Examining Biliary Acid Constituents among Gall Bladder Patients: A Bayes Study Using the Generalized Linear Model

Puja Makkar[1,*], S.K. Upadhyay[1,2], V.K. Shukla[3] and R.S. Singh[4]

[1]Department of Statistics; [2]DST Centre for Interdisciplinary Mathematical Sciences; [3]Department of General Surgery, Institute of Medical Sciences, Banaras Hindu University, Varanasi, India; [4]Department of Statistics, University of Guelph, Canada

Abstract: The generalized linear model is an important class of models that has wide variety of applications mainly because of its inherent flexibility and generality. The present paper provides an important application of GLM in order to examine different constituents of bile acid in the development of gallstones as well as carcinoma among the gallbladder patients. These constituents may be broadly categorized as primary and secondary bile acids. The paper, in fact, considers two particular cases of GLM based on normal and gamma modelling assumptions and provides the complete Bayes analysis using independent but vague priors for the concerned model parameters. It then analyzes a real data set taken from SS Hospital, Banaras Hindu University, with primary (secondary) bile acids as response variables and secondary (primary) bile acids as the predictors. The authenticity of the assumed models for the given data set is also examined based on predictive simulation ideas.

Keywords: Generalized linear model, vague priors, posterior distribution, biliary acids, gallbladder diseases, predictive simulation, Bayes information criterion.

1. INTRODUCTION

We may often come across a kind of epidemiological study where the interest centres on studying the changing pattern of the results of a few tests taken from different groups of patients available in the form of case-control scenarios so that one can easily try to find an association, if any, between the test constituents and the disease. Normally, the medical practitioners are interested to know if the particular test results are primarily responsible for the development of a disease but such answers are difficult, in general, as the data may not often be available in a way we require to answer these issues.

The medical practitioners undertaking such studies suggest the patients to undergo for a number of tests so that they may establish a link between the outcome of various tests and the disease. These tests may be based on physical examinations of the patients or may also be based on a number of pathological or other investigations often suggested by the practitioners depending on his/her expertise. Let us consider, for an instance, a group of patients suffering from non-small cell lung cancer (NSCLC) or leukemia. In order to assess what actually caused the disease a number of factors such as the age at diagnosis, sex of the patient, marital status, occupation, education, smoking habits, blood pressure, status of any other disease other than the leukemia or NSCLC, etc. may be recorded. Besides, the patients may also be examined for blood, urine, etc. and sometimes even for histological types of cells like Adenocarcinoma, Squamous cell carcinoma, large cell carcinoma, Adenosquamous carcinoma, etc.

The linking between the outcome of various tests and the disease may be important in several ways. First, it may help in ascertaining the actual cause of the disease and, second, it may help in diagnostics as well. For researchers in the medical field, such a linking may be equally important as it helps in understanding how the test results change from a normal group of persons to a diseased group. It is, therefore, important for the practitioners to know the results of various test constituents so that they may attempt to know the relationship, if any, among the test constituents and the disease or among the test constituents from a normal group to a diseased group.

In statistical terms, any such linking can be regarded as the problem of regression analysis. Variables in the study may include demographic factors, physiological factors, histological factors, etc. to name a few. The regression analysis actually defines a relationship between the main (response) variable of interest and several other concomitant variables so that the quantitative effect of the causal variables upon the variable that they influence can be ascertained. Based on the types of relationships several regression models have been defined in the literature. These may include linear, non-linear, generalized linear, and generalized non-linear models, etc. A detailed discussion of these

*Address correspondence to this author at the Department of Statistics, Banaras Hindu University, Varanasi, India; Tel: (91-542) 6702892; Fax: (91-542) 2368174; E-mail: puja.stats@gmail.com

E-ISSN: 1929-6029/15

relationships is beyond the scope of the present paper; however, the interested readers may refer to Draper and Smith (1998), Congdon (2003), Gelmam *et al*. (2006), Vittinghoff *et al*. (2005), [1-4] etc. The present paper mainly concerns the generalized linear model (GLM) and, therefore, our discussion will henceforth focus on GLM only although we shall finally switch over to two particular cases of GLM based on normal and gamma modelling assumptions for the response variables. The primary objective of the paper is to focus on three groups of patients suffering from gallbladder disease including carcinoma who were examined for four different constituents of bile acid so that the effect of different constituents on the disease can be studied.

The incidence of the carcinoma of biliary tract varies greatly in different parts of the world. The high incidence areas are Israel (7.5 per 100, 000 population for males and 13.8 per 100, 000 population for females) and American Indians (5.1 for males and 8.7 for females; the incidence in rest part of the USA is 1.0 per 100, 000). In the Varanasi region of India, it accounts for 4.44% of all the malignancies. The efficacy of the disease may be based on the different concentration levels of basic constituents of biliary acid which are believed to be primarily responsible for the development of gallbladder stones and then to cancer (see, for example, Shukla *et al*. (1993)) [5] although a statistical authenticity of such a conclusion is difficult due to the limitation of the study. These constituents are termed as colic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA) and lithocholic acid (LCA). Since it has been argued that gallstones are the major risk factor for biliary neoplasm (see Fraumeni (1975)) [6], it is not surprising that the gallbladder is the most common site of the biliary tract cancer (see Shukla *et al.* (1985)) [7]. Therefore, attempts have always been made to examine the question of carcinogenesis by comparing the distribution of primary (CDCA and CA) and secondary (DCA and LCA) bile acids in the patients with carcinoma of the gall bladder, cholelithiasis, and in a control group. Shukla *et al*. (1993) [5] is an important reference where a successful attempt has been made in this direction based on a realistic data obtained from SS hospital, Banaras Hindu University. Their treatment was, however, classical and mostly based on significance testing. A careful review of their work provides an impression that it has enough scope to offer for further statistical exploration.

The present paper analyzes the work of Shukla *et al*. (1993) [5] in a Bayesian framework under the assumption of normal and gamma based regression models, both of which can be considered to be particular cases of GLM. It is well known that GLM is an important class of statistical models that allows us to sort out many complications that cannot be handled within the framework of familiar linear models. The fitting of such models has been the subject of a great deal of research over the past decade. In fact, the model offers a unifying class which is widely used in the regression analysis incorporating variety of application areas. Although initially introduced in a classical framework (see, for example, Nelder and Wedderburn (1972), McCullagh and Nelder (1989), etc.) [8, 9], the past decade has witnessed rapid growth employing these models in the Bayesian context as well. This is not only due in part to their attractiveness with the familiar hierarchical structuring but also due in part to the wide availability of high speed computing resources to implement simulation based fitting of these models.

The GLM can be considered as a generalization of the general linear model. In its simplest form, a linear model specifies the (linear) relationship between a response variable Y and a set of predictor variables, say X's. However, there are many relationships that cannot be adequately summarized by simple linear equation mainly because of the two reasons. First, the response variable of interest may have a non-continuous distribution, and thus, the predicted values should also follow the respective distribution; any other predicted values are not logically possible. The second reason might be based on the fact that the effect of the predictors on the response variable may not be linear in nature.

In statistics, there are two types to model one which has simplicity and other that has completeness, simple models are not very cumbersome to understand and computationally more tractable but often show odds with the data, whereas complicated models are better fit to the data and capture more realistic scenarios, but they can be computationally awkward or intractable. When they are too complicated, they are hard to replicate. in a realistic circumstances, the classical model like as the general linear model have had to be computationally manageable. Initially Many things that were impossible before—iterative algorithms such as Monte Carlo methods, repeated tests, the whole range of Bayesian approaches—now can be routine (or nearly so). Nevertheless, this model is widely acceptable in classical as well as Bayesian paradigm. We do not intend to provide further details on GLM due to space restriction although the interested readers may refer to McCullagh and Nelder (1989) [9].

The inferences to the GLMs are available in bulk. The evolution of these models as well as the details regarding model fitting, model checking and inferences in a classical framework are thoroughly documented by McCullagh and Nelder (1989) [9]. The other important classical references that deal with the estimation and testing for the concerned model parameters include Nelder and Wedderburn (1972), Breslow (1984), Lawless (1987a, b), Smyth (1989), Fahrmeir and Tutz (1991), Lindsey (1997), etc. (see also Dey and Ravishanker (2000)) [8, 10-16]. Besides, Breslow and Clayton (1993), and Vittinghoff *et al*. (2005) [4, 17] are two other significant references where the authors have explored this model by introducing random effects in addition to the fixed effects.

Bayes inferences to the model are also available in bulk. This happened only after the fact that the computational complications were well taken care of by MCMC based approaches (see, for example, Gelfand *et al.* (1999), Delaportas and Smith (1993), etc) [18, 19]. Among important Bayesian references, one can refer to Ibrahim (1990), Ibrahim and Laud (1995), Das and Dey (2006) [20-22], and more recently, Das and Dey (2007) [23], etc. The last reference is quite elaborative and besides providing a detailed accountability of various important works, it also details the Bayesian analysis entirely in a new perspective. Some other important Bayesian references worth to mention include Robert and Casella (1996), Gelfand and Sahu (1999), and Marin and Robert (2007), [24-26] etc.

The plan of the paper is as under. The next section briefly outlines the GLM and the associated discussions. The section divided in two subsections then provides two particular cases of GLM based on normal and gamma assumptions for the response variables. The corresponding Bayesian modelling formulations for the two particular cases are also given in these subsections. It is to be noted that the normal based regression model could have been described without a reference to GLM but the main advantage in switching from GLM is to maintain the uniformity of presentation. Section 3 provides a review discussion of Bayes information criterion (BIC), a criterion that has been used in the latter sections to deal with the model comparison. This section may appear as an odd combination although it makes the paper self-content. Section 4 provides a real data example taken from SS hospital, Banaras Hindu University, on different constituents of bile acid from three categories of gallbladder patients. Section 5 provides Bayes analysis of the real data example using the modelling formulations given in subsections 2.1 and 2.2. The section divided in three subsections finally examines the compatibility of the two regression models for the data in hand using predictive simulation ideas and uses the results to examine the changing pattern of different constituents of bile acid from one category of patients to another. The results of model comparison based on BIC have also been given. It is worth mentioning that the same data set was initially reported and analyzed by Shukla *et al*. (1993) [5] in a classical framework based on simple and non-validated modelling assumptions. Finally, a brief conclusion and recommendation are given at the end.

2. THE GENERALIZED LINEAR MODEL

The linear models are quite important in the regression analysis and they often arise in situations where a linear relationship is expected between a dependent (response) variable and a set of predictor variables. The assumption of linearity is done party because we are more accustomed to visualize linear relationships and partly because of the fact that such relationships are easy to analyze with generally closed form results. Such models normally embody both systematic as well as random (error) components, with the errors often assumed to have normal distributions. The associated analytic technique is based on least squares theory in a classical paradigm and often easily manageable posteriors in a Bayesian paradigm. Techniques for non-normal data include probit analysis involving binomial variate and contingency tables involving multinomial variate with systematic part of the model usually multiplicative in the latter case. In both these situations, there is a linear aspect to the relationship although sometimes it may not be visualized directly. In another extension to linear relationship, certain transformation might be desired to obtain linearity which is otherwise not obvious (see, for example, Nelder (1968)) [27]. A further class of models based on chi-square or gamma distributions arises in the estimation of variance components with systematic component of the model again having somewhat linear structure (see, for example, Nelder and Wedderburn (1972)) [8].

The GLM is a generalization of the general linear model which includes all the above examples (and a few more) as special cases and, as a matter of fact, enjoys a great deal of interest from the viewpoint of both statistical researchers and the practicing data analysts. The class of GLM was introduced by Nelder

and Wedderburn (1972) [8] who also provided a unified procedure for fitting them based on the likelihood. In fact, their procedure can be regarded as a generalization of the procedure described by Finney (1952) [28] for maximum likelihood estimation in probit analysis.

GLM considers the response variable belonging to the exponential family with covariates and it is capable of modelling various kinds of data whether they are available in the form of binomial, count, or even continuous structure. A regression model determines the structure of the covariate information with the link function specifying the relationship between the regression model and the expected values of the observation. To provide a mathematical framework for the model, let us assume that y_1, y_2,…, y_n are independent observations with y_i having the density from the natural exponential family given by

$$f(y_i \mid \theta_i) = \exp\{(\theta_i y_i - \psi(\theta_i)) + c(y_i)\}, \quad (1.1)$$

where the density in (1.1) is parameterized by the canonical parameter θ_i (i = 1, 2, ..., n). We further assume that the functions $\psi(\cdot)$, and $c(\cdot)$ are known and θ_i's are related to the regression coefficients by the link function

$$\theta_i = h(\eta_i), \quad (1.2)$$

and

$$\eta_i = x_i\beta \quad (1.3)$$

is the systematic component of the GLM. In (1.3), $x_i = (1, x_{i1}, x_{i2}, ..., x_{ip-1})$ is a 1×p vector denoting the i^{th} row of a (n , p) matrix of covariates X, $\beta = (\beta_0, \beta_1, \beta_2, ..., \beta_{p-1})'$ is a p-vector of regression coefficients, and $h(\cdot)$ is monotonic differentiable function (which we call the link function) (see also Nelder and Wedderburn (1972), and Das and Dey (2007)) [8, 23].

The model given by (1.1)-(1.3) is called GLM. As mentioned earlier, the Gaussian, logistic, binomial, Poisson, gamma, and inverse Gaussian regression models, etc. are all special cases of the GLM (see, for example, McCaullagh and Nelder (1989) [9] for further details). In fact, the application of the model to individual cases is done in accordance with the availability and the nature of the data. For example, if the available data is in the form of counts we restrict to a Poisson model; if it is in the form of categorical observations, we restrict to binomial assumption; and for the continuous scenarios gamma, Gaussian, log normal and inverse Gaussian relationships can be thought of. The choices, however, depend on the form of link functions that are defined to link the model parameters with the corresponding covariates.

The GLM can be computationally difficult in a Bayesian framework especially when the data advocate for the non-Gaussian modelling assumptions. In such cases the unusual form of link functions may add further computational intricacies (see, for example, Delaportas and Smith (1993)) [19]. Most of such situations typically require the MCMC implementation but there is no unified strategy for the same. One often requires a problem oriented, tailored, and hybrid algorithm to achieve the desired computational efficiency. Recently, Das and Dey (2007) [23] considered GLM absolutely in a new perspective which often resulted in closed form estimators for the corresponding regression parameters. Their contribution is certainly significant from the viewpoint of applied researchers although we shall not go into the details of their strategy rather stick to the model that has been used in the paper and proceed in a usual way.

Before we end the section, we shall make a brief comment on residuals or error terms. The residuals for simple linear model are well defined and straightforward to visualize. However, the situation is not that straightforward with the GLM. Here the concept of residuals can be defined and illustrated both at systematic component level and at random component level. We do not intend to go into the details of these various concepts in a general framework rather stick to our own particular cases. The interested readers are, however, referred to Das and Dey (2007) [23] (see also McCullough and Nelder (1989)) [9] for details.

2.1. Normal Regression Model as a Particular Case of GLM

As mentioned earlier, the normal regression model is, in fact, a general linear model considered extensively in the literature and most of the expressions based on it are obtainable in closed forms. The model can also be defined using a general structure of GLM by considering an identity link function. To begin with, let us assume the response variables y_1, y_2, …, y_n as n independent observations from $N(\mu_i, \sigma^2)$ and let us use the identity link as $E(y_i)= \mu_i= \eta_i = x_i\beta$ to establish a connection with the explanatory variables $x_i = (1, x_{i1}, x_{i2}, ..., x_{ip-1})$. The assumption of common variance σ^2 has been taken for simplicity only and it appears natural as well for the

intended analysis of gallbladder patients. Accordingly, our normal based regression model can be written as

$$y_i \sim N(\mu_i, \sigma^2) \tag{1.4}$$

$$E(y_i \mid \beta, x_i) = \mu_i = \eta_i = x_i.\beta\,. \tag{1.5}$$

$$Var(y_i \mid \beta, x_i) = \sigma^2 \tag{1.6}$$

To complete the modelling structure, we finally assume that the errors $\varepsilon_i = y_i - E(y_i \mid \beta, X)$ are independent normally distributed with mean zero and common variance σ^2 (i =1,…, n).

Thus, in a matrix form, we can consider $y = (y_1,\ldots y_n)'$ as a vector of response variables having mean vector Xβ and the variance-covariance matrix $\sigma^2 I$ where I is n x n identity matrix. Obviously, the components of y are assumed to be independent with common variance as mentioned earlier. The error term **ε**, which is also a vector of n components and is assumed to be homoscedastic, can be written as ε = y-E(y | β, X).

Obviously, (1.4)-(1.6) define a general linear model that has also been visualized as a particular case of GLM by considering the identity link. The likelihood function based on this modelling assumption can, therefore, be written as

$$L\left(y \mid \beta, X, \sigma^2\right) \propto \left(\frac{1}{\sqrt{\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right], \tag{1.7}$$

which, after simplification, reduces to

$$L\left(y \mid \beta, X, \sigma^2\right) \propto \left(\frac{1}{\sigma^2}\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right\}\right]. \tag{1.8}$$

To complete the Bayesian model formulation, let us assume the joint non-informative prior for β and σ^2 as

$$g(\beta, \sigma^2) \propto \frac{1}{\sigma^2},\ -\infty < \beta_i < \infty,\ i = 0,\ldots, p-1;\ \sigma^2 > 0. \tag{1.9}$$

Combining (1.8) with (1.9) *via* Bayes theorem yields the joint posterior that can be specified up to proportionality as

$$p(\beta, \sigma^2 \mid y, X) \propto \left(\frac{1}{\sigma^2}\right)^{n/2+1} \exp\left[-\frac{1}{2\sigma^2}\left\{\nu s^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})\right\}\right]. \tag{1.10}$$

From (1.10), it can be seen that the conditional posterior distribution (see Zellner (1971)) [29] of β given σ^2 is a p-dimensional multivariate normal with mean vector $\hat{\beta}$ and variance-covariance matrix $(X'X)^{-1}\sigma^2$ where

$$\hat{\beta} = (X'X)^{-1}X'y, \tag{1.11}$$

is the least squares estimate of β,

$$s^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\nu}, \tag{1.12}$$

and ν = n-p. Further, a simple algebraic manipulation can be used to establish that the marginal posterior $p(\beta \mid y, X)$ is a multivariate student's t distribution with degrees of freedom ν, location parameter $\hat{\beta}$ and the scale parameter $s^2(X'X)^{-1}$. Similarly, it can be shown that the marginal posterior distribution of σ^2 is inverted gamma with shape parameter (n-p)/2 and scale parameter $\nu s^2/2$. Thus, we have closed form expressions for the posteriors of both β and σ^2 which can easily provide the Bayes estimates of β as the corresponding least squares estimator and that of σ^2 as the corresponding posterior mode (say) of inverted gamma distribution.

2.2. Gamma Regression Model as a Particular Case of GLM

The previous subsection considered a simple form of normal regression model as a particular case of GLM although it could have been entertained directly as a linear model. As mentioned earlier in Sections 1-2, there can be several particular cases of GLM. One such important subcategory is gamma based regression model in a GLM framework that has been considered by a number of authors (see, for example, Aitkin *et al*. (1989), Green and Silverman (1994), Lindsey (1997), Gill (2001), Dobson (2001), etc.) [30-34] and that has often been advocated when the response variables are positive. To provide a mathematical framework, let us assume that the response variables $y = (y_1,\ldots, y_n)'$ are n independent observations from a gamma distribution with probability density function given by

$$f(y_i \mid \nu, \lambda_i) = \frac{1}{\Gamma(\nu)}(\lambda_i)^{\nu} y_i^{\nu-1} \exp\left[-y_i\lambda_i\right], \tag{1.13}$$

where λ_i is the scale parameter and ν determines the shape of the distribution. Thus we have assumed that the scale parameter changes from variate to variate

whereas the shape parameter of the distribution remains constant. The mean and variance of the distribution can be written as

$$E(y_i) = \mu_{gi} = \frac{\nu}{\lambda_i},$$

$$V(y_i) = \frac{\nu}{\lambda_i^2} = \frac{\mu_{gi}}{\lambda_i}.$$

A simple reparameterization in the model (1.13) results in the probability density function given as

$$f(y_i \mid \nu, \mu_{gi}) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu_{gi}}\right)^{\nu} y_i^{\nu-1} \exp\left[-y_i\left(\frac{\nu}{\mu_{gi}}\right)\right], \quad (1.14)$$

which now has mean μ_{gi} and variance $\frac{\mu_{gi}^2}{\nu}$. The coefficient of variation based on the model is constant and, therefore, we should roughly require the data to have this characteristic.

For gamma based regression model, there can be several link functions which can combine $E(y_i)$ with the corresponding covariates. One such function is logarithmic link function that can be written as

$$\log\{E(y_i \mid \beta, x_i)\} = \log(\mu_{gi}) = \eta_i = x_i.\beta \quad (1.15)$$

Obviously, the link function has been chosen such that the negative values μ_{gi} can be avoided. Moreover, it is to be noted that the gamma model originally given in (1.13) cannot be directly considered to provide the particular case of GLM but the form given in (1.14) does the thing routinely with the link function given in (1.15).

The error term (residual) can be defined similarly as it has been done earlier for normal regression case. The likelihood function based on this modelling assumption can, therefore, be written as

$$L(y \mid \nu, \beta, X) = \prod_{i=1}^{n}\left[\left(\frac{1}{\Gamma(\nu)}\right)\{\nu/\exp(x_i\beta)\}^{\nu} y_i^{\nu-1} \exp\left[-\nu\{y_i/\exp(x_i\beta)\}\right]\right] \quad (1.16)$$

To complete the Bayesian model formulation, let us consider the normal priors for the regression coefficients and uniform prior for the shape parameter (see, for example, Congdon (2007)) [35], that is,

$$g(\beta_i) \propto N(\varphi, \delta^2),\ i = 0,1,\ldots,p-1$$

$$g(\nu) \propto U(0, \omega), \quad (1.17)$$

where φ, δ, and ω are the hyperparameters. It is to be noted that if we consider large values of δ and ω, the resulting prior affect will be nearly vague. In addition, we have assumed all β_i's have same means and variances a priori although one can remove this restriction and can consider different means and variances as well.

Combining (1.16) with (1.17) *via* Bayes theorem yields the joint posterior of the parameters but unfortunately that cannot be specified in a closed form except that it can be written up to proportionality as a product of (1.16) and (1.17). The solution can, however, be obtained *via* Markov chain Monte Carlo simulation and a simple code in WINBUGS or R can do the job. We do not specify the posterior up to proportionality although the same can be routinely written.

3. BAYES INFORMATION CRITERION

The BIC also known as Schwarz criterion is a well-known tool for comparing the models. According to this criterion, a model is recommended if it minimizes the corresponding term given by

$$BIC = -2(\log(L(\hat{\theta})) + p \log(n) \quad (1.18)$$

where $L(\hat{\theta})$ denotes the maximized likelihood function corresponding to a model indexed with parameter θ, n denotes the total number of observations and p is the dimension of the concerned model. It is to be noted that the first term in the right hand side of (1.18) supports a more complex model and second term supports a simpler model having low dimensions. Thus BIC, being free from any prior information, penalizes the complexity of the model according to its dimension. It is a consistent measure in the sense that the probability of selecting the correct model tends to unity as the number of observations approaches infinity although it suffers from a disadvantage that it is a valid measure only for a well-behaved model (see, for example, Ghosh *et al*. (2006)) [36]. The quantity $\hat{\theta}$ in (1.18) can be replaced by posterior mode if the prior is vague.

4. A REAL DATA EXAMPLE

The real data example, initially reported and analyzed by Shukla *et al*. (1993) [5], is related to bile acid distributions in 10 controls, 10 gallstone patients and 10 carcinoma patients in a study carried out at SS Hospital, Banaras Hindu University. In the two study groups, the patients underwent cholecystectomy for gallstones. Age and sex matched patients in the control

group underwent laparotomy for appendicectomy, peritoneal tuberculosis or intestinal obstruction and did not have any biliary tract procedure. The patients with advanced carcinoma of the gall bladder had stones in their gall bladder as observed from their previous records. The data sets in the form of four important constituents, namely, Cholic Acid (CA), Chenodeoxycholic Acid (CDCA), Deoxycholic Acid (DCA), and Lithocholic Acid (LCA) are shown in Tables **1-3** for the three categories of patients. These constituents were recorded with utmost care to avoid any kind of contamination of bile samples with blood. The details of the procedure used to obtain these observations and other related details on bile acid constituents are given in Shukla *et al*. (1993) [5] so we skip these discussions due to space restriction.

Among several important studies conducted by Shukla *et al*. (1993) [5], they also considered gas-liquid Chromatographic study using Chemito 3800 CC model, a 5-foot column with an inside diameter of 0.025'' containing 3% OV-17. They observed that there was a change in the peaks of individual bile acids, in general, from one group to another though the change in LCA and CDCA was comparatively meager as compared to other two constituents. While there was almost no change in LCA and CDCA from control group to cholelithiasis, there was only a minor change from cholelithiasis to carcinoma. DCA and CA were showing rapid change in their peaks from one group to another. The change in the peaks of DCA (CA) was from low (high) to high (low) as we moved from control group to cholelithiasis and then to carcinoma. In general, their finding based on Chromatographic study can be narrated as follows. Both the primary acids exhibit a decreasing trend whereas both the secondary acids exhibit an increasing trend as we move from control

Table 1: Distribution of Bile Acid Constituents (mg/ml) in the Control Group

Patient Number	CA	CDCA	DCA	LCA
1	15.2	19.00	7.42	0.00
2	28.3	21.56	1.07	0.08
3	22.5	27.14	1.62	0.03
4	16.7	12.92	0.45	0.06
5	12.9	13.52	2.62	0.41
6	18.6	22.06	0.75	0.04
7	24.6	22.86	2.07	0.51
8	12.3	16.82	6.42	0.02
9	17.3	15.30	4.23	0.07
10	25.4	24.80	3.43	0.05
Total	193.8	195.98	30.08	1.27
Mean	19.38	19.59	3.00	0.12

Table 2: Distribution of Bile Acid Constituents (mg/ml) in the Cholelithiasis Group

Patient Number	CA	CDCA	DCA	LCA
1	8.50	9.92	3.37	0.22
2	6.80	14.44	2.08	0.78
3	8.70	11.40	4.44	0.29
4	10.20	11051	1.31	0.29
5	7.30	12.38	0.64	0.16
6	4.10	4.51	2.67	0.16
7	6.00	6.69	1.85	0.29
8	10.50	10.65	3.57	0.27
9	13.30	9.18	3.37	0.69
10	7.90	12.92	8.25	0.29
Total	83.30	103.60	31.55	3.44
Mean	8.33	10.36	3.15	0.34

Table 3: Distribution of Bile Acid Constituents (mg/ml) in the Carcinoma Group

Patient Number	CA	CDCA	DCA	LCA
1	2.40	8.56	12.13	4.20
2	6.20	8.67	13.24	4.04
3	4.20	13.68	15.51	1.47
4	3.00	9.81	16.36	1.11
5	2.20	11.01	11.31	1.18
6	4.50	8.47	11.16	3.63
7	6.90	7.27	7.61	3.02
8	3.20	4.75	3.33	0.59
9	7.30	9.78	10.01	2.45
10	5.50	6.58	8.74	2.67
Total	45.40	88.58	109.40	24.36
Mean	4.54	8.85	10.94	2.43

group to cholelithiasis and then to carcinoma although there was no statistical testimony to this fact.

Shukla *et al*. (1993) [5] further considered classical significance testing and they observed that the patients with gall bladder carcinoma had significantly higher concentration of secondary bile acids in comparison to the patients in other two groups. There was no significant difference in the secondary bile acids concentration among the control and the cholelithiasis groups. Among other findings, they also observed a significant correlation co-efficient between CA and LCA and between DCA and CDCA.

5. BAYES ANALYSIS OF THE REAL DATA EXAMPLE

For the purpose of numerical illustration, we consider the real data sets given in Tables **1-3** and analyze the same in a Bayesian framework using the modelling formulations given in subsections 2.1-2.2. The main objective of the analysis is to see how the different constituents of bile acid change from one category of patients to another. We, therefore, assume each of the primary variables as a response variable and the two secondary acids as the predictor variables. Similarly, we assume each of the two secondary variables as the response variable and the two primary acids as the predictor variables. That is, we consider only three constituents of biliary acid at a time. This strategy has been adopted for visualizing the changing pattern of each biliary acid constituent from one category of patients to another that has been shown in Section 5.3 based on predictive simulation technique. It is to be noted here that the preliminary investigation of the data for the search of an appropriate model may not be too fruitful as the number of observations in each category is too small to guess any authentic behaviour from the data. A few of the preliminary results, however, exhibited both symmetric and slightly skewed behaviour for various bile acid constituents which are to be taken as response variables.

The size of the data is so small, by the central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution. That is, suppose that a sample is obtained containing a large number of observations, each observation being randomly generated in a way that does not depend on the values of the other observations, and that the arithmetic average of the observed values is computed. If this procedure is performed many times, the central limit theorem says that the computed values of the average will be distributed according to the normal distribution (commonly known as a "bell curve").

It is to be noted here that the preliminary investigation of the data for the search of an appropriate model may not be too fruitful as the number of observations in each category is too small to guess any authentic behavior from the data. A few of the preliminary results, however, exhibited both symmetric and slightly skewed behavior for various bile acid constituents which are to be taken as response variables.

Also based on the predictive simulation technique predicted values are independently and identically distributed and converge to normal distribution.

Therefore the impact of the correlation in a small sample would be vanished when drawing large number of predictive samples from simulation. And, to illustrate this technique we have used two different families' normal and gamma in generalized linear model.

5.1. Analysis Based on Normal Regression Model

We first proceed to obtain the Bayes estimators of various parameters involved in the model. For the sake of clarity, we use a subscript N with β to specify the estimates corresponding to normal regression model. Since the Bayes estimators of β corresponding to normal regression model are available in closed form, we easily obtained the same for each of the data values given in Tables **1-3**. The Bayes estimate of σ for each of the three categories of patients was obtained using ordinary Monte Carlo simulation by generating 5000 inverted gamma variates with shape parameter (n-p)/2 (n=10, p=3) and scale parameter $\nu s^2/2$ and retaining the sample based estimate in the form of posterior median and mode. The estimates corresponding to control group, cholelithiasis group, and carcinoma group are shown in Table **4**. The values with asterisk represent the posterior modes. In the table β_{N0}, β_{N1}, β_{N2}, and σ (β'_{N0}, β'_{N1}, β'_{N2}, and σ′) are the parameters associated with the situation when the primary acid CA (CDCA) is regressed with the secondary bile acids DCA and LCA. Similarly, $\beta^s_{N0}, \beta^s_{N1}, \beta^s_{N2}$ and σ^s ($\beta'^s_{N0}, \beta'^s_{N1}, \beta'^s_{N2}$ and σ'^s) are the parameters associated with the situation when the secondary bile acid DCA (LCA) is regressed with the primary acids CA and CDCA.

Since β_{N0}, β'_{N0}, β^s_{N0}, and β'^s_{N0} are the intercepts, they can simply be regarded as the expected value of the corresponding response variable when all the other predictors are zero. It is obvious from the results of β_{N0} and β'_{N0} that the expected values of primary acids (CA and CDCA) decrease from Control group to Cholelithiasis group and then to Carcinoma group when both DCA and LCA are zero (see the values of estimated intercepts in Table **4**). These findings may be interpreted in two ways although the interpretation may not often be relevant. First, decrease in the levels of primary bile acid, on average, causes gallstones and then causes carcinoma among the patients who have very small (almost zero) values of DCA and LCA. Second, the gallstones and the carcinoma reduce the levels of CA and CDCA when DCA and LCA values are very small (close to zero). We definitely cannot use this finding to conclude that carcinoma of gallbladder is caused by decrease in primary bile acids (due to limitation of our available data) although we are tempted to say that the decrease in the levels of CA and CDCA and occurrence of gallstones and carcinoma may be related, either one causing the other, when DCA and LCA are almost zero. A similar conclusion can be drawn based on β^s_{N0} and β'^s_{N0} when the two secondary acids are considered as response variables (see Table **4**) although we do not find a specific trend when LCA is regressed with CA and CDCA. A word of remark: the validity of the conclusion based on intercepts is correct only if the predictor variables are all zero (or quite close to zero), a situation which has no relevance in true sense.

Similarly, the regression coefficients β_{N1}, β_{N2}, β'_{N1}, β'_{N2}, $\beta^s_{N1}, \beta^s_{N2}$, β'^s_{N1}, and β'^s_{N2} represent the expected change in the concerned response variable per unit change in the associated predictor variable at a constant value of other predictor variable. We shall not be able to speak much based on these estimated

Table 4: Bayes Estimators of Normal Regression Parameters Under the Assumption of Each Biliary Acid Constituent as a Response Variable

Parameters	Bayes estimates corresponding to		
	Control group	Cholelithiasis group	Carcinoma group
β_{N0} (β'_{N0})	23.095 (21.409)	6.246 (7.802)	4.367 (4.132)
β^s_{N0} (β'^s_{N0})	5.801 (0.179)	1.502 (-0.057)	0.765 (1.654)
β_{N1} (β'_{N1})	-1.176 (-0.416)	0.166 (0.271)	-0.113 (0.535)
β_{N2} (β'_{N2})	-1.508 (-4.615)	4.502 (5.146)	0.581(-0.448)
β^s_{N1} (β'^s_{N1})	-0.328 (0.010)	0.067 (0.024)	-0.161 (0.257)
β^s_{N2} (β'^s_{N2})	0.185 (-0.013)	0.103 (0.019)	1.232 (-0.046)
σ (σ′)	5.729 (5.584) *4.905 (*4.893)	2.817 (3.218) *2.357(*2.773)	1.993 (1.616) *1.701(*1.417)
σ^s (σ'^s)	2.363 (0.205) *2.028(*0.176)	2.500 (0.225) *2.115(*0.192)	2.721 (1.427) *2.316(*1.220)

values except that we can point out the kind of relationship between the corresponding response variable and the associated predictor variable for each category of patients. Say, for example, estimated values of both β_{N1} (β'_{N1}) and β_{N2} (β'_{N2}) are showing a negative relationship of CA (CDCA) with DCA and LCA in the control group. A similar conclusion can be drawn based on other estimated values of regression coefficients for different categories of patients. The estimated σ, σ', σ^s, and σ'^s show the variability of the corresponding response variable for the three categories of patients. It can be seen that the variability, in general, decreases from Control group to Cholelithiasis and then to Carcinoma group for CA and CDCA whereas it, in general, increases for DCA and LCA.

5.2. Analysis Based on Gamma Regression Model

We next considered analyzing the same data sets (Tables **1-3**) using the gamma based regression modelling formulation discussed in subsection 2.2. The study was conducted exactly in a way as it was done earlier for normal regression modelling case. We use the subscript G with β to specify the estimates based on gamma regression model. It is to be noted here that the posterior corresponding to this situation (see subsection 2.2) is not available in closed form and, therefore, we used WINBUGS software to simulate the posterior and to obtain the estimates of the concerned parameters. The estimates corresponding to control group, cholelithiasis group, and carcinoma group are shown in Table **5**. The values with asterisk represent the posterior modes. In the table β_{G0}, β_{G1}, β_{G2}, and ν (β'_{G0}, β'_{G1}, β'_{G2}, and ν') are the parameters associated with the situation when the primary acid CA (CDCA) is regressed with the secondary bile acids DCA and LCA. Similarly, $\beta^s_{G0}, \beta^s_{G1}, \beta^s_{G2}$ and ν^s ($\beta'^s_{G0}, \beta'^s_{G1}, \beta'^s_{G2}$ and ν'^s) are the parameters associated with the situation when the secondary bile acid DCA (LCA) is regressed with the primary acids CA and CDCA.

We can conclude similarly on the basis of the estimated intercepts and the regression coefficients but most of the conclusions have their own limitations as it was seen in the normal case. Say, for example, the intercept provides the expected value of the corresponding response variable only if the associated predictors are all zero, a situation that can be hypothetical only since we shall never have the predictors zero for the situation under consideration. Similarly, one can conclude on the basis of estimated regression coefficients possibly for the purpose of exploring the kind of relationships among the different variables. The estimated shape parameters, however, do provide an important conclusion. It can be seen that these estimates are often quite large and, therefore, advocate for the use of normal model as the gamma model with large shape parameter can be approximated by a normal model. In situations where

Table 5: Bayes Estimators of Gamma Regression Parameters Under the Assumption of Each Biliary Acid Constituent as a Response Variable

Parameters	Bayes estimates corresponding to		
	Control group	Cholelithiasis group	Carcinoma group
β_{G0} (β'_{G0})	2.887 (2.915) *2.896(*2.939)	1.768 (2.015) *1.786(*2.039)	0.963 (1.511) *0.976(*1.489)
β^s_{G0} (β'^s_{G0})	1.908 (-1.169) *1.917(*-1.607)	0.360 (-2.442) *0.254(*-2.462)	0.847 (0.454) *0.907(*0.567)
β_{G1} (β'_{G1})	0.023 (0.026) *0.020(*0.025)	0.050 (0.049) *0.039(*0.300)	0.022 (0.066) *0.021(0.068)
β_{G2} (β'_{G2})	0.206 (-0.041) *0.141(*-0.122)	0.586 (0.512) *0.556(*0.457)	0.133 (-0.031) *0.144(*-0.032)
β^s_{G1} (β'^s_{G1})	0.040 (0.267) *0.039(*0.198)	0.087 (0.082) *0.058(*0.067)	0.044 (0.135) *0.020(*0.097)
β^s_{G2} (β'^s_{G2})	-0.074 (-0.300) *-0.064(*-0.276)	0.011 (0.065) *0.021(*0.070)	0.148 (-0.016) *0.145(*-0.026)
ν (ν')	11.560 (15.200) *10.860(*14.900)	11.950 (11.270) *9.340(*10.930)	6.583 (32.310) *5.183(*33.810)
ν^s (ν'^s)	1.492 (0.479) *1.195(*0.406)	2.547 (5.049) *2.041(*4.059)	13.470 (3.334) *11.560(*2.659)

the estimated shape parameters are not too large, one can possibly think for skewed regressor variables and, therefore, for gamma model as an important candidate.

5.3. Real Data Analysis Continued: Predictive Simulation and Model Comparison

So far, we considered the problem of analyzing the two regression models for the assumed data sets so that the parameters involved in the models can be estimated and, as such, the models can be made completely specified. Our primary task is definitely beyond this. Actually, we need to visualize the changing patterns, if any, among the different biliary acid constituents from one category of patients to another. Before we comment on this important issue, let us examine the compatibility of the assumed models with the data in hand so that the entertained models can be justified before we proceed.

For studying the model compatibility, we first generated 1000 predictive samples each of size 10 separately for each of the four response variables using the values of the corresponding predictor variables. This was done independently for all the three categories of patients. To clarify, let us consider, for example, the control group with CA as the response variable and DCA and LCA as predictor variables. So we generated 1000 predictive samples each consisting of 10 observations (one observation corresponding to one patient) for CA using the corresponding estimates of β_{N0}, β_{N1}, β_{N2}, and σ (respectively reported as 23.095, -1.176, -1.508, 5.729 in Table **4**) and successive values of DCA and LCA in the same category. The predictive samples for CA was also obtained for Cholelithiasis and Carcinoma groups of patients using corresponding estimates of β_{N0}, β_{N1}, β_{N2}, σ and the corresponding values of the predictor variables DCA and LCA. The procedure was similarly repeated for all the four

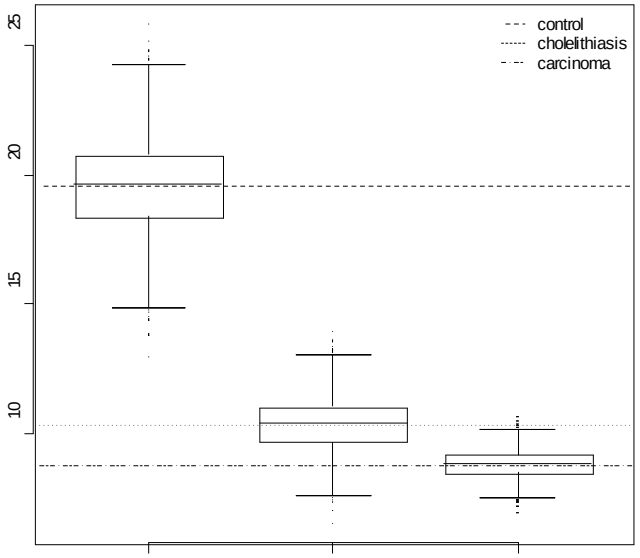

Figure 2: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when CDCA is the response variable under normal regression model (horizontal lines are the corresponding observed data means).

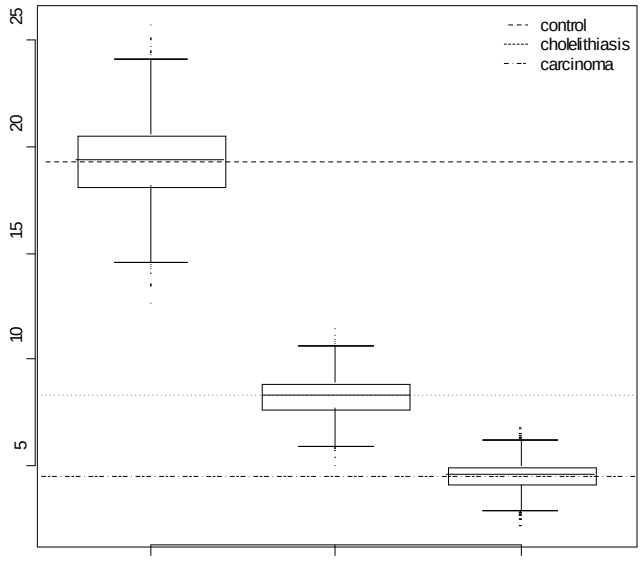

Figure 1: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when CA is the response variable under normal regression model (horizontal lines are the corresponding observed data means).

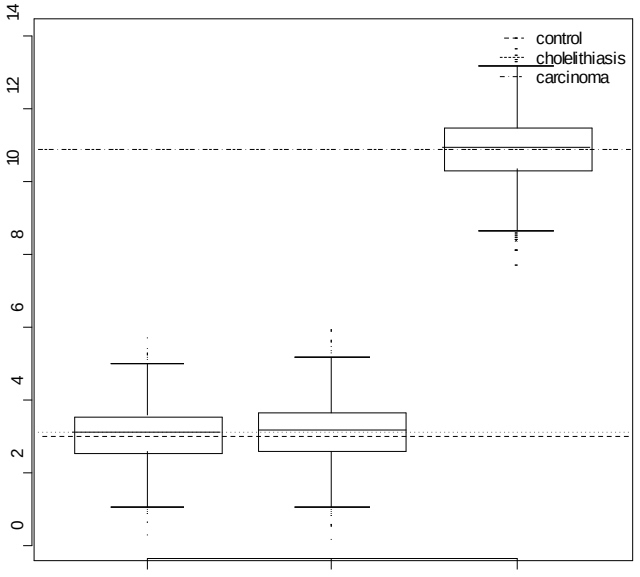

Figure 3: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when DCA is the response variable under normal regression model (horizontal lines are the corresponding observed data means).

response variables and all the three categories of patients using corresponding estimates of β's and corresponding values of predictor variables (see also subsection 4). We lastly calculated 1000 predictive means based on 1000 simulated predictive samples under each considered category of response variable and obtained the density estimates of predictive means using boxplot representations.

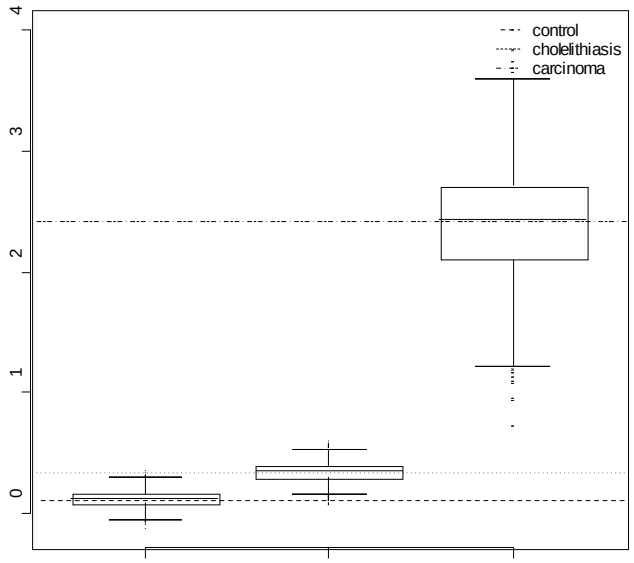

Figure 4: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when LCA is the response variable under normal regression model (horizontal lines are the corresponding observed data means).

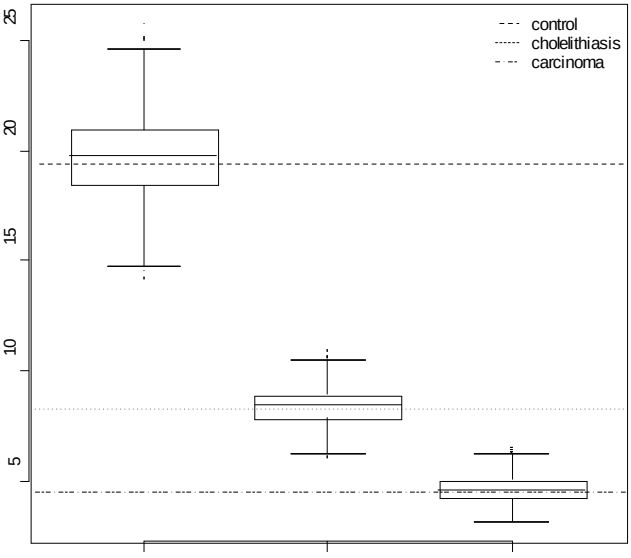

Figure 5: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when CA is the response variable under gamma regression model (horizontal lines are the corresponding observed data means).

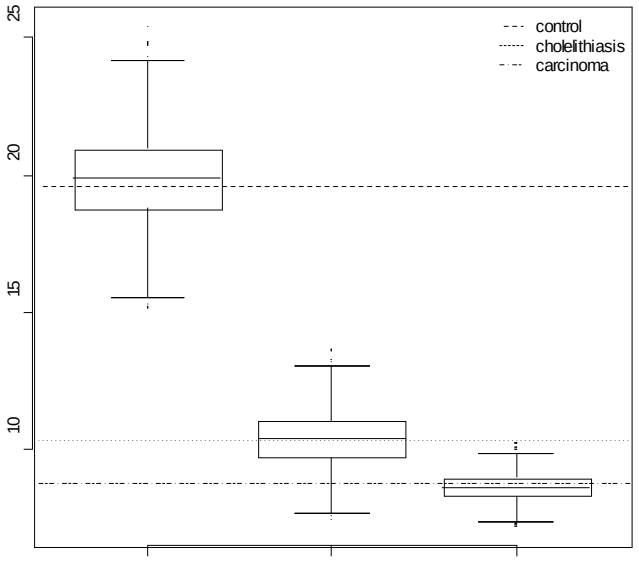

Figure 6: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when CDCA is the response variable under gamma regression model (horizontal lines are the corresponding observed data means).

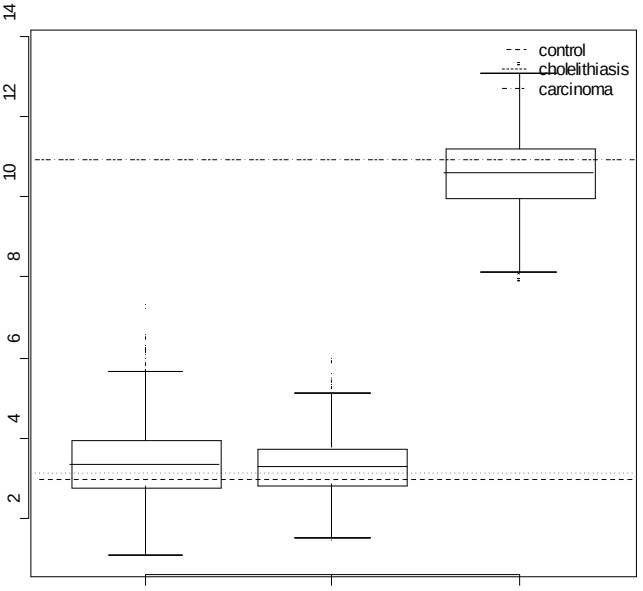

Figure 7: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when DCA is the response variable under gamma regression model (horizontal lines are the corresponding observed data means).

The predictive density estimates in the form of boxplots are shown in Figures **1-4** when each bile acid constituent is considered as a response variable. These figures are based on normal modelling assumptions for the associated response variables. Similarly, the boxplots based on gamma modelling assumption for the response variables are shown in Figures **5-8**. In each of these boxes the dotted

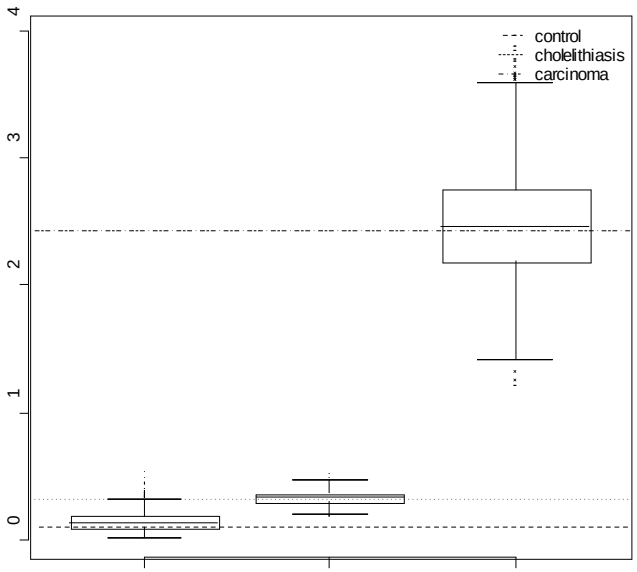

Figure 8: Density estimates showing the predictive means for Control, Cholelithiasis, and Carcinoma groups (left to right) when LCA is the response variable under gamma regression model (horizontal lines are the corresponding observed data means).

horizontal lines represent the corresponding observed data means. In addition, each figure shows three boxes. These three boxes correspond to Control group, Cholelithiasis group, and Carcinoma group of patients (from left to right). It can be seen from the figures that the mean values for all the observed data lines pass through the central regions in the corresponding boxes of predictive data means and, therefore, both the models are quite compatible with the data under consideration. A simple remark: both normal and gamma modelling assumptions are almost similar, the latter because of the logarithmic transformation for the link function, except that the gamma model may be able to cover skewness in the data, if any, as well.

We also drew predictive density estimates of individual observations based on matplots with superimposed observed data points in the form of bold dots. The plots are displayed in Figures **9**, **10** corresponding to normal and gamma regression models when CA is the response variable and DCA and LCA are the predictor variables. These figures correspond to Carcinoma patients. It is to be noted here that we drew similar plots for all other categories of response variables and different categories of patients (similar to what we have presented in Figures **1-8**). These pictures appeared to be more or less similar in appearance to those shown in Figures **9**, **10**. We, therefore, omit the other figures due to space restriction though the conclusion remains the same that both the models are quite compatible with all the three categories of data (see Tables **1-3**).

The Figures **1-8** do provide yet another important message when considered without horizontal observed data lines. It can be clearly seen that the density estimates in the form of boxplots show a clear-cut trend for the three categories of patients. Let us consider, for example, the primary acids CA and CDCA. It can be seen that these two biliary acids decrease, in general, from Control group to Cholelithiasis group and then to Carcinoma group (see Figures **1**, **2, 5** and **6**) whatever modelling assumption is used for the response variable. Similarly it can be seen that the secondary acids DCA and LCA, in general, increase from Control group to Cholelithiasis group and then to Carcinoma group irrespective of the choice of models for the response variables (see Figures **3**, **4** and Figures **7**, **8**). This may be interpreted in several ways. For instance, one can say that the decreasing levels of primary acids and simultaneously increasing levels of secondary acids are the main causes for developing gallstones and finally carcinoma. The other interpretation can also be likewise given. That is, the development of gallstones and carcinoma results in lowering down the primary acids and raising the secondary acids. This latter conclusion, not so clearly revealed, has often been provided by the medical practitioners based on Chromatographic or other statistical studies and they claim that the gallstones and carcinoma cause some bacterial degradation in the patients and, as a result, such changes in biliary acid constituents may occur. Our data, however, do not support any such conclusion due to independent set of observations in the three categories. In either case, one finding is quite obvious. That is, the decreasing levels of CA and CDCA (Figures **1**, **2**, **5** and **6**) and increasing levels of DCA and LCA (Figures **3**, **4**, **7** and **8**) may be considered as clear indicative of gallstones which may perhaps finally result in carcinoma. Therefore, once this tendency starts appearing and the patient is diagnosed with gallstones, he or she should immediately become cautious so that any further change in the levels of biliary acid may not occur and the carcinoma of gallbladder may be avoided. Such suggestions are often given by the medical practitioners based on their experiences or some preliminary investigations but the easily established analytical support to such finding can be considered as the chief feature of our Bayes analysis.

Before we end the section, let us provide some results on model comparison as well although the

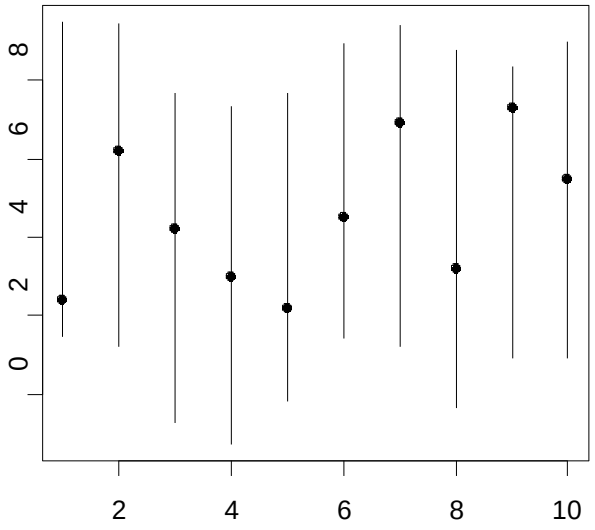

Figure 9: Density estimates showing each individual predictive data for Carcinoma group of patients when CA is the response variable under the assumption of normal regression model (dots exhibit the corresponding observed data points).

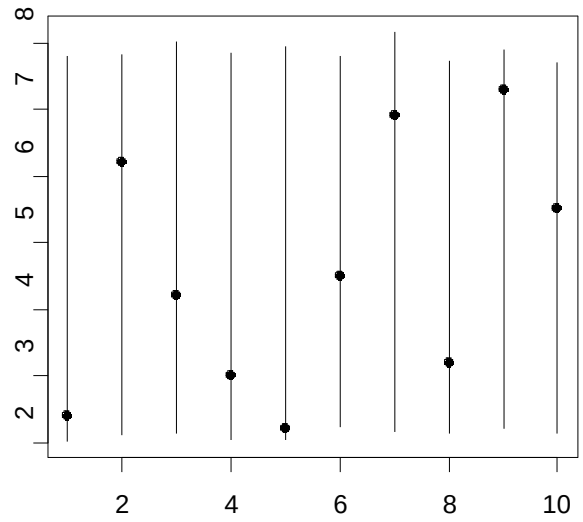

Figure 10: Density estimates showing each individual predictive data for Carcinoma group of patients when CA is the response variable under the assumption of gamma regression model (dots exhibit the corresponding observed data points).

conclusions based on the two models especially with regard to predictive inferences are more or less similar and, therefore, convey that any of the two models can be used. For the purpose of comparison, we simply obtained BIC based on the two models. It is to be noted that the choice of BIC as a model comparison criterion was merely for the sake of ease and one can go for other sophisticated measures as well (see Upadhyay and Mukherjee (2008)) [37]. The values of BIC are shown in Table **6** for different modelling assumptions on response variables. It is obvious that the two models are equally good if CA and CDCA are the response variables whatever kind of patients one considers since the values of BIC corresponding to the two models are quite close. Similarly, for the situations when DCA and LCA are the response variables, the gamma model can be recommended for Control and Cholelithiasis groups whereas for Carcinoma group either of the two models can be taken (see Table **6**). As a word of final remark: since the two models are equally complex with regard to involving same number of unknown parameters and the inferences with which we are concerned are more or less same, we advocate the use of any of the two models for the assumed data sets. The comparatively large differences in the values of BIC corresponding to the two models especially for Control and Cholelithiasis groups may be attributed to the fact that the observations are quite small (see Tables **1**, **2**), sometimes even zero or close to zero, and, as a result, BIC is largely affected in these two groups.

6. CONCLUSION AND RECOMMENDATION

The paper provides an extensive study of different categories of gallbladder patients to examine how the biliary acid constituents change from Control group to Cholelithiasis group and then to Carcinoma group. Two special cases of GLM are used for the intended study. A real data based study reveals that the two primary acids, namely CA and CDCA, decrease from Control group to Cholelithiasis group and further decrease among the Carcinoma group of patients. The tendency is reversed on case of secondary biliary acids, namely DCA and LCA. Model validation and model comparison have also been taken up using predictive simulation ideas and BIC, respectively.

The findings given in the paper are often reported by the medical practitioners based on the Chromatographic study but we do not find any statistically established results of such analyses especially using Bayes paradigm. The decreasing tendency of primary biliary acids and the increasing tendency of the secondary biliary acids as we move from Control to Cholelithiasis and then to Carcinoma group can be an important message to medical practitioners dealing with gallbladder diseases and who are primarily focused to take appropriate preventive measures to avoid developing of Carcinoma. There is, however, a need to generate appropriate data so that more remarkable conclusions such as the following can be thought out. Say, for example, if the decreasing (increasing) trend in CA, CDCA (DCA, LCA) can be

Table 6: BIC for Normal and Gamma Regression Models on Different Response Variables in the Three Groups of Gallbladder Patients

Modelling assumption	Disease group	BIC when the response variable is			
		CA	CDCA	DCA	LCA
Normal	Control	70.858	68.799	53.947	2.723
	Cholelithiasis	55.857	58.595	52.422	5.853
	Carcinoma	49.233	58.056	66.410	42.153
Gamma	Control	70.381	68.271	50.183	-17.614
	Cholelithiasis	55.380	60.010	48.227	1.644
	Carcinoma	48.376	59.585	68.827	41.670

considered to be important factors for developing Cholelithiasis or even Carcinoma or if the diseases such as Cholelithiasis and/or Carcinoma cause a decrease in the levels of CA, CDCA and increase in the levels of DCA, LCA. It is to be noted that such conclusions are really meaningful from the viewpoints of medial practitioners.

REFERENCES

[1] Draper NR, Smith H. Applied Regression Analysis (3rd edition). Wiley, New York 1998. http://dx.doi.org/10.1002/9781118625590

[2] Congdon P. Applied Bayesian Modelling. Wiley: New York 2003. http://dx.doi.org/10.1002/0470867159

[3] Gelman A, Hill J. Data Analysis Using Regression and Multilevel/ Hierarchical Models. Cambridge University Press, Cambridge 2006. http://dx.doi.org/10.1017/cbo9780511790942

[4] Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Regression methods in Biostatistics: Linear, Logistic, Survival, and repeated measures models. Spinger- Verlag, New York 2005.

[5] Shukla VK, Tiwari SC, Roy SK. Biliary bile acids in cholelithiasis and carcinoma of the gallbladder. Eur J Cancer Prev 1993; 2: 155-160. http://dx.doi.org/10.1097/00008469-199303000-00008

[6] Fraumeni JF. Cancer of pancreas and biliary tract: epidemiological considerations. J Cancer Res 1975; 35: 3437-3446.

[7] Shukla VK, Khandelwal C, Roy SK, Vaidya MP. Primary carcinoma of the gallbladder: A review of a 16 year period at the University Hospital. J Surg Oncol 1985; 28: 32-35. http://dx.doi.org/10.1002/jso.2930280109

[8] Nelder JA, Wedderburn RWM. Generalized linear models. J Roy Statist Soc Ser A 1972; 135: 370-384. http://dx.doi.org/10.2307/2344614

[9] McCullough P, Nelder JA. Generalized Linear Models. Chapman and Hall, London 1989. http://dx.doi.org/10.1007/978-1-4899-3242-6

[10] Breslow NE. Extra Poisson variation in log-linear models. Appl Statist 1984; 33: 38-44. http://dx.doi.org/10.2307/2347661

[11] Lawless JF. Negative binomial and mixed Poisson regression. Canadian J Statistics 1987a; 15: 209-225. http://dx.doi.org/10.2307/3314912

[12] Lawless JF. Regression methods for Poisson process data. J Am Stat Assoc 1987b; 82: 808-815. http://dx.doi.org/10.1080/01621459.1987.10478502

[13] Smyth GK. Generalized linear models with varying dispersion. J Roy Statist Soc Ser B 1989; 51: 47-60.

[14] Fahrmeir L, Tutz G. Multivariate Statistical Modeling Based on Generalized Linear Models. Spinger, New York 1991.

[15] Gelfand AE, Hills S, Racine-Poon A, Smith AFM. Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. J Am Stat Assoc 1990; 85: 972-985.

[16] Dey DK, Ravishanker N. Bayesian approaches for overdispersion in generalized linear models. Dey DK, Ghosh SK, Mallick BK, Eds. Generalized Linear Models: A Bayesian Perspective, Chapman & Hall/CRC Biostatistics Series 2000; pp. 73-88.

[17] Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. J Am Stat Assoc 1993; 88: 9-25.

[18] Lindsey JK. The uses and limits of linear models. Statistics and Computing 1995; 5: 87-89.

[19] Dellaportas P, Smith AFM. Bayesian inference for generalized linear and proportional hazards models *via* Gibbs sampling. Appl Statist 1993; 42: 443-460. http://dx.doi.org/10.2307/2986324

[20] Ibrahim JG. Incomplete Data in Generalized Linear Models. J Am Stat Assoc 1990; 85: 765-769. http://dx.doi.org/10.1080/01621459.1990.10474938

[21] Laud PW, Ibrahim JG. Predictive model selection. J Roy Statist Soc Ser B 1995; 57: 247-262.

[22] Das S, Dey DK. On Bayesian analysis of generalized linear models using jacobian technique. Am Statist 2006; 60: 264-268. http://dx.doi.org/10.1198/000313006X128150

[23] Das S, Dey DK. On Bayesian analysis of generalized linear models: A new perspective, Technical Report, Statistical and Mathematical Sciences Institute, North Carolina, USA 2007.

[24] Casella G, Robert CP. Rao-Blackwellization of sampling schemes. Biometrika 1996; 83: 81-94. http://dx.doi.org/10.1093/biomet/83.1.81

[25] Gelfand AE, Sahu S. Gibbs Sampling, identifiability and improper priors in generalized linear mixed models. J Am Statist Assoc 1999; 94: 247-253. http://dx.doi.org/10.1080/01621459.1999.10473840

[26] Marin JM, Robert CP. A Practical Approach to Computational Bayesian Statistics. Springer-Verlag, New York 2007.

[27] Nelder JA. Weighted regression, quantal response data, and inverse polynomials. Biometrics 1968; 24: 979-985. http://dx.doi.org/10.2307/2528884

[28] Finney DJ. Probit Analysis. University Press, Cambridge 1952.

[29] Zellner A. An introduction to Bayesian Inference in Econometrics. Wiley, New York 1971.

[30] Aitkin M, Anderson D, Francis B, Hinde J. Statistical Modelling in GLIM. Clarendron Press, New York 1989.

[31] Green PJ, Silverman BW. Nonparametric Regression and Generalized Linear Models. Chapman & Hall, London 1994. http://dx.doi.org/10.1007/978-1-4899-4473-3

[32] Lindsey JK. Applying Generalized Linear Models. Springer, New York 1997.

[33] Gill J. Generalized Linear Models: A Unified Approach. Sage University Press, California 2001.

[34] Dobson AJ. An introduction to generalized linear models. Chapman & Hall, London 2001. http://dx.doi.org/10.1201/9781420057683

[35] Congdon P. Bayesian Statistical Modelling. Wiley: New York 2007.

[36] Ghosh JK, Hjort N, Messan AA, Ramamoorthi RV. Bayesian Bivariate Survival Estimation. Journal of Statistical Planning and Inference 2006; 136: 2297-2308.

[37] Upadhyay SK, Mukherjee B. Assessing the value of the threshold parameters in the Weibull distribution using Bayes paradigm. IEEE Trans Reliab 2008; 57: 489-497. http://dx.doi.org/10.1109/TR.2008.928196

Received on 23-04-2015 Accepted on 09-05-2015 Published on 21-05-2015

http://dx.doi.org/10.6000/1929-6029.2015.04.02.9