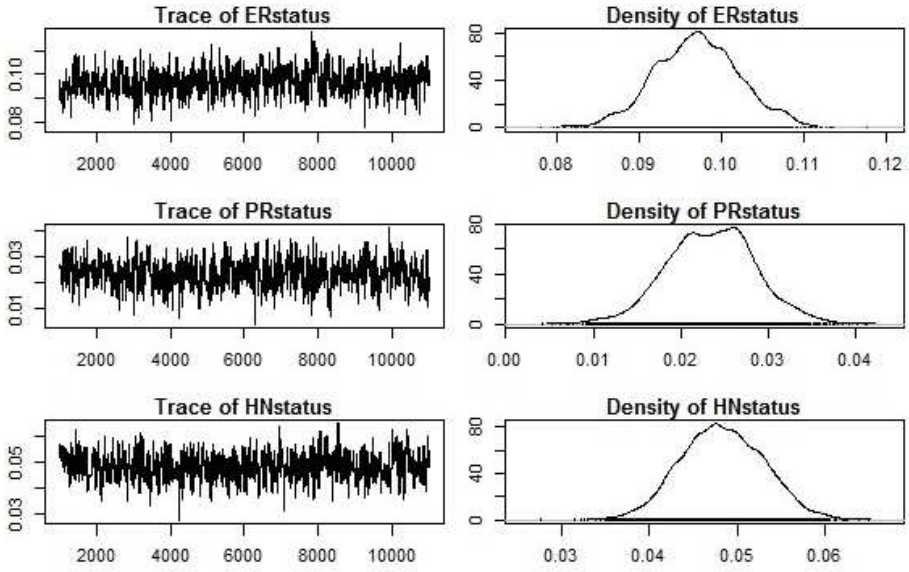

Fig. 2: Trace plots of ERstatus for convergence diagnostics and marginal posterior kernel density plots.

5 Discussion

In this paper we have tried to determine the predictors associated with the involved lymph nodes in breast cancer patients. The number of involved nodes is most important therapeutic and prognostic factor for breast cancer [9]. In fact, it plays a very important role in assessing the severity and progression of disease stage. Generally, the clinicians need to predict the number of involved nodes in breast cancer patients in order to improve health outcomes. Many studies have been carried out to predict the nodal status (presence or absence) in breast cancer patients, but a few authors have highlighted the prediction of the number of involved nodes using statistical models for count data. Number of involved nodes which is a discrete variable exhibits count data, hence Poisson and Negative Binomial regression models can be the best choice for modeling these types of data. Guern and Vin Hung (2008) [23] found that Negative Binomial model better predicts the involved number of nodes than Poisson regression model. Another study by Rodriguez et al.(2009) [26] shows that Negative Binomial provides a better fit to the total number of involved nodes as compared to Poisson process in meta analysis. All these studies fit various statistical models over count data and compare them but to the best of our knowledge none of them explored these models under Bayesian setting. In this paper we have demonstrated the applications of Bayesian regression approach under Poisson regression model and Negative Binomial regression models assuming multivariate normal prior. The rational behind Bayesian approach is that, it incorporate the prior information on the parameters in the model and hence obtain a much refined set of posterior estimates. Earlier results states that Negative Binomial model describes better the number of nodal involvement than the Poisson model due to excess variability (over dispersion) [16, 23]. Our findings also support that, in a Bayesian analysis Negative Binomial regression model performs better than the Poisson regression model. The predictors viz., tumor grade (III), tumor size, PR status and CA 15-3 are found to be statistically significant for involved lymph nodes across both the models. Also we find that the larger the tumor size the increased is the risk of involved number of nodes.
We acknowledge some limitations of our study that may be considered. Firstly, granted the additional knowledge of predictor?s namely menstrual status, parity, types of surgery etc. one could be able to provide better prediction of involved number of nodes. Secondly, the study is a single centre study and may not represent the majority of the populations considering the regional diversity of India.

6 Conclusion

The Bayesian Negative Binomial regression is a viable approach to describe the nodal distribution than the Bayesian Poisson regression. The predictors? viz., tumor size, tumor grade, CA 15-3 marker and progesterone receptor status are significantly associated with the involved lymph nodes of the breast cancer patients. Focusing in these predictors will help the medical practitioner to start the early diagnosis of breast cancer patients.

Acknowledgement

We are grateful to the anonymous reviewers for their insightful comments and constructive suggestions that improved the article.

References

[1] B.B. Yeole and A.P. Kurkure, An epidemiological assessment of increasing incidence and trends in breast cancer in Mumbai and other sites in India, during the last two decades, Asian Pacific Journal of Cancer Prevention, **4(1)**, 51-56 (2003).
[2] C.H.Yip, N.A. Taib and I. Mohamed, Epidemiology of Breast cancer in Malaysia, Asian Pac. J. Cancer Prev. **7**, 369-374 (2006).
[3] I. J. Michael and A. Jernal, Cancer epidemiology, prevention and screening, Cancer medicine, Hollan. Frei. American Cancer Society, Philadelphia: BC Decker Inc, 367-81 (2003).
[4] I. Ali, Rahis-ud-din, K.Saleem, H.Y. Aboul-Enein and M.A. Rather, Social Aspects of Cancer Genesis, Cancer Therapy, **8**, 6-14 (2011).
[5] M.Akbari , K.Abachizadeh , M.Khayamzadeh, et. al, Iran cancer report, Cancer J. Clin., **55**, 74-108 (2008).
[6] A. Khoshkar, T. Koshki and B. Mahaki, Investigating the Incidence of Prostate Cancer in Iran 2005-2008 using Bayesian Spatial Ecological Regression Models, Asian Pacific journal of cancer prevention: APJCP, **16(14)**, 5917 (2015).
[7] T. Nemoto, N. Natarajan, R. Bedwani et. al., Breast cancer in the medial half; results of the 1978 national survey of the American College of Surgeons, Cancer ,**51**, 1333-1338 (1983).
[8] R.A. Saez, W.L. McGuire, G.M. Clark, Prognostic factors in breast cancer, Semin Surg Oncol , **5**, 102-110 (1989).
[9] C.A. Hernandez-Avila, C. Song,L. Kuo, H. Tennen, S.Armeli, H.R.Kranzler, Targeted versus daily naltrexone: Secondary analysis of effects on average daily drinking Alcoholism, Clinical and Experimental Research; **30(5)**, 860?865 (2006).
[10] D.J. Slymen, G.X. Ayala, E.M. Arredondo, J.P. Elder, A demonstration of modeling count data with an application to physical activity. Epidemiologic Perspectives & Innovations, **3(3)**, 1? [PubMed: 16390556] (2006).
[11] Y. Akifumi , M. Toshiki and I. Makio, Bayesian Analysis of Lymphatic Spreading Patterns In Cancer Of The Thoracic Esophagus*, Ann. Inst. Statist. Math. **45(3)**, 401-418 (1993).
[12] N.J. Horton, E.Kim, R. Saitz, A cautionary note regarding count models of alcohol consumption in randomized controlled trials, BioMed Central Medical Research Methodology, **7(9)**, 1?9 (2007).
[13] W.San, J. Sun.et. al., Number of negative lymph nodes is associated with disease-free survival in patients with breast cancer, BMC Cancer, 1-7 (2015) .
[14] A.C.Cameron, P.K. Trivedi, Econometric Society Monograph. New York: Cambridge University Press; Regression Analysis of Count Data (1998).
[15] A. Dwivedi., S.,Dwivedi, S. Deo, R. Shukla, E. Kopras, Statistical models for predicting number of involved nodes in breast cancer patients.,Health (Irvine Calif)..July,**2(7)**, 641?651 (2010).
[16] W. Kendal, The number distribution for lymph nodes in cancer, Mathematical Biosciences **205**, 32?43 (2007).
[17] W. H. Yeh, Y.Jiang, ,L. Garrard, Y. Lei & B.Gajewski, A Bayesian model for censored positive count data in evaluating breast cancer progression, Model Assisted Statistics and Applications, **8(2)**, 143-150 (2013).
[18] J. F.Angers and A. Biswas, A Bayesian analysis of zero-inflated generalized Poisson model, Computational statistics & data analysis, **42(1)**, 37-46 (2003).
[19] S. K. Ghosh, P. Mukhopadhyay and J. C. J. Lu, Bayesian analysis of zero-inflated regression models, Journal of Statistical planning and Inference, **136(4)**, 1360-1375 (2006).
[20] D.J. Spiegel halter, N.G. Best, B.P. Carlin, A. vanderLinde, Bayesian measures of model complexity and fit, Journal ofthe Royal Statistical Society Series B, **64**, 583?639 (2002).
[21] M.Cianfrocca and L.Goldstein, Prognostic and Predictive Factors in Early-Stage Breast Cancer, The Oncologist; **9**, 606-616 (2004).
[22] C. Yiangou, S .Shousha and H.D. Sinnett, Primary tumour characteristics and axillary lymph node status in breast cancer, British Journal of Cancer **80(12)**, 1974?1978 (1999).
[23] A.S. Guern, V.Vinh-Hung, Statistical distribution of involved axillary lymph nodes in breast cancer, Bull Cancers; **95(4)**, 449?455 (2008).
[24] S. Martino and H. Rue, Implementing approximate Bayesian inference using Integrated Nested Laplace Approximation: A manual for the inla program, Department of Mathematical Sciences, NTNU, Norway (2009).
[25] A. D. Martin and K. M. Quinn, MCMCpack: Markov chain Monte Carlo (MCMC) Package, 2005. URL http://mcmcpack.wustl.edu. R package version 0.6-3.
[26] A. Rodriguez,B. Manrique-Espinoza,S.G.Sosa-Rubi, Statistical analysis for count data: Use of health services applications, Salud Publica Mex. , **51(5)**, 397?406 (2009).

Prafulla Kumar Swain, is a assistant professor in the P.G. Department of Statistics, Utkal University, Bhubaneswar, India. His area of interest is Biostatistics and Survival analysis. He has more than 15 research publications in the field of HIV/AIDS and cancer survival modeling.

Gurprit Grover, is a associate professor at Department of Statistics, Faculty of Mathematical Sciences, University of Delhi, India. Her area of interest is Biostatistics. She has more than 32 years of teaching experience and has also produced more than 15 Ph.D. during this period. She is also working as a reviewer for several reputed journals.

Sangeeta Chakravorty, is working as a associate professor at Institute of Economic Growth, Delhi, India. Her specialization is Econometrics and Bayesian Inferences.

Komal Goel is a Ph.D. scholar in Department of Statistics, Delhi University, India. Her area of interest is bio-statistics.

Vikas Singh, MBBS (Gold Medal), MS (Gold Medal), DNB (Surgical Oncology). Currently working as a Senior Resident in the Department of surgical oncology, Max Superspeciality Hospital, New Delhi, India.