A Bayes Study of Bile Acid Constituents on Cholelithiasis and Carcinoma of the Gallbladder

Richa Srivastava¹*, Satyanshu K. Upadhyay² and Vijay K. Shukla³

¹Department of Statistics, Banaras Hindu University, Varanasi-221005, India
²Department of Statistics and DST Centre for Interdisciplinary Mathematical Sciences, Banaras Hindu University, Varanasi-221005, India
³Department of General Surgery, Institute of Medical Sciences, Banaras Hindu University, Varanasi-221005, India

Abstract: Background: Significantly high concentration of secondary bile acids and low concentration of primary acids are reported by the surgeons in patients with gallbladder carcinoma (GBC) compared to control and cholelithiasis groups.

Aim: To examine the effect of primary and secondary bile acids on the development of cholelithiasis and GBC and to investigate the association, if any, among the two groups of bile acids.

Methods: The study uses two groups of patients at a time selected in accordance with some pre-fixed inclusion and exclusion criteria. Informed consent was obtained from all patients. The demographic characteristics such as mean age, sex ratio and body mass index, etc. are obtained for the selected groups of patients. The study defines dichotomous responses and the four bile acid constituents, namely cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA) and lithocholic acid (LCA), as the predictors. It then assumes logistic regression model to associate the binary responses with the predictors by using probability scores. Bayes analysis is developed using Markov chain Monte Carlo (MCMC) pack in R software for the posterior simulation.

Results: Twenty one cholelithiasis patients and twenty patients in each of control and GBC groups are studied. It is seen that a unit decrement in the level of CA (CDCA) increases the log (odds ratio) for cholelithiasis by an amount of 0.49 (0.14) and odds ratio by almost 1.5 (1.12). Similarly, a unit increment in the level of DCA (LCA) provides the log (odds ratio) for cholelithiasis as 0.18 (1.3) and odds ratio as 1.16 (2.95). Comparing GBC with control population, it is noted that a unit decrease in the level of CA (CDCA) in the control population increases the log (odds ratio) for GBC by an amount of 1.16 (0.26) and odds ratio by almost 2.63 (1.24) times. Similarly, the log(odds ratio) for GBC increases by 0.77 (1.94) and the odds ratio increases by 1.9 (5.0) for the unit increment in the level of DCA(LCA).

Conclusions: The study observes relatively high variations in the primary and secondary bile acids in the cholelithiasis and GBC groups as compared with the control group. It, in turn, reflects strong association among the two categories of bile acids in gallbladder diseases.

Keywords: Bile acid constituents, Cholelithiasis, Gallbladder carcinoma, Logistic regression model, Vague prior, Odds ratio, Posterior simulation.

1. INTRODUCTION

The incidence of the gallbladder diseases varies greatly in different parts of the world. The high incidence areas are Chile, Israel and American Indians although the disease is observed in rest of US as well. Among the various diseases, cholelithiasis is perhaps the commonest which requires surgical removal of the gallbladder of the patient with high rate of success. GBC, on the other hand, has attracted maximum attention of surgeons as it often involves life risk of the patients. Pandey and Shukla [1] noted that GBC is the most common malignancy of the biliary tract and perhaps the fifth frequently diagnosed malignancy of the gastrointestinal tract. It is the third most common gastrointestinal malignancy in the north-eastern part of India [2] and, as a result, it will not be illogical to say that the disease divides the country in two parts. A crude incidence rate of GBC in north-eastern part of India was recently estimated to be 15.0 per 100,000 females and 5.9 per 100,000 males [3].

Primary GBC was first documented by deStoll [4] and still management of the disease is a challenge for the clinicians. The reason may be mainly attributed to the unclear etiology and, in majority of cases, the late diagnosis.

Etiological study of GBC suggests that gallstone is perhaps the most important risk factor. A number of studies have been carried out in the past and it was found that gallstone was observed in nearly 40-100% of the cases of GBC. Adson [5] documented an association of cholelithiasis in more than 75% cases.
Similar associations were also observed by Hart et al. [6] and Hart and Madan [7] in separate studies based on their observations. Some recent studies include the role of mustard oil and nicotine in gallbladder carcinoma [8, 9].

Etiology of the biological factors responsible for the evolution of the disease includes some specific distribution of bile acid constituents, namely cholic acid (CA), chenodeoxycholic acid (CDCA), deoxycholic acid (DCA) and lithocholic acid (LCA). Among these different constituents, the first two are categorized as the primary bile acids whereas the last two are regarded as the secondary bile acids.

Studies were carried out to explore the association of cholelithiasis and GBC with different concentration levels of bile acid constituents. There is a large body of evidence in favour of the hypothesis that different levels of bile acid constituents may be treated as carcinogenic tumor promoters [10, 11]. For instance, an increased concentration level of secondary bile acids and the decreased concentration level of primary bile acids may be treated as tumor promoters and initiators. Shukla et al. [12] reported such a finding based on GBC patients at University hospital, Varanasi. Their reporting was, however, based on chromatographic study and some elementary classical statistical tools. Extensive statistical studies for such evidences can also be seen in the literature. An important reference includes Makkar [13] that provides a complete Bayesian study based on bile samples of control and diseased groups where diseased groups include observations from both cholelithiasis and carcinoma patients. The author used an alternative modelling formulation based on generalized linear model (GLM). It was observed that the increased level of secondary bile acids and the decreased level of primary acids may be treated for the development of both gallstones and GBC [12]. Although the model was properly justified based on some exploratory data tools and using the ideas of predictive simulation [13], the choice of GLM remained unappealing. Besides, the study used limited data resources.

The present study is carried out to examine the role of bile acid constituents as the causal factors for cholelithiasis and carcinoma of the gallbladder. For the purpose, we have used an alternative modelling formulation other than the one considered by Makkar [13] and performed the complete Bayes analysis based on the data from the three groups of patients, namely the control, cholelithiasis and GBC groups. Our modelling formulation uses logistic regression model with independent vague priors for the model parameters and it has an advantage of being easily implemented by a number of software. It is to be noted that the logistic regression model is proposed by associating dichotomous variable with each patient considering two groups together. The primary objective, therefore, includes identifying the role, if any, of bile acid constituents in the patients with the gallbladder diseases and the same is attempted using the estimated values of logistic regression parameters, the latter can also be interpreted as the log odds ratio.

2. METHODS

2.1. Patients and Samples

The study considers bile samples from 61 cases admitted to the University Hospital, Varanasi. We had three groups, namely, control, cholelithiasis and those diagnosed with GBC. The cholelithiasis group consisted of 21 patients whereas remaining two groups were each having 20 patients. It is to be noted here that the earlier studies [12, 13] considered only 30 observations (10 in each category) and, as such, the present study can be considered more appealing statistically.

The inclusion criterion for the subjects of this study include histologically confirmed cases of gallbladder cancer, ultrasonographically confirmed cases of gallstone and cases with diagnosis other than gallstones and GBC. Similarly, the exclusion criterion for the subjects include patients below 18 years of age, pregnant women, patients of cancer being treated with chemotherapy, patients suffering from cancer other than GBC, patients undergoing hormone therapy, patients with history of small bowel resection, patients on immunosuppressants and patients with prolonged fasting or parenteral nutrition.

The study was approved by ethics committee of the Institute of Medical Sciences, Banaras Hindu University. The patients were recruited for the study after their written informed consents.

The mean age of the control, cholelithiasis and GBC patients were 39.5, 38.72 and 45.10 years, respectively. Control patients included both females and males in the ratio 7:3. However, the females to males ratio in cholelithiasis and GBC groups were 9.5:1 and 9:1, respectively. Thus females accounted for bulk of cases in both the groups. Body mass index (B.M.I.)
was also calculated in the study resulting into mean B.M.I. for cases and controls as 28.23 ± 3.44 and 23.14± 5.47 kg/m², respectively, with standard ranges between 18 and 30.

The bile samples were collected by the fine needle aspiration of the gallbladder from the three groups. Care was taken to exclude the contamination of bile samples with blood. These samples were kept at −20°C until observations were recorded. Concentration (mg/ml) of bile constituents, namely CA, CDCA, DCA and LCA, were determined separately for each patient belonging to the three groups. Concentration of CA in mg/ml was determined independently for each sample by the method of Irvin et al. [14] and concentration of remaining bile constituents were determined according to their ratio to CA from gas-liquid chromatogram. The gas-liquid chromatogram used Chemito 3800 CC model, a 5 foot column with an inside diameter of 1/8 inch, containing 3% OV 17. Nitrogen was taken as the carrier gas and temperature was kept at 160°C; the injection temperature was 250°C [12].

2.2. The Data

Table 1 shows the mean distribution of primary and secondary bile acids in the three groups of patients. The values in the parentheses represent the corresponding standard deviations.

Table 1: Mean Distribution of Primary and Secondary Bile Acids in Control, Cholelithiasis and Carcinoma Groups (mg/ml)

<table>
<thead>
<tr>
<th>Bile constituents</th>
<th>CA</th>
<th>CDCA</th>
<th>DCA</th>
<th>LCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>control</td>
<td>20.12</td>
<td>17.57</td>
<td>2.86</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>(5.83)</td>
<td>(6.60)</td>
<td>(1.80)</td>
<td>(1.29)</td>
</tr>
<tr>
<td>cholelithiasis</td>
<td>13.05</td>
<td>12.06</td>
<td>4.97</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td>(5.26)</td>
<td>(3.84)</td>
<td>(4.83)</td>
<td>(1.67)</td>
</tr>
<tr>
<td>GBC</td>
<td>5.90</td>
<td>9.02</td>
<td>12.71</td>
<td>2.87</td>
</tr>
<tr>
<td></td>
<td>(5.15)</td>
<td>(4.54)</td>
<td>(6.41)</td>
<td>(2.17)</td>
</tr>
</tbody>
</table>

We, however, consider two groups of patients at a time for the purpose of analysis entertained in the paper and associate three outcome variables \(y^j, j=1,2,3\) as the response variables and the four bile acid constituents, namely CA, CDCA, DCA and LCA, as the predictor variables, certain levels of which may be considered as the risk factors for developing cholelithiasis and GBC. Considering the structure as discussed, the likelihood functions corresponding to the three dichotomous variables \(y^j, j=1,2,3\) can be written as

\[L_j(y^j \mid p_j) \propto \prod_{i=1}^n p_j^{y_{ji}} (1 - p_j)^{1-y_{ji}}, \quad j=1,2,3\]  

where \(p_j\) is the probability that the dichotomous variable \(y^j\) takes value unity. The assumption of linear logistic regression model leads to the link function

\[
\log \frac{p_j}{1-p_j} = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}, \quad j=1,2,3
\]

where \(\alpha_0\) is the intercept and \(\alpha_j\) is the regression coefficient associated with the corresponding regressors variable \(x_{il}, l=1, ..., 4\). The parameter \(\alpha_j\) can also be interpreted as the log odds ratio associated with the corresponding predictor variable for developing the disease. A general interpretation of logarithm of odds ratio associated with a particular bile acid constituent can be given as under. A negative (positive) value

Similarly, considering cholelithiasis and carcinoma groups together, the corresponding outcome variables \(y^j\) can be defined, for each patient as

\[y_j^{(1,2)} = \begin{cases} 1 & \text{if patient } i \text{ is in cholelithiasis (carcinoma) group} \\ 0 & \text{if patient } i \text{ is in control group} \end{cases} \]  

(2)

Different levels of bile constituents CA, CDCA, DCA and LCA in the corresponding groups can then be considered as the regressors variables. Thus \(y_j^{(1,2)}\) will take value unity for all the patients in cholelithiasis (carcinoma) group and value zero for all the patients in control group. Similarly considering carcinoma and cholelithiasis groups together, \(y^3\) will take value unity for all the patients in the former group and value zero in the latter group.

2.3 Statistical Analysis

We compare each of the two categories with control group of patients, a group which is exposed to the risk factor but not developed the disease yet. For completeness, we also consider comparing the cholelithiasis group with GBC although this comparison is not desired as long as the control population is same. Our study uses the dichotomous variables \(y^j, j=1,2,3\) as the response variables and the four bile acid constituents, namely CA, CDCA, DCA and LCA, as the predictor variables, certain levels of which may be considered as the risk factors for developing cholelithiasis and GBC.
indicates the protective (causal) effect of the predictor variable in the occurrence of the disease for the unit increment in the same. On the other hand, a value close to zero indicates no or very little association of the disease with the associated risk factor. It is to be noted that the subscripts 1, 2, 3 and 4 for \( x_i \) denote the bile acids CA, CDCA, DCA and LCA, respectively.

Obviously, \( 1 - p_j \) denotes the probability that the dichotomous variable \( y^j \) takes value zero. Also, for each patient in the control group, we have assumed that the probability for getting into diseased group remains the same. This assumption seems justified as the patients belong to the same region and are exposed to the similar risks. Moreover, the probability \( p_j \) can be defined as

\[
p_j = \frac{\exp(X^j \alpha^j)}{1 + \exp(X^j \alpha^j)}, \quad j = 1, 2, 3
\]  

(5)

where \( X^j \alpha^j \) denotes the right hand side of (4).

In order to complete the Bayesian model formulation, we consider normal prior distribution for each regression coefficient as well as for the intercept. Thus

\[
\alpha^j_k : N(\mu^j_k, \sigma^2_k), \quad \forall \quad k = 0,1,...,4 \quad \text{and} \quad j = 1, 2, 3
\]

(6)

where \( \mu^j_k \), \( \sigma^2_k \) \( \forall \quad k = 0,1,...,4 \) and \( j = 1, 2, 3 \) are the prior hyperparameters. Since we do not have any specific information to specify these hyperparameters, we propose to assume prior means as zero in each case. These values indicate that any level of bile constituents do not have any causal or protective effect, on average, in developing the gallbladder diseases. Similarly, we may consider large variances for the normal priors so that the priors may remain proper but vague in each case. We considered a number of large choices for the normal variances (leading to vague prior distributions) although the results below are shown for a value 10.0 in each case.

Combining likelihoods with the corresponding priors via Bayes theorem yields the resulting posterior distributions that can be written up to proportionality as.

\[
\Pi_j \left[ \exp \left( \frac{\alpha^j_k - \mu^j_k}{2\sigma^2} \right) \right] \forall \quad j = 1, 2, 3
\]

(7)

Although logistic-normal combination is a standard one, posterior distributions, as shown above, are not available in closed forms. An important option may include simulation from these updated posterior distributions, which can be done using Markov chain Monte Carlo technique, in particular, the Gibbs sampler algorithm. Some important references are Upadhyay et al. [15, 16].

Using the choices as given above, the Gibbs sampler algorithm was implemented separately on each of the three posteriors using a single long run of the chain. For initial values, we considered least squares estimates of different parameters, which were obtained using R software. The convergence monitoring was done based on ergodic averages. Once the convergence monitoring was done, we took samples of size 1000 for each of the regression coefficients and intercepts. These samples were taken by choosing the equidistant values (every 10th) from the generating chains to minimize serial correlation among the generating variates. We have used MCMC pack in R (http://mcmcpack.wustl.edu/) for simulating from such posteriors. This software is freely available for use.

3. RESULTS

A preliminary inspection of the data in Table 1 shows that there is some variability in the mean levels of primary and secondary bile acids as revealed by the test results. The mean concentration levels of both CA and CDCA decrease from control to cholelithiasis and then to carcinoma group. This finding is more authenticated for CA than for CDCA as the variability also decreases for the former throughout although for the latter, the variability slightly increases in carcinoma group as compared to the same for cholelithiasis group. Secondary bile acids, DCA and LCA, also show variability in the bile samples of patients in all the three groups but in a reverse manner. Thus DCA and LCA have lowest mean concentration levels in the control group and then increasing trend from cholelithiasis to GBC groups. The observed standard deviations further authenticate this finding for the two secondary bile acids.
Table 2: Estimated log (odds ratio) and the Corresponding 95% Credible Interval in Parenthesis

<table>
<thead>
<tr>
<th>Bile constituents</th>
<th>Control vs Cholelithiasis $(\alpha_i^1, l = 1, \ldots, 4)$</th>
<th>Control vs GBC $(\alpha_i^2, l = 1, \ldots, 4)$</th>
<th>Cholelithiasis vs GBC $(\alpha_i^3, l = 1, \ldots, 4)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA</td>
<td>-0.49 $(0.84, -0.22)$</td>
<td>-1.16 $(-3.93, 0.54)$</td>
<td>-3.79 $(-8.09, 0.25)$</td>
</tr>
<tr>
<td>CDCA</td>
<td>-0.14 $(-0.41, 0.05)$</td>
<td>-0.26 $(-4.50, 2.69)$</td>
<td>-0.51 $(-1.52, 0.31)$</td>
</tr>
<tr>
<td>DCA</td>
<td>0.18 $(-0.12, 0.58)$</td>
<td>0.77 $(-2.74, 5.27)$</td>
<td>3.22 $(-0.19, 6.78)$</td>
</tr>
<tr>
<td>LCA</td>
<td>1.30 $(0.24, 2.53)$</td>
<td>1.94 $(-25.79, 30.38)$</td>
<td>1.36 $(-2.40, 5.96)$</td>
</tr>
</tbody>
</table>

The above observation although somewhat striking may not be considered to provide any statistically authenticated summary on the role of bile acid constituents on the cholelithiasis and carcinoma. We, therefore, performed the full Bayesian analysis as detailed above. The results are shown in Table 2 in the form of different a posteriori estimated log (odds ratios), $\alpha_i^j$ $j=1, 2, 3$.

It can be seen that negative estimates of $\alpha_i^1$ and $\alpha_i^2$ (Table 2) indicate that depreciation of primary acids increases the risk of gallstone. Moreover, since the absolute value of estimated $\alpha_i^1$ is, in general, larger than the corresponding value of $\alpha_i^2$, one can comfortably proclaim that depreciation of CA in bile has greater role on developing the gallstone than that of CDCA. It can be seen that a unit decrement (in mg/ml) in the level of CA increases the log (odds ratio) for cholelithiasis by an amount of 0.49 and odds ratio by almost 1.5. A similar interpretation can be given for the secondary bile acids but in a reverse manner. That is, positive signs of the regression coefficients of secondary bile acids suggest the causal role of their increments on the occurrence of gallstone in controls. It is clear that LCA has a greater role than DCA on developing the gallstone with a unit increment in the level causing the log (odds ratio) for cholelithiasis by an amount of 1.30 and odds ratio by almost 2.95.

Similarly, negative estimates of $\alpha_i^1$ and $\alpha_i^2$ (Table 2) suggest the role of decrement in the level of primary bile acids (CA and CDCA) for causing carcinogen. For instance, a unit decrease in the level of CA in the control population increases the log (odds ratio) for GBC by an amount of 1.16 and odds ratio by almost 2.63 times. A similar interpretation can be given for CDCA though it is not as vital as CA. Secondary bile acids once again provide positive estimates indicating that the increment in their values may be the causal factors for developing GBC. The odds ratio increases almost 1.9 (5.0) times for the unit increment in the level of DCA (LCA). Thus increase in the value of LCA may be considered a more significant causal factor for GBC than the corresponding increase in the value of DCA.

A comparison of malignant group with that of cholelithiasis is also shown in Table 2. Once again the estimated regression coefficients corresponding to the primary bile acids are negative whereas those corresponding to the secondary bile acids are positive giving a similar conclusion that was drawn earlier. The mean odds ratios for the regression coefficients are, however, in general higher than the corresponding values in control versus GBC indicating that the patients already developed gallstones are more likely to develop carcinogenesis than those who are in the control population.

Table 2 also provides estimated 95% credible interval. These values provide further insight into the estimated log (odds ratios) although a few lower (upper) limits might appear different from what we have concluded above. This may not be considered as a deterrent issue as such intervals also weigh towards the actual estimates reported in to the table.

The posterior density estimates of various regression coefficients are also shown by means of box plot representations in Figures 1-3. It is obvious that the density estimates are more or less symmetric and close to normality, a conclusion that otherwise seems obvious since the natural logarithm of odds ratio can be shown to be approximately normal [17]. It can be further seen that most of the estimated densities especially for $\alpha_i^1$’s and $\alpha_i^2$’s are highly condensed except for the regression coefficients corresponding to LCA which show, in general, large posterior variability. On the other hand, the estimated densities of $\alpha_i^3$’s are mostly dispersed except for the one corresponding to primary acid CDCA. A number of other conclusions can be like wise drawn and this is perhaps the apparent...
advantage of sample based approaches to Bayesian computation.

4. DISCUSSION AND CONCLUSION

Carcinoma of the gallbladder is the fifth most common malignancy of the gastrointestinal tract with poor survival rate. The management of the disease still remains a challenge for the surgeons as well as for the patients. A lot of literature is available on the causal factors of gallbladder diseases but none convey the exact message in favour of or against a factor with high degree of belief. In about 40% to 90% of the cases

Figure 1: Boxplots showing the density estimates of $\alpha^l$, $l=1, \ldots, 4$ (left to right).

Figure 2: Boxplots showing the density estimates of $\alpha^l$, $l=1, \ldots, 4$ (left to right).
stones are found in the gallbladders of carcinoma patients. This is perhaps the reason that gallstones are suggested as one of the main causal factors for promoting carcinoma.

Variation in the concentration levels of bile constituents are also observed in cholelithiasis and carcinoma patients in comparison to controls. Literature suggests that the reason for this variation is bacterial degradation which is ultimately responsible for the formation of secondary bile acids. It has been remarked that CDCA when altered by bacteria will get converted into LCA which is highly influential in both human and animal carcinogenesis [18]. Moreover, Cook [19] observed that methylcholanthrene, a highly potent carcinogen, could be obtained chemically from CA or dehydronorcholene although the role of CDCA cannot be ignored. A parallel study was done by Hill [20] but with reference to colon cancer. The author observed the incidence of colon cancer in patients having high proportion of anaerobes which is capable of degrading primary bile acids into secondary bile acids. There is a large body of evidence that bile acids are also tumor promoters if the normal proportion of the constituents is disturbed in a specific manner although the earlier studies are based on samples of size 10 only [12, 13].

The present study analyzes the distribution of bile acids in three groups of patients, namely the control, cholelithiasis and carcinoma groups. The study is first of its kind in the sense that it visualizes the data in an alternative manner by incorporating dichotomous variables and then analyzes the same by justifiably assuming logistic regression model. Among other things this assumption makes the analysis routinely manageable.

It is observed that the levels of primary acids (CA, CDCA) decrease in the cholelithiasis and carcinoma groups whereas those of secondary bile acids increase. Moreover, this decrease is even more appreciable in carcinoma group as compared to the gallstone group. Besides, it is also noted that the primary acid CA is more appreciably affecting the scenario than CDCA. This is obvious because the variability, as we move from the control group to cholelithiasis and then to carcinoma, is also highly affected in CDCA than in CA giving more precise and authenticated estimates for the latter than for the former (Table 1-2, Figures 1-3). Similarly the secondary bile acid LCA appears to be more influential than the secondary bile acid DCA. In fact, DCA and LCA, which represent small proportion in normal bile, are found to be higher in the concentration in the bile samples of gallstone patients and highest among the carcinoma patients. We can, however, notice that the LCA estimates are more trustworthy than DCA estimates as the former have small variability as compared to the latter.

Figure 3: Boxplots showing the density estimates of $\alpha_i^k$, $k=1, \ldots, 4$ (left to right).
The analysis based on control verses carcinoma and cholelithiasis verses carcinoma (Table 2) also convey the association of gallstone with malignancy. In fact, the possibility of developing GBC among those having cholelithiasis is much higher for unit depreciation of primary bile acids and unit augmentation in concentration level of secondary bile acids.

Overall conclusion suggests that relatively high variations in the primary and secondary bile acids in the cholelithiasis and GBC groups as compared with the control group reflect an association among these two bile acids. It might be possible that calculi induce an element of stasis, promoting chronic infection leading to increased turnover of primary bile acids to secondary bile acids, which are known tumor promoters and initiators [21]. The structure of the available data however leaves an issue unanswered and it needs to be examined thoroughly to get the exact scenario. One can certainly raise a doubt that if the decreasing (increasing) levels of primary bile (secondary bile) causes GBC and gallstone or if these diseases are responsible for the changes in the levels of bile constituents that we have observed. However, it requires an extensive study taking large follow up data.

**REFERENCE**