

# Determination of Causal Relationship Between Bilirubin and Other Liver Biomarker in Case of Hepatitis C

Proloy Banerjee<sup>1,\*</sup>, Anirban Goswami<sup>2</sup>, Shreya Bhunia<sup>1</sup>, Sudipta Basu<sup>1</sup>

<sup>1</sup>Department of Mathematics and Statistics, Aliah University, Kolkata, India

<sup>2</sup>Regional Research Institute of Unani Medicine, Guzri, Patna City, Patna, India

## Email address:

proloy.stat@gmail.com (P. Banerjee)

\*Corresponding author

## To cite this article:

Proloy Banerjee, Anirban Goswami, Shreya Bhunia, Sudipta Basu. Determination of Causal Relationship Between Bilirubin and Other Liver Biomarker in Case of Hepatitis C. *Biomedical Statistics and Informatics*. Vol. 6, No. 2, 2021, pp. 23-31. doi: 10.11648/j.bsi.20210602.11

**Received:** March 25, 2021; **Accepted:** April 12, 2021; **Published:** April 26, 2021

---

**Abstract:** *Background:* Liver works as one of the most versatile organs in the human body. But any kind of disturbance occurs in the liver may cause the liver disease. One of the most common liver infections is hepatitis C which is caused by the Hepatitis C Virus (HCV). It is well known that liver is the largest solid organ in the human body and also it is called the exocrine gland as it secretes bile into the intestine. *Aim:* The aim of this study is to evaluate the causal relationship of Bilirubin with each liver biomarker using the advanced regression techniques. *Methods:* We use two advanced regression techniques, namely Joint Generalized Linear Model (JGLM) and Generalized Additive Model (GAM). For model selection, we check the AIC value, GCV score and adjusted R-square as well as the different diagnostic plots like Q-Q plot, Residual vs. Fitted plot etc. are displayed. *Results:* Bilirubin, a human liver disease biomarker, is a brownish yellow substance found in bile and it is produced in the liver when the old red blood cells break down. The present study reveals that Bilirubin is positively associated ( $p\text{-value} < 0.05$ ) with Aspartate Aminotransferase (AST), Creatinine (CREA), Gamma-Glutamyl Transpeptidase (GGT), Protein (PROT), Alkaline Phosphatase (ALP)\*Albumin (ALB) and marginally associated with Choline Esterase (CHE)\*Cholesterol (CHOL) ( $p\text{-value} = 0.0591$ ). While it is negatively associated ( $p\text{-value} < 0.05$ ) with Age, Sex, Alkaline Phosphatase (ALP), Alanine Aminotransferase (ALT), Choline Esterase (CHE), Cholesterol (CHOL), Albumin (ALB), Creatinine (CREA)\*Gamma-Glutamyl Transpeptidase (GGT) under JGLM. Besides of that, Bilirubin is positively associated with AST, CREA, GGT, (CREA\*GGT), (CHE\*CHOL) whereas it is negatively associated with Sex, ALT, CHE, CHOL. Also, ALB is highly positively significant as a non-parametric smoothing term ( $p\text{-value} < 0.001$ ) under GAM. *Conclusion:* Both the advanced regression models JGLM and GAM explain the association between Bilirubin with other liver diseases biomarker in case of Hepatitis C.

**Keywords:** Bilirubin, Hepatitis C, Joint Generalized Linear Model, Generalized Additive Model, Gamma Distribution, Link function

---

## 1. Introduction

Hepatitis is the most common liver disease that occurs once the tissue cells within the liver get inflamed owing to alcohol or poison. In case of autoimmune hepatitis, some infectious diseases are spread when the immune cells within the body attack the liver in such a way so that the liver cells can be damaged severely. Hepatitis C is a liver infection caused by the Hepatitis C Virus (HCV) [30] and nowadays, a majority of people becomes infected with the hepatitis C

virus by sharing needles or alternative instrumentality used to prepare and inject medicine. Hepatitis C is a short-term condition for certain individuals, but for more than half of the people who become infected with the hepatitis C virus, may be exposed to a chronic and long-term illness. WHO has set a target to reduce new viral hepatitis infections by 90% and reduce deaths due to viral hepatitis by 65% by 2030 [34].

Generally, there are no symptoms in many individuals with hepatitis C. However, after the virus enters the bloodstream, between 2 weeks and 6 months abnormalities in human body such as dark urine, fatigue, jaundice, joint pain, nausea,

stomach pain, vomiting are noticed [29]. These complications can be prevented by early diagnosis and treatment of chronic hepatitis. For identifying the liver disease, many traditional diagnostic tests were measured on albumin (ALB), bilirubin (BIL), choline esterase (CHE),  $\gamma$ -glutamyl-transferase (GGT), aspartate amino-transferase (AST), alanine amino-transferase (ALT), Protein (PROT), Creatinine (CREA), Alkaline Phosphatase (ALP), Cholesterol (CHOL). Bilirubin levels in the blood goes up and down in patients with viral hepatitis C. It has been observed that when bilirubin levels remain high for long period of time, this usually means that major liver failure and likely cirrhosis are present. In several articles, it is found that there is a correlation between any two biomarkers of liver disease [26-28]. The dataset, collected from various patients together with some physical and interesting characters is supposed to be a multivariate data. As a result, a simple correlation between any two markers is useless, instead of that partial correlation is preferable. It is recommended to use an effective modeling to investigate the relationship between any biomarker with the remaining markers. Most of the previous research relied upon only on simple or multiple regression analysis, which is ineffective for the current dataset since the response variable are non-normal, heteroscedastic and positive. A number of studies using the simple statistical techniques are attempted to identify the factor related to liver diseases in case of Hepatitis C [31-33].

The objective of this study is to evaluate the causal relationship of Bilirubin with each liver biomarker using advanced statistical models. Remaining part of this paper has

been organized as follows, in section 2, we describe the data and statistical methodologies - Joint Generalized Linear Gamma Model and Generalized Additive Model. Section 3 represents the results of data analysis and Diagnostic plots. Findings and discussions are presented in section 4. Finally, in section 5 we conclude the article.

## 2. Materials and Statistical Methodology

### 2.1. Materials

This article is based on the secondary data on 615 individuals with a proven serological and histopathological diagnosis of hepatitis C [25] and it can be downloaded from “<https://archive.ics.uci.edu/ml/datasets/HCV+data>”. In this dataset, liver biopsies and blood samples for the examination of biochemical measurands on 615 hepatitis C patients with 14 cofactors were taken at the same time. The description of the covariates, factors and their levels with the summarized statistics such as the mean, standard deviation, range and proportion of the levels are provided in Table 1. Here we mainly focus on 589 hepatitis C patients with all non-missing information considering 12 cofactors by excluding two variables (Patient Id and Category) as we assume all the patients have hepatitis C virus. Among the cofactors, 11 are of continuous type and one is binary variable. In this analysis, we have considered BIL (Bilirubin) as the dependent variable and the remaining are treated as the independent or explanatory variables.

**Table 1.** Summary Statistics of variables in the analysis.

Variable	Abbreviation of variable	Mean $\pm$ SD / Frequency (%)	Range
Age of the patient (Year)	Age	47.42 $\pm$ 9.93	23 – 77
Gender (Male=1, Female=2)	Sex	1=363 (61.63%) 2=226 (38.37%)	-
Albumin	ALB	41.62 $\pm$ 5.76	14.9 - 82.2
Bilirubin	BIL	11.02 $\pm$ 17.41	0.8 – 209
Choline Esterase	CHE	8.2 $\pm$ 2.19	1.42 - 16.41
Gamma-Glutamyl Transpeptidase	GGT	38.2 $\pm$ 54.3	4.5 - 650.9
Aspartate Aminotransferase	AST	33.77 $\pm$ 32.87	10.6 – 324
Alanine Aminotransferase	ALT	26.58 $\pm$ 20.86	0.9 - 325.3
Protein	PROT	71.89 $\pm$ 5.35	44.8 - 86.5
Creatinine	CREA	81.67 $\pm$ 50.7	8 - 1079.1
Alkaline Phosphatase	ALP	68.12 $\pm$ 25.92	11.3 - 416.6
Cholesterol	CHOL	5.39 $\pm$ 1.13	1.43 - 9.67

### 2.2. Statistical Methods

Generally, in classical linear regression models, the standard assumption about response variable (Y) is that the variance is constant over the entire range of parameter values [3]. But in literature, it has been found that this assumption is not always satisfied by the real-life dataset [1-2]. For example, medical science data are usually heterogeneous in nature. To stabilize the heteroscedasticity of the data, the

techniques that has been proposed by Box [6] is the log-transformation of the response variable. But in practice, a single data transformation fails to satisfy the various model assumptions and thus variance may not be always positive [2; Table 2.7, p. 36].

*Joint Generalized Linear Model (JGLM):*

Distributions those are useful for the analysis of some continuous positive characteristic of the dataset, have non-Normal error distributions and are usually incorporates within

the class of GLM. The exponential and gamma are the such distributions which are often useful for modeling the positive data that have variance with mean relationship and the variance of the response variable is non-constant [4, 5]. Here we describe gamma JGLM, which is used to analyze the HCV dataset. Nelder and Lee [7] have suggested JGLMs for the mean and dispersion in case of heteroscedastic positive data  $y_i$ 's. A brief overview of the JGLM is given here:

$$E(y_i) = \mu_i \text{ and } \text{var}(y_i) = \sigma_i^2 V(\mu_i)$$

$V(\cdot)$  is the variance function and  $\sigma_i^2$ 's are the dispersion parameters. There are two components of variance in GLMs, one is  $V(\mu_i)$  which depends on mean function, and the other is  $\sigma_i^2$ , which is independent of mean function. The variance function  $V(\mu)$  specifies the distribution of GLM family i.e., if  $V(\mu)=1$  then the distribution is Normal, if  $V(\mu)=\mu$  then the distribution is Poisson and when  $V(\mu)=\mu^2$  the distribution is Gamma etc. The JGLM for the mean and dispersion parameters are

$$\eta_i = g(\mu_i) = x_i^t \beta \text{ and } \varepsilon_i = h(\sigma_i^2) = w_i^t \gamma,$$

where  $g(\cdot)$  and  $h(\cdot)$  are GLM link functions for the mean and the dispersion respectively; and  $x_i^t$ ,  $w_i^t$  are the row vectors for the regression models of mean and dispersion respectively, based on the levels of control variables. Maximum likelihood (ML) method is used to estimate the mean model parameters, and Restricted ML (REML) estimators are used for the dispersion model [8, 9]. An extensive discussion on GLM approach is found in [10-13].

#### Generalized Additive Model:

A more general and advanced regression technique GAM was developed and popularized by Hastie and Tibshirani in 1990 [14]. Schimek (2000) [15] and Wood (2006) [16] gave a comprehensive overview of the method and associated techniques, as well as the algorithms used to analyze the dataset.

GAM extends the linear and generalized linear models where the modeling of the mean functions overlooks the assumption of linearity. In spite of the fact, it is assumed that the additivity of the mean function belongs to the covariates. However, for some covariates mean functions may be considered to be linear, whereas non-linear mean functions are modelled with smoothing methods like kernel smoothers, lowess, smoothing splines, or regression splines. [3, 17]. In general, the model has the following structural form:

$$g(\mu) = \alpha_0 + \sum_{j=1}^p f_j(x_j)$$

Where,  $\mu=E(Y)$  for a response variable with some exponential family distribution,  $g(\cdot)$  is the link function and  $f_j$  are some smooth functions of the covariates  $x_j$  for each  $j=1, 2, \dots, p$ . GAM has greater stability than GLMs because they eliminate the linear dependency hypothesis between the

covariates and the expected value of the response variables. Estimation of the smooth function  $f_j$  increases the complexity of the GAM model and there are number of ways to address it. Quite possibly the most well-known choices depend on splines, which permit the GAM assessment to be decreased to the GLM context [18]. Smoothing splines [19] utilizes however many knots as unique values of the covariate  $x_j$  and control the model's smoothness by adding a penalty to the least squares fitting objective [19, 20].

### 3. Statistical Results and Graphical Analysis

In this article, two advanced regression modeling techniques have been carried out to study the impact of BIL along with the others cofactors to liver patients' diagnostic. Firstly, the response variable BIL is modeled through JGLM adopting the Gamma distribution. Apart from this, GAM is also adopted to identify the risk factors. The result of both the models are summarized in Tables 2 and 3 respectively. In both the JGLM and GAM model, some interaction terms along with main effects are considered. In Regression and Design of experiment, interaction terms are very much popular as it implies the cofactors have a joint influence on response variable. All the selected effects in a model are not necessarily always significant [22]. But sometimes insignificant effects are also retained in the model according to the marginality rule. In epidemiology, statistical insignificant included factors or variables in the fitted models are known as confounder [23]. For JGLM model, diagnostic plots for mean model and dispersion model are separately given in Figures 1 and 2 respectively. Figure 3 represents the diagnostic plot for GAM model.

#### 3.1. Results of Joint Generalized Linear Model (JGLM)

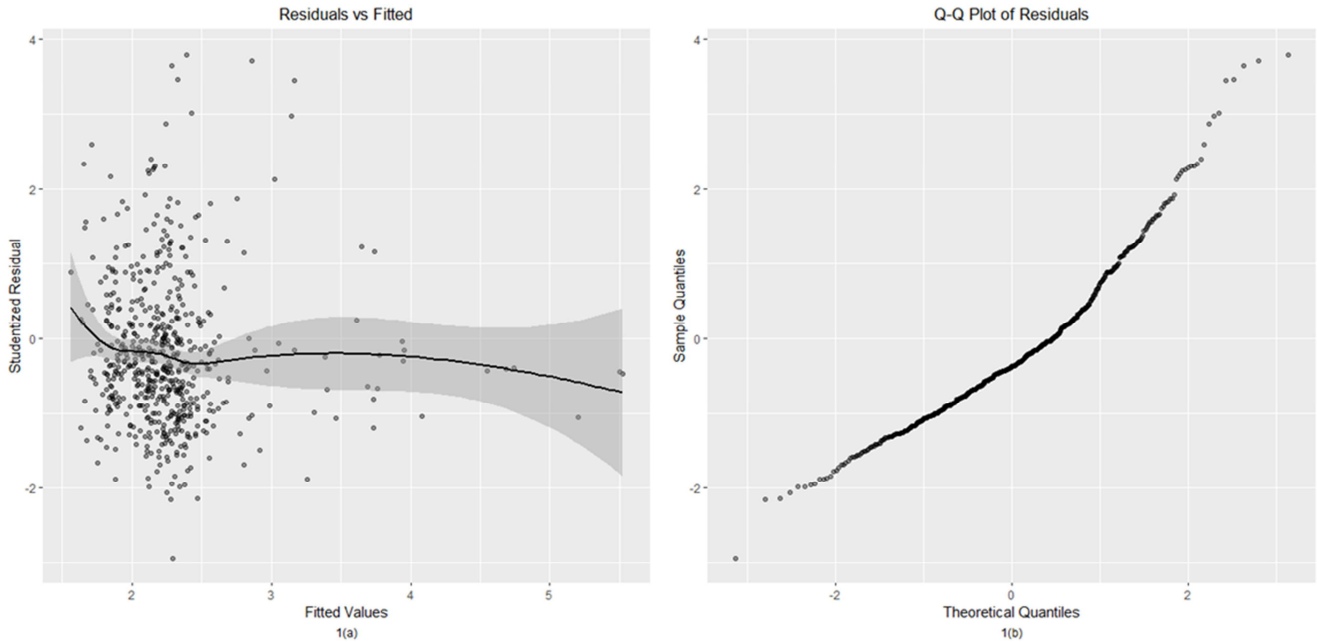
Log-Normal or Gamma models are generally used to analyze the positive data. Here the response variable BIL has been modeled using Gamma distribution through JGLM technique. The models will be finalized based on the smallest AIC. It is well known that the AIC selects a model which minimizes the predicted additive errors and squared error loss [7]. Also, the separate diagnostic plot for mean model and dispersion model are taken into account for model selection. Figure 1a and 2a represent the absolute residuals plot with respect to the fitted values of the Gamma models respectively. Both the figure displays the straight flat diagram, implying that the variance is constant with running means. Normal probability plot for fitted mean and dispersion model are displayed in Figure 1b and 2b respectively. From these figures, it can be interpreted that there is no lack of fit or departure from symmetry.

BIL Gamma fitted mean ( $\hat{\mu}$ ) model (from Table 2)

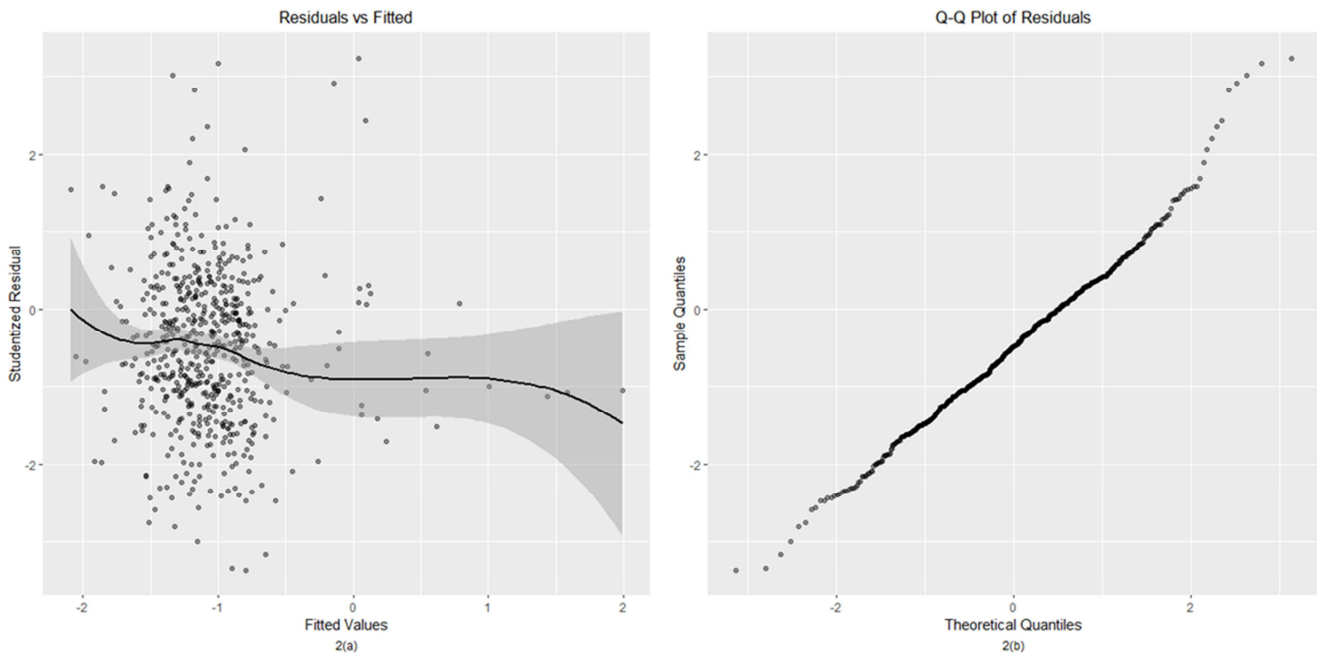
$$\hat{\mu} = \exp(2.721 - 0.294 \text{ SEX} - 0.0006 \text{ ALP} - 0.0046 \text{ ALT} + 0.0083 \text{ AST} - 0.174 \text{ CHE} - 0.2464 \text{ CHOL} + 0.0049 \text{ CREA} + 0.0048 \text{ GGT} + 0.016 \text{ PROT} + 0.0219 (\text{CHE} * \text{CHOL}) - 0.00004 (\text{CREA} * \text{GGT}))$$

and BIL Gamma fitted variance ( $\hat{\sigma}^2$ ) model (from Table 2)

$$\hat{\sigma}^2 = \exp(2.687 - 0.015 \text{ AGE} - 0.0245 \text{ ALP} - 0.0677 \text{ ALB} + 0.0079 \text{ AST} - 0.105 \text{ CHE} + 0.0007 (\text{ALP} * \text{ALB}))$$



**Figure 1.** Diagnostic plot for the joint Gamma fitted Mean models of BMI. 1 (a) Absolute residuals plot with respect to the fitted values. 1 (b) the normal probability plot of Residuals.



**Figure 2.** Diagnostic plot for the joint Gamma fitted Dispersion models of BMI. 2 (a) Absolute residuals plot with respect to the fitted values. 2 (b) Normal Q-Q plot of Residuals.

**Table 2.** Mean and Variance models of Gamma fit results of HCV Dataset.

Model	Covariate	Estimate	Standard Error	t-value	p-value
Mean Model	Constant	2.72141	0.73358	3.70976	0.000207
	Sex	-0.29454	0.07176	-4.10431	0.000040
	ALP	-0.00068	0.00159	-0.42885	0.668027
	ALT	-0.00460	0.00196	-2.34805	0.018871
	AST	0.008354	0.00187	4.45103	0.0000085
	CHE	-0.17452	0.06397	-2.72805	0.0063708

Model	Covariate	Estimate	Standard Error	t-value	p-value
Dispersion Model	CHOL	-0.24646	0.10987	-2.24314	0.02488
	CREA	0.00498	0.00222	2.23746	0.02525
	GGT	0.00487	0.00207	2.35455	0.01854
	PROT	0.01618	0.00676	2.39221	0.01674
	CHE*CHOL	0.02197	0.01164	1.88698	0.05916
	CREA*GGT	-0.000044	0.000020	-2.19677	0.02803
	Constant	2.68725	1.10428	2.43347	0.02085
	AGE	-0.01545	0.00597	-2.58448	0.01433
	ALP	-0.02456	0.01247	-1.96814	0.05767
	ALB	-0.06777	0.02643	-2.56374	0.01510
	AST	0.00792	0.00174	4.54851	0.00002
	CHE	-0.10509	0.02877	-3.65187	0.00054
	ALP*ALB	0.000735	0.00033	2.22098	0.03406

In Table 2, summarized forms of the obtained Gamma fitted mean and variance models of BIL are given. Therefore, the following interpretations are drawn based on the Gamma models.

- 1) The mean BIL (MBIL) is negatively associated with sex (Male=1; Female=2) with p-value < 0.001, concluding that BIL is higher for male than female.
- 2) ALT is negatively associated with MBIL with p-value 0.018, which indicates that if ALT level in the blood increases BIL decreases.
- 3) MBIL and AST have high positive significant association between them with p-value < 0.001. It implies that if AST increases in Bloodstream BIL also increases. This result supports the real-life situation also.
- 4) MBIL is negatively associated with CHE with p-value 0.006, indicating that BIL rises as CHE decreases.
- 5) MBIL is negatively associated with CHOL having p-value 0.024, implying that BIL increases in Bloodstream as CHOL decreases.
- 6) MBIL is positively associated with the interaction effect (CHE\*CHOL) with p-value 0.059, interpreting that BIL may rise in Blood cells, as (CHE\*CHOL) jointly increases. Here one thing is to be noted that, though the marginal effect of CHE and CHOL are negatively associated with BIL, the joint effects of these two cofactors have some positive significant association.
- 7) CREA have positive association with MBIL with p-value 0.025. It means that if CREA increases BIL level also increases.
- 8) MBIL have positively associated with the cofactor GGT with p-value 0.018, implying that BIL rises as GGT increases.
- 9) MBIL is negatively associated with the interaction effect (CREA\*GGT) with p-value 0.02, indicating that BIL increases as (CREA\*GGT) decreases. Note that marginal effects of CREA and GGT are positively associated with mean BIL, while their interaction effect is negatively associated with BIL. This situation implies that CREA and GGT increases BIL level, but their joint effect decreases BIL.
- 10) MBIL have positive significant association with PROT with p-value 0.016, implying that BIL increases as

PROT increases.

- 11) Variance of BIL (VBIL) is negatively associated with Age with p-value 0.0143, interpreting that BIL Variance is higher at younger patients than older.
- 12) VBIL have negative association with ALP with p-value 0.0576, indicates that BIL variance increases as ALP decreases.
- 13) VBIL is negatively associated with ALB with p-value 0.015, indicates that BIL variance increases as ALB decreases.
- 14) VBIL have very high positive significant association with AST, with p-value < 0.001. That implies that BIL variance rises as AST increases.
- 15) The cofactor CHE have highly negative significant association with VBIL with p-value < 0.001. It indicates that BIL variance increases when CHE also increase in Bloodstream.
- 16) VBIL is positively and significantly associated with the interaction effect (ALP\*ALB) with p-value 0.034, interpreting that variance of BIL increases as (ALP\*ALB) increases. Although the marginal effects ALP and ALB are both negatively associated, still the joint effects of the cofactors are positive association with variance of BIL.

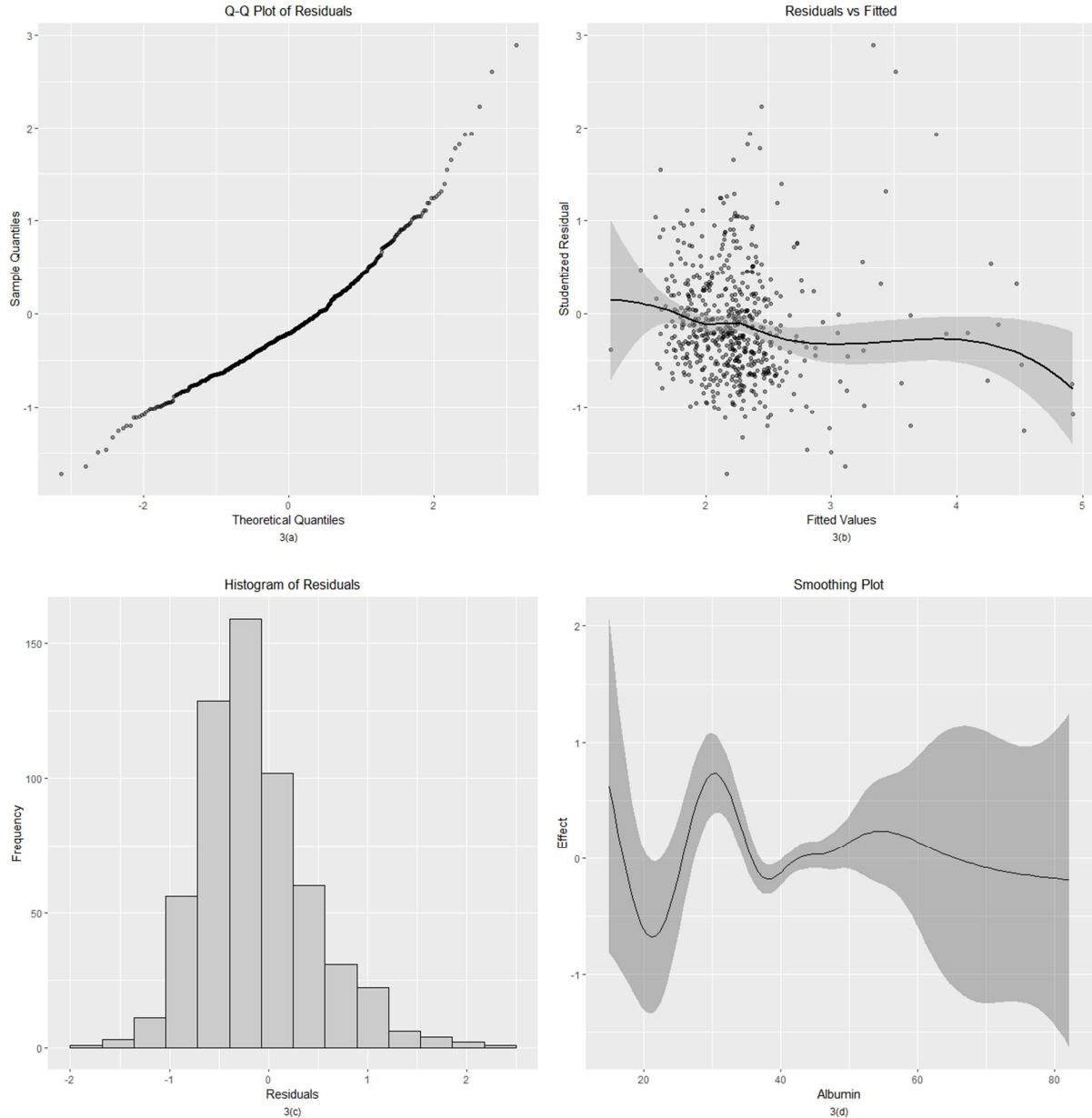
### 3.2. Results of Generalized Additive Model (GAM)

The response variable BIL has been modeled through GAM using Gamma distribution and logarithm link. GCV value along with different model diagnostic plot has been used for choosing the best GAM model. GAM consists of two part of estimation methods, namely Parametric estimation for those cofactors enter in the model parametrically and Non-parametric estimation used for smoothing cofactors. Heterogeneity and the non-linear relationship between response variable and the other cofactors are handled through this Non-parametric estimation part of GAM. This property gives the GAM models more flexible among the others. Both of these part of estimation results is summarized in Table 3. Besides the main effects, two second order interaction effects and one smoothing cofactor have to be included in the final model to point out the true relationship between BIL and other cofactors.

To represent the above results graphically, four different diagnostic plots are also been examined. In Figure 3a,

Normal probability plot shows the theoretical quantiles which are plotted against deviance residuals. Figure 3b displays the residuals values which are plotted against the fitted values of GAM model. It is almost a flat diagram with the running means. This residual plot is the indication of the constant variance with the respective means. Histogram of

the residuals are plotted in Figure 3c, reveals that the residuals are normally distributed. Figure 3d represents the smoothing factor ALB with confidence belt. It shows that after some level, linearity has been attained with respect to its smoothness.



**Figure 3.** Regression diagnostic plot of GAM for HCV dataset. 3 (a) Normal Q-Q Plot of Residual. 3 (b) Residual vs Fitted plot. 3 (c) Histogram of Residuals. 3 (d) Plot of Albumin as a smoothing term.

**Table 3.** Results for GAM of HCV data analysis using Gamma distribution with 'log' link.

Estimation of Parametric Coefficients				
Covariates	Estimate	Standard Error	F-Value	Pr(> t )
Constant (Intercept)	3.473	0.7852	4.423	0.0000117***
Sex	-0.3033	0.06963	-4.356	0.0000157***
ALT	-0.003628	0.001674	-2.167	0.030658 *
AST	0.004522	0.001178	3.838	0.000138 ***
CHE	-0.2150	0.06352	-3.385	0.000760 ***
CHOL	-0.2713	0.1052	-2.579	0.010151 *
CREA	0.003861	0.001842	2.096	0.036543 *

Estimation of Parametric Coefficients				
Covariates	Estimate	Standard Error	F-Value	Pr(> t )
GGT	0.004484	0.001520	2.949	0.003321 **
PROT	0.0114	0.007773	1.433	0.152330
CHE*CHOL	0.02678	0.01145	2.338	0.019722 *
CREA*GGT	0.00003814	0.00001608	-2.372	0.018036 *
Approximate Significance of smooth terms (Non-Parametric)				
Smooth Covariate	Edf	Ref. df	Chi.sq	p-value
ALB	8.655	8.964	3.365	0.000337

Edf: Estimated degrees of freedom; Ref.df: Degrees of freedom before smoothing; Chi. Sq: Chi square value. Significance Level: '\*\*\*' 0.001; '\*\*' 0.01; '\*' 0.05; '.' 0.1.

AIC=3594.69; Deviance Explained: 44.7%; GCV Score: 0.39336; Scale Estimate: 0.5417; n=589

From Table 3, final selected GAM fitted model of the BIL is shown below:

$$\text{Log (BIL)} = 3.473 - 0.303 \text{ SEX} - 0.003 \text{ ALT} + 0.004 \text{ AST} - 0.215 \text{ CHE} - 0.271 \text{ CHOL} + \text{CREA} 0.003 + \text{GGT} 0.004 + 0.0114 \text{ PROT} + 0.026 (\text{CHE} * \text{CHOL}) + 0.000038 (\text{CREA} * \text{GGT}) + f(\text{ALB})$$

In the above formula, except ALB, all the cofactors entered in this additive model parametrically. The only smoothing expression whose approximate significance is determined using a non-parametrical approach (Chi-square test) is ALB.

#### Results of estimation of Parametric coefficients:

From Table 3, the results and interpretation of the cofactors for parametric estimation are described below:

- (1) The factor Sex (Male=1; Female=2) has very high negative significant relationship with BIL with p-value < 0.001. It indicates that BIL is higher for male than female.
- (2) BIL has partial negative significant association with ALT (test measures the amount of this enzyme in the blood), having p-value 0.03 which means that if the ALT level increases then the BIL value decreases. In real phenomena, also lower than normal levels of bilirubin aren't a problem but higher level of ALT in blood indicates liver diseases or a viral hepatitis infection.
- (3) AST and BIL have very high positive significant association between each other having p-value < 0.001. If the value of AST increases BIL also increases. Both the situation is the indication of some liver damage may be due to the hepatitis or cirrhosis.
- (4) CHE (an enzyme required for function in the nervous system and is responsible for breaking down acetylcholine) has high negative significant association with BIL having p-value < 0.001. If the value of CHE increases BIL decreases.
- (5) BIL has partial negative significant association with CHOL having p-value 0.01. If Cholesterol level decreases BIL may also increase. This result also supports the real phenomena.
- (6) In this GAM fitted model although the marginal effects of CHE and CHOL both are negatively associated with BIL but their interaction effect (CHE\*CHOL) have positive association with BIL, where p-value is 0.019. So, if both CHE and CHOL increase then BIL also increases.
- (7) CREA has partial positive association with BIL with p-

value 0.036. Implying that if CREA increases, BIL also increases.

- (8) The factor GGT have some high positive association with BIL having p-value < 0.001. It means that if BIL increases then GGT level also increases. This result may suggest that there have some complications in the liver function.
- (9) The interaction effect (CREA\*GGT) is positively significantly associated with BIL with p-value 0.018.

#### Results of Non-parametric estimation for approximate significance of Smooth term:

In this GAM fitted model, ALB is the only smoothing cofactor. To test the hypothesis, Chi-square test statistic is used as it is a nonparametric method of estimation. From Table 3, it is observed that smoothness of the cofactor ALB is highly significant with p-value 0.0003. Also observed that, the GAM fitted model has an AIC value 3594.69. GCV score and scale estimate are 0.3933 and 0.5417 respectively, which are also very low as compared to the other models.

## 4. Discussion

In this article, association between BIL with other liver disease biomarker from Hepatitis C patient has been studied. Bilirubin, a biochemical parameter responsible for liver infection, is treated here as a response variable. We tried to model this BIL variable which is a continuous random variable with non-constant variance and non-normal distribution pattern. To model this, we use two statistical modelling techniques namely JGLM and GAM with the assumption of Gamma distribution and logarithm as a link function. The variable description and the fitted results are presented in Table 1, 2 and 3 respectively. The model checking plots such as normal probability plot, absolute residual plots are presented in Figures 1, 2 and 3 respectively.

In JGLM model, mean model of BIL is expressed by Sex, ALP, ALT, AST, CHE, CHOL, CREA, GGT, PROT, (CHE\*CHOL), (CREA\*GGT) whereas variance of BIL is expressed by Age, ALP, ALB, AST, CHE, (ALP\*ALB). Mean model is more complicated with comparison to dispersion model as it contains two interaction effects. Here

things to be noted that the natures of marginal effect of CHE and CHOL (negative), CREA and GGT (positive) in mean model and ALP and ALB (negative) in dispersion model are different from their respective interaction effect (CHE\*CHOL) (positive), (CREA\*GGT) (negative) in mean model whereas (ALP\*ALB) (positive) in dispersion model. Also, ALP is insignificant in mean model while it is partially significant (p-value 0.05) in dispersion model. Age is significant (p-value 0.014) cofactor in dispersion model but it is not included for explanation of mean model. PROT is a significant cofactor (p-value 0.016) in the mean model but it is not used in dispersion model. AST has high positive correlation (p-value < 0.001) with BIL both in mean and dispersion model. Earlier research also supports this result [24]. Similarly, there is also high negative correlation (p-value < 0.001) between CHE and BIL.

In case of GAM model, the relationship of BIL with other biochemical factors as explained considering Sex, ALT, AST, CHE, CHOL, CREA, GGT, PROT, (CHE\*CHOL), (CREA\*GGT) and one smoothing cofactor ALB. Interaction effect (CHE\*CHOL) are positive in nature while the marginal effects are opposite in nature, i.e., negatively associated. On the other hand, both marginal effect of CREA and GGT and interaction effect (CREA\*GGT) have identical role on BIL; i.e., both are positively correlated with BIL. PROT has no significant effect on BIL, which is also found in earlier research work [24]. In this model, ALB having a positive significant association with BIL and this report suggests that this cofactor having some nonlinear associations with BIL.

From all the models it is conclusive that AST and CHE, both the enzymes are produced by liver have high significant impact on Bilirubin level in bloodstream. One has positive impact (AST) and other (CHE) is negatively associated. Best of our knowledge, there are a few research articles examine the association of BIL with other liver Biomarker of Hepatitis C patient, simultaneously with JGLM and GAM in regression framework.

## 5. Conclusion

The association and effects of several liver biomarker of Hepatitis C patients have been clearly explained throughout this article. The current results reported in Tables 2 and 3 are not completely conclusive rather acquaintance. The developed relationship of BIL with other cofactors are based on following regression analysis criteria: (1) The determinants are selected based on JGLMs and GAM fitted model assuming the response variable follows the Gamma distribution, which is appropriate for modeling positive and continuous data. (2) Final model is selected according to the smallest AIC and in case of GAM model, smallest GCV value is the criterion. (3) Regression diagnostic plot of JGLM and GAM are verified for final models. (4) From Tables 2 and 3, it is observed that the standard error of the estimates is very small and it is the indication that the explanatory factors of BIL are stable [21].

Therefore, the current report has some stability and it may expect that it will help the medical practitioners and the researchers in this field to know the mathematical relationship between the different Biomarkers of the hepatitis C patient. The current result is focused on many interesting conclusions. The present analysis includes the liver Biomarker CHE and it is significantly associated with BIL. Also, the high level of ALT and AST is the indication of liver cirrhosis.

## Abbreviations

JGLM: Joint Generalized Linear Model, GAM: Generalized Additive Model, AST: Aspartate Aminotransferase, CREA: Creatinine, GGT: Gamma-Glutamyl Transpeptidase, CHE: Choline Esterase, CHOL: Cholesterol, ALT: Alanine Aminotransferase, ALB: Albumin, PROT: Protein, ALP: Alkaline Phosphatase, BIL: Bilirubin, MBIL: Mean Bilirubin, VBIL: Variance Bilirubin, HCV: Hepatitis C Virus, GCV: Generalized Cross-Validation, AIC: Akaike Information Criterion.

## Conflicts of Interest

The Authors have no conflict of interest to declare.

## References

- [1] McCullagh P, Nelder JA. Generalized linear models. London: Chapman & Hall, 1989.
- [2] Myers RH, Montgomery DC, Vining GG. Generalized linear models with applications in engineering and the sciences. New York: John Wiley & Sons, 2002.
- [3] Mukherjee S. Statistical Modeling for Findings the Determinants of Neonates' Very Low Birth Weight. *American Journal of Clinical and Experimental Medicine* 2018; 6: 125-135.
- [4] Das RN. Determinants of Diabetes Mellitus in the Pima Indian Mothers and Indian Medical Students. *The Open Diabetes Journal* 2014; 7: 5-13.
- [5] Das RN, Devi RS, Kim J. Mothers' Lifestyle Characteristics Impact on Her Neonates' Low Birth Weight. *International Journal of Women's Health and Reproduction Sciences* 2014; 2: 229-235.
- [6] Box GEP. Signal-to-noise ratios, performance criteria, and transformations (with discussion). *Technometrics* 1988; 30: 1-40.
- [7] Nelder JA, Lee Y. Generalized linear models for the analysis of Taguchi-type experiments. *Applied Stochastic Models and Data Analysis* 1991; 7: 107-120.
- [8] Lee Y, Nelder JA, Pawitan Y. Generalized Linear Models with Random Effects (Unified Analysis via H-likelihood). Chapman & Hall, London 2006.
- [9] Das RN, Lee Y. Log normal versus gamma models for analyzing data from quality-improvement experiments. *Qual Eng* 2009; 21: 79-87.

- [10] Lee Y, Nelder JA. Generalized linear models for the analysis of quality improvement experiments. *Can J Statist* 1998; 26: 95-105.
- [11] Das RN. Robust Response Surfaces, Regression, and Positive Data Analyses. Chapman & Hall, London 2014.
- [12] Qu Y, Tan M, Rybicki L. A unified approach to estimating association measures via a joint generalized linear model for paired binary data. *Commun Statist-Theory Methods* 2000; 29: 143-56.
- [13] Lesperance ML, Park S. GLMs for the analysis of robust designs with dynamic characteristics. *J Qual Tech* 2003; 35: 253-63.
- [14] Hastie T, Tibshirani R. Generalized Additive Models, first ed. London: Chapman and Hall 1990.
- [15] Schimek MG. Estimation and inference in partially linear models with smoothing splines. *Journal of Statistical Planning and Inference* 2000; 91: 525-540.
- [16] Wood SN. Generalized Additive Models: An Introduction with R. London: Chapman and Hall 2006.
- [17] Mukherjee S, Kapoor S, Banerjee P. Diagnosis and Identification of Risk Factors for Heart Disease Patients Using Generalized Additive Model and Data Mining Techniques. *J Cardiovasc Disease Res* 2017; 8: 137-44.
- [18] Currie I, Durban M, Eilers P. Generalized linear array models with applications to Multi-dimensional smoothing. *Journal of the Royal Statistical Society B* 2006; 68: 259-280.
- [19] Green P, Silverman B. Nonparametric Regression and Generalized Linear Models. London: Chapman & Hall 1994.
- [20] Ruppert D. Selecting the number of knots and for penalized splines. *Journal of Computational and Graphical Statistics* 2002; 11: 735-757.
- [21] Eilers P, Marx B. Flexible smoothing with B-splines and penalties. *StatSci* 1996; 11: 89-121.
- [22] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning, Springer-Verlag, 2001.
- [23] Nelder JA. The statistics of linear models: back to basics. *Statistics and Computing* 1994; 4: 221-234.
- [24] Das RN, Lee Y, Mukherjee S, et al. Relationship of Alanine Aminotransferase (SGPT) with other Liver Biomarkers. *Journal of Gastroenterology, Pancreatology & Liver Disorders* 2019; 7: 1-5.
- [25] Hoffmann G, Bietenbeck A, Lichtinghagen R, Klawonn F. Using machine learning techniques to generate laboratory diagnostic pathways—a case study. *Journal of Laboratory and Precision Medicine* 2018; 3: 58-67. doi: 10.21037/jlpm.2018.06.01.
- [26] Das RN, Lee Y, Sengupta S, et al; Complex Association of Albumin with Other Liver Biomarkers. *EC Gastroenterology and Digestive System* 2019; 6: 132-137.
- [27] Das RN. Determinants of total bilirubin for liver patients. *Journal of Liver Research, Disorders & Therapy* 2018; 4: 127-131.
- [28] Das RN, Mukherjee S, Sharma I. Alkaline Phosphatase Determinants of Liver Patients. *Journal of the Pancreas* 2018; 19: 18-23.
- [29] Smith MW. Hepatitis C and the Hep C Virus. [Online] <https://www.webmd.com/hepatitis/digestive-diseases-hepatitis-c> (2019, accessed 20 February 2021).
- [30] Center for Diseases control and prevention. Viral Hepatitis [Online]. <https://www.cdc.gov/hepatitis/hcv/index.html> (accessed 18 February 2021).
- [31] Yoshimasu Y, Furuichi Y, Kasai Y, et al. Predictive Factors for Hepatocellular Carcinoma Occurrence or Recurrence after Direct-Acting Antiviral Agents in Patients with Chronic Hepatitis C. *Journal of Gastrointestinal and Liver Diseases* 2019; 28: 63-71.
- [32] Pohl A, Behling C, Oliver D, et al. Serum Aminotransferase Levels and Platelet Counts as Predictors of Degree of Fibrosis in Chronic Hepatitis C Virus Infection. *The American Journal of Gastroenterology* 2001; 96: 3142-6.
- [33] Williams ALB, Hoofnagle JH. Ratio of Serum Aspartate to Alanine Aminotransferase in Chronic Hepatitis Relationship to Cirrhosis. *Gastroenterology*. 1988; 95: 734-739.
- [34] Hepatitis C. [Online]: <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c#:~:text=Hepatitis%20C%20is%20a%20liver,major%20cause%20of%20liver%20cancer> (2020, accessed 20 February 2021).