American Journal of Clinical and Experimental Medicine
2018; 6(6): 125-135
http://www.sciencepublishinggroup.com/j/ajcem
doi: 10.11648/j.ajcem.20180606.11
ISSN: 2330-8125 (Print); ISSN: 2330-8133 (Online)

Statistical Modeling for Findings the Determinants of Neonates' Very Low Birth Weight

Sabyasachi Mukherjee

Department of Mathematics, NSHM Knowledge Campus, Durgapur, India

Email address:
sabyasachi99@gmail.com

To cite this article:
Sabyasachi Mukherjee. Statistical Modeling for Findings the Determinants of Neonates' Very Low Birth Weight. *American Journal of Clinical and Experimental Medicine*. Vol. 6, No. 6, 2018, pp. 125-135. doi: 10.11648/j.ajcem.20180606.11

Received: November 20, 2018; **Accepted**: December 8, 2018; **Published**: January 14, 2019

Abstract: Aim of the epidemiological research is to identify a causal relationship between the risk factors and the disease. The present study aims to identify the determinates of neonates' very low birth weight which have significant effects on their birth weight using generalized additive probabilistic modeling. In this present study very low birth weight (BWT) is the response variable with heterogeneity and non-normality in nature which can be modeled through either by the Log-normal or by the gamma models. A well-known method is joint modeling of mean and variance (JGLM) to handle this heterogeneity and non-normality but this study introduced the most advanced regression techniques namely generalized additive model (GAM). Materials and Methods: The present article is based on the secondary data on 174 very low birth weighted neonates along with 26 explanatory factors/ variables. The very low birth weight of 174 neonates is heterogeneous, positive, and gamma distributed. Therefore, generalized additive model with gamma distribution and log link function has been introduced to analyze this very low birth weight. The Native American neonates has the smaller birth weight (BWT) than the other racial neonates namely black, white or oriental (P-value = 0.009). The BWT is smaller for those neonates who were outsiders from Duke (P-value = 0.05) and it also decreases to the female births (P- value=0.09). BWT is higher for the pneumothorax infants than non-pneumothorax with P-value <0.001. It is smaller for those infants for whom oxygen supply had been introduced in between 30 days of his/her birth (P-value=0.002). The neonates who were not been survived they also have the smaller birth weight than the alive neonates (P-value= 0.07). Besides these factors, lowest pH in first 4 days of neonates' life (P-value=0.09), Apgar score at one minute (P-value=0.03) and the interaction effects of lowest pH and Apgar score (0.03) are the significant variables (continuous cofactors) for neonates' very low birth weight. Hospital stays in number of days, platelet counts and gestational age in weeks are also highly significant factors and these are entered in the GAM model as a non – parametric smoothing term (P-value<0.001). The birth weight of each new born baby is identified as heterogeneous and gamma distributed (possible). Most of the present findings, especially the Apgar score in one minute, occurrence of pneumothorax in neonates, oxygen supply given to new born baby within 30 days of life and smoothing term determinants namely number of days hospital stay, platelet counts and gestational age in weeks (non-parametric smoothing terms) are significant factors for neonate's birth weight and are completely new in the literature.

Keywords: Very Low Birth Weight, Generalized Additive Model, Gamma Distribution, Non-constant Variance

1. Introduction

Neonatal death is a serious concern, both in the developing and in the developed countries. While infant mortality rates have been decreasing steadily all over the world, changes in neonatal mortality rate have been much slower. One of the commonest causes of neonatal mortality in the world is prematurity and low birth weight (LBW) [1-4]. Generally, it is recognized that low birth weight can be caused by many factors [5-7]. Neonate low birth weight has long been a subject of clinical and epidemiological investigations and a target for public health intervention. Low birth weight is defined by WHO as a birth weight less than 2500 g (before 1976, the WHO definition was less than or equal to 2500 g), since below this value birth-weight-specific infant mortality begins to rise rapidly [2, 8-11].

Across the world, neonatal mortality is 20 times more likely for LBW babies compared to normal birth weight

(NBW) babies (>2.5 kg) [12]. It is now a well-recognized fact that birth weight is not only a critical determinant of child survival, growth, and development, but also a valuable indicator of maternal health, nutrition, and quality of life [13].

The incidence of LBW is estimated to be 16% worldwide, 19% in the least developed and developing countries, and 7% in the developed countries. The incidence of LBW is 31% in South Asia followed by East and North Africa (15%), Sub-Saharan Africa (14%), and East Asia and Pacific (7%). Asia accounts for 75% of worldwide LBW followed by Africa (20%) and Latin America (5%).

Another important issue that should to be considered here is very low birth weight (VLBW) of neonates, which emphasizes birth weight of a new born baby having less than 1500 grams. In a developed or developing countries neonatal deaths due to VLBW is very rare but still in worldwide (mostly in under developed countries) it has a significant figure. But the main concern is about the mortality rate of VLBW babies, because it has almost 10 times higher mortality rate than the LBW babies.

The main cause of a baby having VLBW is being born too early. This is called preterm or premature birth. Premature means a baby is born before 37 gestational weeks of pregnancy. Very-low-birth-weight babies are often born before 30 weeks of pregnancy. A premature baby has less time in the mother's uterus to grow and gain weight. Much of a baby's weight is gained during the later part of pregnancy. Another cause of very low birth weight is when a baby does not grow well during pregnancy. This is called intrauterine growth restriction (IUGR). It may happen because of problems with the placenta, the mother's health, or birth defects. Most very low birth weight babies who have IUGR are also born early. They are usually very small and physically immature.

There are so many complications arise regarding VLBW which can be described briefly here. Babies with a very low birth weight have a greater risk of developing problems. Their tiny bodies are not as strong as babies of normal weight. They may have a harder time eating, gaining weight, and fighting infection. They have very little body fat. So they often have trouble staying warm in normal temperatures.

Most babies with a very low birth weight are also premature. This can make it hard to separate the problems caused by the prematurity from the problems of just being so tiny. In general, the lower the baby's birth weight, the greater the risks for complications. Here are some of the most common problems of babies with VLBW – (i) Low oxygen levels at birth (ii) Trouble staying warm, feeding and gaining weight (iii) Infection (iv) Breathing problems because of immature lungs (respiratory distress syndrome) (v) Nervous system problems, such as bleeding inside the brain or damage to the brain's white matter (vi) Serious digestive problems, such as necrotizing enterocolitis and (vii) Sudden infant death syndrome (SIDS). *(https://www.urmc.rochester.edu/encyclopedia/content.aspx?ContentTypeID=90&ContentID=P02424).*

In particular, considerable attention has been focused on the causal determinants of birth weight, and especially of low birth weight (LBW), in order to identify potentially modifiable factors. Many researches have focused on factors with well-established direct causal impacts on intrauterine growth include infant sex, racial/ethnic origin, maternal height, pre-pregnancy weight, paternal weight and height, maternal birth weight, parity, history of prior low-birth-weight infants, gestational weight gain and caloric intake, general morbidity and episodic illness, malaria, cigarette smoking, alcohol consumption, and tobacco chewing [1, 7, 10]. Many research works has been done for very low birth weight of neonates [14-15, 19-21] and mainly these research works are based on the mortality rate of that neonates. Note that these factors were identified based on preliminary statistical methods such as frequency distribution, odds ratio, simple regression analysis, logistic regression etc. These methods may not identify the determinants correctly in medical systems, demography and quality engineering process, as the variance of the response may be non-constant, and the variance may have some relationship with the mean [16- 22]. Generally, the above methods identify insignificant factors as significant and vice versa, which is a serious error in any data analysis. This present article tried to find the main determinates who are directly responsible for this very low birth weight through advanced statistical analysis.

In the statistical analysis of clinical trials and observational studies, the identification and adjustment of prognostic factors is an important activity in order to get valid outcome. The failure to consider important prognostic variables, particularly in observational studies, can lead to errors in estimating treatment differences. In addition, incorrect modeling of prognostic factors can result in the failure to identify nonlinear trends or threshold effects on survival. This article describes how 'Generalized Additive Models' (GAM) [33], a flexible statistical method has been used to identify and characterize the effect of potential prognostic factors on disease endpoints.

2. Materials & Statistical Methodology

2.1. Materials

The present article is based on the secondary data on 671 very low birth weighted neonates along with 26 explanatory factors/ variables. This data of 671 infants with very low (<1600 grams) birth weight from 1981-87 were collected at Duke University Medical Center by Dr. Michael O'Shea, now of Bowman Gray Medical Center [15]. The data set can be downloaded from http://biostat.mc.vanderbilt.edu/DataSets". The description of the covariates, factors and their levels are described in Table 1. The summarized statistics such as the mean, standard deviation, and proportion of the levels are given in Table 1. In the main dataset, there are 671 neonates with many missing information on the 26 factors /variable. We have considered only 174 neonates (from 671 neonates) with all non-missing information on the 26 factors/ variables. The description of the patient population and the data

collection method is not reproduced herein as the length of the paper will be increased.

The current data contains 10 continuous variables (2 variables date of birth and death are not been used in the model) and 16 attribute characters. The description of each variable or attribute character, attribute levels, and how they are operationalized in the present report is displayed in Table 1. Here we have considered the birth weight of new born baby as the dependent variable, and the remaining others are treated as the independent or explanatory variables.

Table 1. *Operationalization of variables with the analysis & summarized statistics.*

Variable name	Operationalization	Mean	Standard deviation	Proportion (Percentage %)
birth	Date of birth (admission) (Not used in model)	85.98	0.78	----
exit	Date of death or discharge (Not used in model)	86.11	0.79	----
hospstay	Hospital stays in number of days	51.95	43.6	----
lowph	Lowest pH in first 4 days of life	7.23	0.12	----
pltct	Platelet count (*10^9/L)	200.99	84.11	----
race	Black=1; White=2; Native American=3; Oriental=4	----	----	1=56.32%; 2=41.38%; 3=1.72%; 4=0.52%
bwt	Birth weight in gram	1112.6	241.9	----
gest	Gestational age in weeks	29.28	2.36	----
inout	Born at Duke or transported, Duke=1; Transported=2	----	----	1=97.70%; 2=2.30%
twn	Multiple gestation, No=1, Yes=2	----	----	1=75.86%; 2=24.14%
lol	Duration of labor in hours	6.75	16.78	----
magsulf	Mother treated with $MgSO_4$. No=1, Yes=2	----	----	1=82.76%; 2=17.24%
meth	Mother treated with beta-methasone. No=1, Yes=2	----	----	1=50.57%; 2=49.43%
toc	Tocolysis - mother treated with beta-adrenergic drug. No=1, Yes=2	----	----	1=74.71%; 2=25.29%
delivery	Abdominal and Vaginal; Abdominal=1, Vaginal=2			1=59.20%; 2=40.80%
apg1	Apgar (The Apgar score, the very first test given to a newborn) at one minute	5.10	2.69	----
vent	Assisted ventilation used. No=1, Yes=2	----	----	1=43.10%; 2=56.90%
pneumo	Pneumothorax occurred No=1, Yes=2	----	----	1=86.78%; 2=13.22%
pda	Patent ductus arteriosus detected No=1, Yes=2	----	----	1=73.56%; 2=26.44%
cld	On suppl. oxygen at 30 days No=1, Yes=2	----	----	1=75.29%; 2=24.71%
pvh	Periventricular hemorrhage; Absent=1, Definite=2, Possible=3	----	----	1=64.37%; 2=24.14%; 3=11.49%
ivh	Intraventricular hemorrhage; Absent=1, Definite=2, Possible=3	----	----	1=89.08%; 2=10.34%; 3=0.57%
ipe	Periventricular intraparenchymal echodense lesion; Absent=1, Definite=2, Possible=3	----	----	1=87.93%; 2=5.75%; 3=6.32%
year	Study year (Not used in model)	85.98	0.78	
sex	Gender: Male= 1; Female=2	----	----	1=50.57%; 2=49.43%
dead	Live status; Live=1, Dead=2	----	----	1=93.1%; 2=6.9%

2.2. Statistical Methods

The classical linear regression models assume that the response (Y) variance is constant over the entire range of parameter values. However, it is not always true [16-17]. Generally, medical science data are heterogeneous. For example, the birth weights of the new born babies are heterogeneous [14, 19]. To stabilize the heteroscedasticity of a data set, the log-transformation is often recommended, but in practice the variance may not always be stabilized [17; Table 2.7, p. 36].

For the analysis of positive observations with constant variance or constant coefficient of variation, the linear regression models with multiplicative error estimation is performed based on either the gamma or the Log-normal models [18]. However, for the medical science data analysis, neither the variance nor the coefficient of variation needs to be constant, so that these two models do not necessarily give identical results [19-21]. Note that the generalized linear models class includes the distributions which are used in modeling some continuous, non-normal, positive, and heteroscedastic data sets. In the generalized linear models class, the variance of the response may have the relationship with its mean. Then the response variance may be non-constant. In order to analyze the heteroscedastic positive data Y_i *'s,* Nelder and Lee [27] have suggested to use the joint generalized linear models (JGLMs). A detailed discussion of JGLMs is given in [23-25]. Application of JGLMs in medical science data analysis is given in [18-21, 31]. This present article introduced more generalized and flexible method of advanced regression techniques known as generalized additive model (GAM). The main advantage of GAM over JGLMs is that, in JGLMs we have to consider two interconnected models one is for mean and another for dispersion but in GAM only one mean model is enough to control the heteroscedasticity in data. The methods available in Generalized Additive Models are implementations of techniques developed and popularized by Hastie and Tibshirani (1990) [32]. A detailed description of these and related techniques, the algorithms used to fit these models,

and discussions of recent research in this area of statistical modeling can also be found in Schimek (2000) [31]. For ready reference, a short description of GAM is presented in this section.

2.3. Generalized Additive Model (GAM)

GAM [32, 34] is an extension of the Generalized Linear Model (GLM) [17] where the modeling of the mean functions relaxes the assumption of linearity, albeit additivity of the mean function pertaining to the covariates is assumed. Whilst the mean functions of some covariates may be assumed to be linear, the non-linear mean functions are modeled using smoothing methods, such as kernel smoothers, lowess, smoothing splines or regression splines. In general, the model has the following structure

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

where, $\mu = E(Y)$ for Y, a response variable with some exponential family distribution, g is the link function and f_j are some smooth functions of the covariates X_j for each $j = 1,2,\ldots\ldots,p$.

GAMs provide more flexibility than do GLMs, as they relax the hypothesis of linear dependence between the covariates and the expected value of the response variable. The main drawback of GAMs lies in the estimation of the smooth functions f_j, and there are different ways to address this. One of the most common alternatives is based on splines, which allow the GAM estimation to be reduced to the GLM context [35]. Smoothing splines [36], use as 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 [37, 38].

Generalized additive models can be used in virtually any setting where linear models are used. For a single observation (i^{th}) the basic idea is to replace $\sum_{j=1}^{p} x_{ij}\beta_j$, the linear component of the model with an additive component $\sum_{j=1}^{p} f_j(x_{ij})$. In other words, the purpose of generalized additive models is to maximize the quality of prediction of the dependent variable Y from various distributions, by estimating unspecific (non-parametric) functions of the covariates X_j which are "connected" to the dependent variable via the link function g.

A unique aspect of generalized additive models is the non-parametric functions f_j of the covariates X_j. Specifically, instead of some kind of simple or complex parametric functions, Hastie and Tibshirani (1990) discuss various general scatterplot smoothers that can be applied to the X variable values, with the target criterion to maximize the quality of prediction of the (transformed) Y variable values. One such scatterplot smoother is the cubic smoothing splines smoother, which generally produces a smooth generalization of the relationship between the two variables in the scatterplot. Computational details regarding this smoother can be found in Hastie and Tibshirani (1990; see also Schimek, 2000) [31-32].

To summarize, instead of estimating single parameters (like the regression weights in multiple regression), in generalized additive models, we find a general unspecific (non-parametric) function that relates the predicted (transformed) Y values to the predictor values.

Detailed descriptions of how generalized additive models are fit to data can be found in Hastie and Tibshirani (1990) [32], as well as Schimek (2000, p. 300) [31]. In general there are two separate iterative operations involved in the algorithm, which are usually labeled the outer and inner loop. The purpose of the outer loop is to maximize the overall fit of the model, by minimizing the overall likelihood of the data given the model (similar to the maximum likelihood estimation procedures as described in, for example, the context of Nonlinear Estimation). The purpose of the inner loop is to refine the scatterplot smoother, which is the cubic spline smoother. The smoothing is performed with respect to the partial residuals; i.e., for every predictor k, the weighted cubic spline fit is found that best represents the relationship between variable k and the (partial) residuals computed by removing the effect of all other j predictors ($j \neq k$). The iterative estimation procedure will terminate, when the likelihood of the data given the model cannot be improved.

GAM provides the facility to take distributional assumption of the response variable. Here birth weight of new born baby is considered as response variable which is positive, heteroscedastic and expect to follow exponential family distribution. So, it should be modeled using GAM either by the gamma or the Log-normal models [33]. What are the appropriate models for birth weight of new born babies? What are the determinants of neonates' birth weight? What are effects of the determinants on the neonates' birth weight? These will be discussed in the following sections.

3. Results

3.1. Very Low Birth Weight Analysis and Interpretations

In this present section, we have considered the (very low) birth weight as the response or dependent variable, and the remaining others as the independent (or explanatory) factors or variables. The response birth weight has been modeled through generalized additive model with gamma distribution and logarithm link [33]. The best GAM model is identified through the GCV value (Table 2) along with the model checking criteria (Figure 1, 2 & 3). Adjusted R-square value and the percentage of the deviance explained by the model are also very important to choose the best model. But good R- square value may not be adequate for determining the best model [39]. GAM has two parts of estimation methods; one is parametric estimation for those cofactors which entered in model parametrically and non-parametric estimation used for smoothing cofactors. Through this non-parametric smoothing estimation part GAM tries to control the heterogeneity and the non-linearity (complexity) of the relationship between response variable and the cofactors [32-33]. Table 2 shows the result of the estimations of the model. Table 2 shows both of these estimation results of the GAM model. For finding

the true relationship between BWT and the other cofactors, article has to consider one second order interaction effects beside the main effects in the present model. Interaction effects is very much popular in regression and design of experiment, it means cofactors have a joint influence on response variable. In medical science it is very much popular and well known, because two or three bio chemical parameters may have joint influence on the corresponding response variable [41]. Sometimes insignificant effects are also put in the model in order to respect the marginality rule, namely that when an interaction term is significant, all related lower-order interactions and main effects should be included in the model. This article considered the P-values up to approximately 5% level as highly, and more than 10% to approximately 20% as partially significant [19-21, 33, 41].

Table 2. *Results for GAM of Very low birth weight data analysis using Gamma distribution with 'log' link.*

Estimation of Parametric coefficients				
Covariates	**Estimate**	**Standard Error**	**t value**	**p-value**
Intercept	9.18830	1.27394	7.213	<0.001***
Race (white) 2#	0.01612	0.02015	0.800	0.42509
Race (Native American) 3	-0.18675	0.07088	-2.635	0.00932 **
Race (Oriental) 4	0.19703	0.12331	1.598	0.11222
Inout (Transported) 2	-0.11715	0.06105	-1.919	0.05693.
Twn (Yes) 2	-0.04354	0.02378	-1.831	0.06908.
Sex (Female) 2	-0.03240	0.01902	-1.703	0.09060.
Pneumothorax (Yes) 2	0.14817	0.03249	4.560	<0.001***
Cld (Yes) 2	-0.09961	0.03239	-3.076	0.00251 **
Dead (Yes) 2	-0.08502	0.04695	-1.811	0.07221.
LowpH	-0.29999	0.17654	-1.699	0.09138.
Apg1	-0.45419	0.21398	-2.123	0.03546 *
LowpH * Apg1	0.06235	0.02957	2.108	0.03669 *
Approximate Significance of smooth terms (Non-parametric)				
Smooth Covariate	Edf	Ref. df	F value	p-value
s(hospstay)	8.066	8.742	5.193	<0.001***
s(pltct)	2.694	3.412	8.827	<0.001***
s(gest)	2.947	3.702	19.875	<0.001***

Edf: Estimated degrees of freedom; Ref.df: Degrees of freedom before smoothing; F value: F test score.
Significance Level: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1
R-sq.(adj) = 0.74; Deviance explained = 79.8%; GCV = 0.01578; Scale estimate = 0.01317
2# means at their second level of the corresponding factor described in the Table 1.

In order to examine the proper fitting of the GAM fitted model (Table 2), one model checking criteria with four different plots are shown in Figure 1. First plot of Figure 1 shows theoretical quantiles are plotted against the deviance residuals, second plot shows linear predictor plotted against residuals, in third plot histogram of the residuals are plotted and in forth plot fitted values plotted against response values. All these four plots suggested that the fitted model is adequate for this data analysis, especially the histogram of residuals is almost normally distributed which has an indicator of good fit. Figure 2 shown two plots, namely, the absolute residuals plot and the smoothness of variable 'hospital stay'. In Figure 2(a), displays the absolute residual values are plotted with respect to fitted values under GAM fitted model (Table 2), it is almost a flat diagram with the running means, indicating that the variance is constant for the fitted model. GAM has a non-parametric smoothing terms estimation part for betterment of the model fitting. It also has a graphical part in which variable values are plotted against its smoothness along with the estimated degrees of freedom. Figure 2(b) shows the smoothness of variable hospital stay in days with 95% confidence interval, which indicates that after crossing a certain value of hospital stay of the neonates the smooth curve declined. Figure 3(a), shows the smoothness of variable platelet count and 3(b) shows the smoothness curve of the variable gestational age (weeks) with 95% confidence interval. Both of these two smooth curves show initial increment in BWT with respect to platelet counts and gestational age but after crossing a certain point (value) platelet counts curve has very slow rate of increment which indicates that if platelet counts are increasing for a neonate then his or her weight is also increased. For gestational age after 32 weeks there is a mode of change in the curve and it declines with very slow rate.

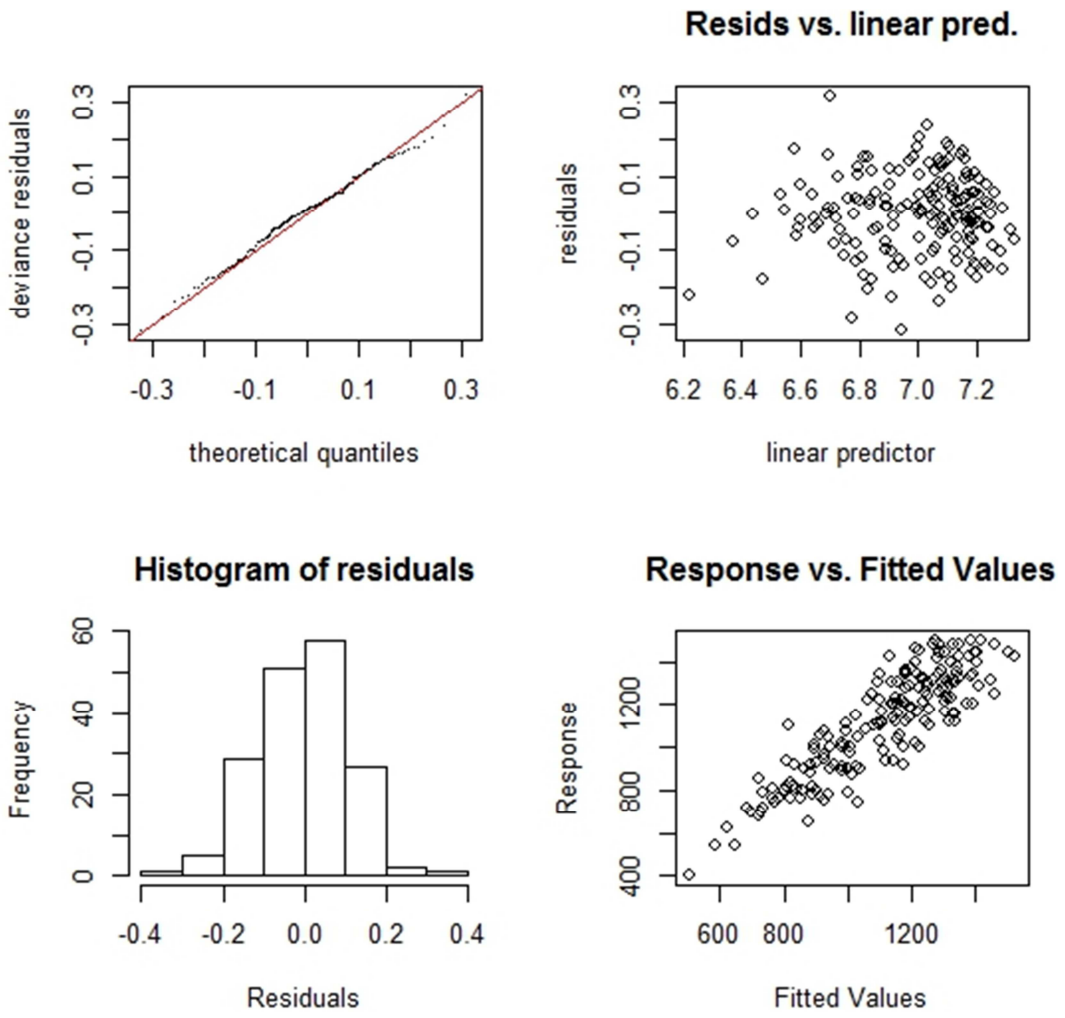

Figure 1. *Regression diagnostic plot of GAM for VLBW data.*

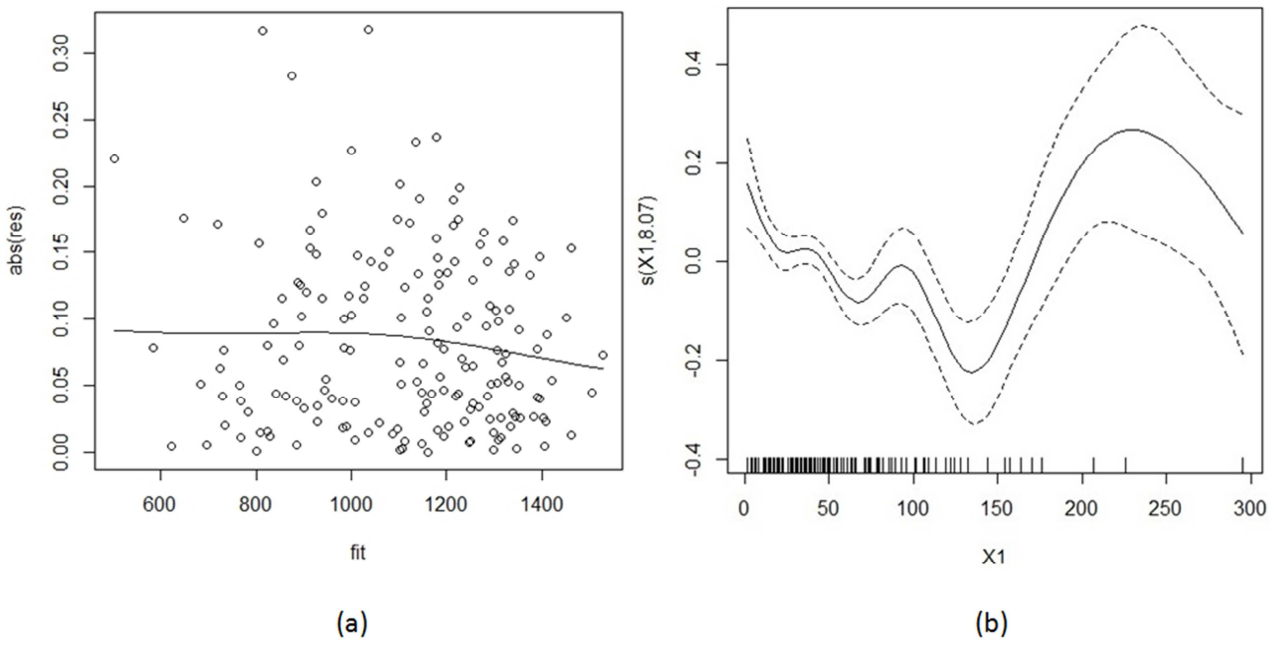

Figure 2. *(a) Residual plot for GAM fitting (b) Plot of hospital stay as a smoothing term of GAM for VLBW data.*

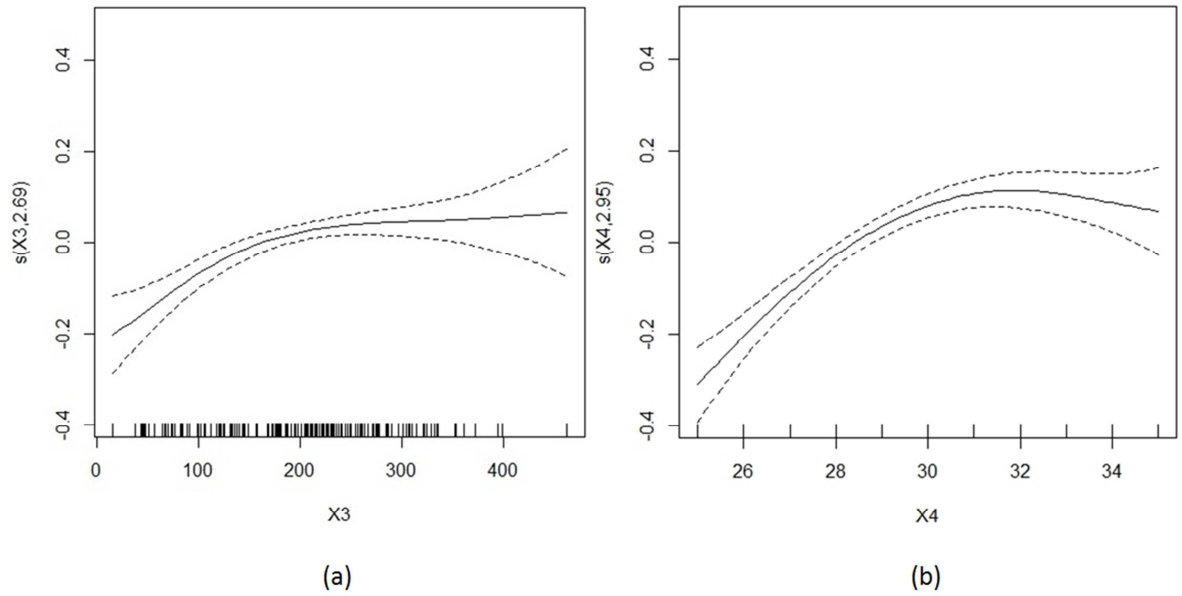

Figure 3. *(a) Plot of platelet count as a smoothing term of GAM (b) Plot of gestational age as a smoothing term of GAM.*

3.2. Results and Interpretations of Low Birth Weight Data Analysis

The results and interpretation of the parametric estimation of cofactors from Table 2 are described as follows,

a. Birth weight (BWT) is high negatively significantly associated with the factor race (Table 1). Out of four categories, Native American neonates have the smallest birth weight than other three racial neonates and Native American category of the factor race is negatively significant with p-value 0.009.
b. In this GAM fitted model, the factor Inout (which stands for whether born at Duke or transported) has a partial negative significant association with BWT with p-value 0.056, which indicates that neonates who are transported from outside Duke have smaller birth weight than who born at Duke.
c. Birth weight (BWT) is partial negatively significantly associated with the factor Twn (Multiple gestations) having p-value 0.07. If the neonates are twins by birth then their birth weights are lower than those neonates who born singly.
d. The factor Sex (male and female) has partial negative significant association with BWT with p-value 0.09. The birth weight of female neonate is less than the male neonates.
e. BWT is highly positively significantly associated with the factor pneumo (means occurrence of pneumothorax) having p-value <0.001. Neonates who suffered from pneumothorax (a diseases which is an abnormal collection of air in the pleural space between the lung and the chest wall) having higher birth weight than who don't have the disease.
f. The factor Cld (indicates requirement of oxygen supply at 30 days of birth) has a high negative significant association with BWT having p-value 0.0025. The neonate who has been required for oxygen supply at 30 days of his/her birth has a smeller birth weight than who don't require this.
g. The factor dead (live status of neonate) has a partial negative significant association with BWT with p-value 0.07. The birth weight of the non-survived neonate is smaller than who had survived.
h. In this GAM fitted model, the variable LowpH (Lowest pH in first 4 days of neonate's life) has a partially negative significant association with BWT having the p-value 0.09. It indicates that if the value of lowest pH in first 4 days of life is increased in neonate's blood then the birth weight (BWT) of that neonate is decreased.
i. Apg1 (The Apgar score, the very first test given to a newborn at one minute) has negative significant association with BWT with p-value 0.03. If the value of Apg1 is increased than BWT value is decreased.
j. Beside main effects described above one interaction effects (LowpH * Apg1) of lowest pH with Apgar score (Apg1) is positively significantly associated with the BWT having p-value 0.03. Though LowpH and Apg1 both are negatively associated with BWT, but the joint effects of these two cofactors are found to be positive. As the interaction effect (LowpH * Apg1) is positively associated with BWT, so if both the LowpH and Apg1 increase then BWT is also increased.

The results and interpretation of the non-parametric estimation of smoothing terms from Table 2 are described as follows,

a. Table 2 shows non-parametric estimation of smoothing terms (cofactors) namely Hospital stays in number of days (Hospstay), Platelet counts (Pltct) and gestational age in weeks (gest). All of these three cofactors entered in the gamma distributed GAM model as smoothing factors. It is observed that F- test statistics has been used for testing this non-parametric smoothness of these

cofactors. The smoothness of all these three cofactors are highly significant with p-value <0.001.

b. It also noticed from Table 2 that, the GAM fitted model has an Adjusted R-square value approximately 0.74 with 80% (approximately) of its deviance explained. The GCV (Generalized cross validation) score is 0.0158 which is also very low compare to other models.

From Table 2, the final selected GAM fitted gamma distributed model of the birth weight (BWT) (y) is shown below

$$\hat{Z} = 9.18 + 0.016Race2 - 0.18Race3 + 0.19Race4 - 0.11Inout2 + 0.04Twn2 - 0.03Sex2 + 0.14Pneumo2 - 0.09Cld2 - 0.08Dead2 - 0.29LowpH - 0.45Apg1 + 0.06(LowpH * Apg1) + f(hospstay) + f(pltct) + f(gest) \quad (2)$$

'*'Denotes the interaction between cofactors and *'f'* denotes the smoothing function.

Where, Z = *ln(y);* ('ln' means Logarithm with base 'e' of y and y is the response variable birth weight of neonate).

4. Discussion

This present article tried to find a relationship between births weights (BWT) of neonate (very low birth weight of neonate) with other covariates (described in Table-1). Birth weight is treated here as a response variable with gamma distribution as an assumption. We tried to model this BWT variable which is a continuous random variable with non-constant variance and non-normal distribution pattern. To model this we introduced generalized additive model popularly known as GAM with a Gamma distributional assumption and logarithm as a link function. The variable descriptions along with their descriptive statistics and the fitted results are presented in Table 1 and 2 respectively. The model checking plots and the other relevant plots such as normal probability plot, absolute residual plot, smoothing term plots are presented in Figure 1, 2, and 3 respectively.

As per our knowledge only few research works has been done using this present dataset (ref) and *M. O'Shea et al.* shows the effects of prenatal factors on the risk of subependymal and lor intraventricular brain haemorrhage in very low birth weight (VLBW) neonates. But this present work has completely another vision – finding the determinants of the neonatal very low birth weight using advanced statistical modeling scheme. Actually, we tried to find those factors which are responsible or affect the neonatal very low birth weight. Discussions regarding the findings of this present article are following below:

a. Table -2 shows that Native American neonates have the smallest birth weight than other racial neonates. Some previous researcher identifies this fact [42-43] but they do not have any mathematical or solid statistical foundation (using some preliminary statistical tools) behind it. Our work proved this fact through mathematical (probabilistic) modeling (equation 6).
b. The gender of the neonate is also significant for determining the very low birth weight; it shows BWT for female is lower than for male. Some early researches also pointed out this fact [44].
c. Neonatal birth place is found to be partially significant in this present study which shows that if the neonate born outside the city of Duke (study place) i.e. *transported* (see Table -1) then the birth weight is decreasing and reasons behind that are completely beyond the scope of our study. In medical research there are no such evidence regarding this finding and we are not claiming any medical justification on it but it's nothing but an additional information to the reader and researchers that may have some other non-medicated reasons.
d. Role of multiple births on neonatal birth weight is often discussible matter in medical sciences and our mathematical model finds this fact [45-46]
e. Many earlier researches pointed out the risk of getting pneumothorax (a diseases which is an abnormal collection of air in the pleural space between the lung and the chest wall) among the new born babies of different birth weights [47-48] but there are no such research article which reports the relationship between birth weight of neonate and the occurrence of pneumothorax. This present article emphasizes this relationship and it has been found that the incident pneumothorax have a significant association with birth weight of a neonate.
f. Another interesting finding of this present work is the relationship between birth weight of a new born and whether the requirement of oxygen supply needed at 30 days of birth. From Table 2 it can be observed that a high negative significant association is present between these two factors. At the time of delivery for very low birth weighted neonates, the supplementary oxygen supply is often required [49-50] but its effects on neonatal birth weight is observed first time in this present work according to our knowledge.
g. Factor 'dead' described that weather the neonate is dead or alive after birth and this present article finds a relationship between neonatal weight with this factor 'dead'. The mortality rate of very low birth weighted neonate is very high so naturally the birth weight of those neonates having death after their birth is low than those are alive. This shows a direct mathematical relation between the mortality and the birth weight, although very low birth weighted neonates having a high rate of mortality, but here we have a clear indication that if the birth weight of the non-survived neonate is smaller than who had survived. Many previous researches are there in support of this but researchers observed this fact in their own way [54].
h. Lowest pH in first four days of neonatal birth is found to be partially significant factor as determinant of birth weight and Table 2 shows that if the lowest pH value is

increased then birth weight of neonate is also decreased. This might be a new finding in the literature of neonatology.

i. One of the most interesting and important finding of this present article is the negative significant association between Apgar1 (The Apgar score, the very first test given to a newborn at one minute) and the neonatal birth weight. Apgar score is the most important indicator for neonatal health and many researchers find the association between Apgar score of neonate with their birth weight [51-53].

j. Another important and interesting finding of this article is the interaction effect of both Apgar score (1 minute) and pH (lowest pH) on birth weight of the neonate. Result indicates that if Apgar score and lowest level pH value is increased then the birth weight of neonate is also increased. This is completely new findings according to best of our knowledge.

k. Platelet counts in new born babies and the gestational age having an association with their birth weight, and this article shows that these two terms having some nonlinear associations with birth weight. This could be identified from the Figure 3(a) and (b). Form Figure 3(a) we can observe that if platelet count is higher than 250 * 10^9/L (upto 400* 10^9/L) having a slow increasing rate in birth weight (BWT) whereas below the range having higher increasing rate in BWT. A platelet count exceeding the upper limit is called thrombocytosis or thrombocythemia [55]. Figure 3(b) reveals that if the gestational age is increased up to 32 weeks then the neonatal birth weight is increasing but after 32 weeks it remains constant. Many researches had been carried out on gestational age and the neonatal birth weight [56-57] and our results support this fact through mathematical modeling.

5. Conclusion

This present article aims to find a relationship between birth weight (BWT) of very low birth weighted neonates and the others cofactors based on a secondary dataset collected at Duke University Medical Center (see material part). Neonatal birth weight is treated here as a response variable following gamma distribution (high positively skewed distribution) as an assumption. We tried to model this BWT variable which is a continuous random variable with non- constant variance and non-normal distribution pattern. To model this we introduced generalized additive model popularly known as GAM with a Gamma distributional assumption and logarithm as a link function. The variable descriptions and the fitted results are presented in Table 1 and 2 respectively. The model checking plots and the other relevant plots such as normal probability plot, absolute residual plot, smoothing term plots are presented in Figure 1, 2, and 3 respectively.

The current reported results (Table 2), though not completely conclusive, are revealing but the determinants of neonate's very low birth weight are derived satisfying the following regression analysis criteria. First, the determinants are selected based on GAM fitted model analyses. Second, the final model is selected based on GCV value. Third, final model is justified based on GAM diagnostic plots [32-34]. Fourth, the standard error of the estimates is very small, indicating that the estimates are stable [39, 41]. Fifth, the final model of the BWT is selected based on locating the appropriate statistical distribution. The BWT distribution is identified herein as the gamma distribution. For more extension regarding this please follow the references [28-30].

To the best of our knowledge, the present models (Results & Discussion section) can be considered as one of the best statistical model under regression framework. The current models may provide a better assistance for researchers and the medical practitioner to know the functional relationship between factors and the birth weight of neonates which is very low and generally below of 1600 grams. The current article finds some very interesting conclusions along with some existing facts in this field. Race of neonate, birth location, multiple births, gender, Pneumothorax occurs or not, oxygen supply needed or not, neonates dead or are the significant categorical variables for birth weights, whereas lowest pH at the 4th day of birth, Apgar score in one minute of birth are the significant continuous variables for BWT. The non-parametric estimation part of this model shows hospital stay in days, platelet counts of neonate and gestational age are the significant smoothing terms. Additionally it is also found in parametric estimation part that, one second order interactions of lowest pH with Apgar score is highly significant for this BWT. Most of these present findings are partially as well as completely new in neonatal health related research literature.

References

[1] Kramer MS. Determinants of low birth weight: methodological assessment and meta-analysis. Bull World Health Organ 1987; 65 (5): 663-737.

[2] Rich-Edwards JW, Buka SL, Brennan RT, Earls F. Diverging associations of maternal age with low birth weight for black and white mothers. Int J Epidemiol 2003; 32: 83-90.

[3] Kumar P, Seshadri R. Neonatal morbidity and growth in very low birth-weight infants after multiple courses of Antenatal Steroids. J Perinatol 2005; 25: 698-702.

[4] Basu S, Rathore P, Bhatia BD. Predictors of mortality in very low birth weight neonates in India. Singapore Med J 2008; 49 (7): 556-560.

[5] Collins JW, David RJ. The different effect of traditional risk factors on infant birthweight among blacks and whites in Chicago. Am J Public Health 1990; 80: 679-681.

[6] David RJ, Collins JW. Differing birth weight among infants of U.S.-born blacks, African-born blacks, and U.S.-born whites. N Eng J Med 1997; 337: 209-214.

[7] Cole C, Binney G, Casey P, Fiascone J, Hagadorn J, Kim C, et al. Criteria for determining disability in infants and children: low birth weight. Evid Rep Technol Assess (Summ) 2002; 70: 1-7.

[8] Chase HC. International comparison of perinatal and infant mortality: the United States and six west European countries. Washington, DC: U.S. Department of Health, Education, and Welfare, Public Health Service; 1967.

[9] Chase HC. Infant mortality and weight at birth: 1960 United States birth cohort. Am J Public Health Nations Health 1969; 59: 1618-1628.

[10] Puffer PR, Serrano CV. Patterns of mortality in childhood: report of the Inter-American investigation of mortality in childhood. Washington, DC: Pan American Health Organization; 1973.

[11] Saugstad LF. Weight of all births and infant mortality. Journal of Epidemiological Community Health 1981; 35: 185-191.

[12] Badshah S, Mason L, McKelvie K, Payne R, Lisboa P. J. G. Risk factors for low Birth weight in the public-hospitals at Peshawar, NWFP-Pakistan, BMC Public Health, 2008; 8: 197.

[13] Roy S, Motghare DD, Ferreira AM, Vaz FS, Kulkarni MS, Maternal determinants of low birth weight at a tertiary care, The Journal of Family Welfare. 2009; 55: 79–83.

[14] 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 (4): 229–235.

[15] O'Shea M, Savitz DA, Hage ML, Feinstein K. A. Prenatal events and the risk of Sub ependymal/intraventricular haemorrhage in very low birth weight neonates. Paediatric and Perinutal Epidemwlogy 1992; 6: 352-362.

[16] McCullagh P, Nelder JA. Generalized linear models. London: Chapman & Hall; 1989.

[17] Myers RH, Montgomery DC, Vining GG. Generalized linear models with applications in engineering and the sciences. New York: John Wiley & Sons; 2002.

[18] Das RN, Lee Y. Log normal versus gamma models for analyzing data from quality- improvement experiments. Qual Eng 2009; 21 (1): 79-87.

[19] Das RN, Mukhopadhya BB. Affects of thyroid function and maternal urinary iodine on neonatal weights. J Neonatal Nurs 2009; 15: 204-211.

[20] Das RN. The role of Iodine in the thyroid status of mothers and their neonates. Thyroid Science 2011; 6 (2): 1-15.

[21] Das RN, Dihidar S, Verdugo R. Infant mortality in India: Evaluating Log-Gaussian and Gamma. Open Demography Journal 2011; 4: 34-41.

[22] Das RN, Park JS. Discrepancy in regression estimates between Log-normal and Gamma: Some case studies. J Appl Stat 2012; 39 (1): 97-111.

[23] Hosmer R, Lemeshow J. Applied logistic regression. 2nd ed. New York: John Wiley & Sons

[24] Box GEP, Cox DR. An analysis of transformations. J R Stat Soc Series B Stat Methodol 1964; 26: 211-252.

[25] Firth D. Multiplicative errors: log-normal or gamma? J R Stat Soc Series B Stat Methodol 1988; 50: 266-268.

[26] Nelder JA, Lee Y. Generalized linear models for the analysis of Taguchi-type experiments. Applied Stochastic Models and Data Analysis 1991; 7: 107-120.

[27] Cox DR, Reid N. Parameter orthogonality and approximate conditional inference. J R Stat Soc Series B Stat Methodol 1987; 49: 1-39.

[28] Lee Y, Nelder JA. Generalized linear models for the analysis of quality improvement experiments. Can J Statist 1998; 26: 95-105.

[29] Lee Y, Nelder JA. Robust design via generalized linear models. J Qual Tech 2003; 35: 2-12.

[30] Das RN, Mukherjee S. Joint Mean-Variance Overall Survival Time fitted Models from Stage III Non-Small Cell Lung Cancer. Epidemiology (Sunnyvale); 2017; 7: 296.

[31] Schimek MG. Estimation and inference in partially linear models with smoothing splines. Journal of Statistical Planning and Inference. 2000; 91 (2): 525-540.

[32] Hastie T, Tibshirani R. Generalized Additive Models, first ed., London: Chapman and Hall 1990.

[33] 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 (4): 137-44.

[34] Wood SN. Generalized Additive Models: An Introduction with R. London: Chapman and Hall; 2006.

[35] Currie I, Durban M, Eilers P. Generalized linear array models with applications to Multi-dimensional smoothing. J R Stat Soc B. 2006; 68: 259–280.

[36] Green P, Silverman B. Nonparametric Regression and Generalized Linear Models. London: Chapman & Hall; 1994.

[37] Ruppert D. Selecting the number of knots and for penalized splines. J Comp Graph Stat. 2002; 11: 735-757.

[38] Eilers P, Marx B. Flexible smoothing with B-splines and penalties. StatSci 1996; 11: 89-121.

[39] S. Chatterjee., A. S. Hadi, Regression Analysis by Example, fifth ed. John Wiley & Sons, New Jersey; 2006.

[40] Hastie T, Tibshirani R. Generalized additive models for medical research. Statistical Methods in Medical Research. 1995; 4: 187-196.

[41] Das RN, Mukherjee S, Sharma I. Alkaline Phosphatase Determinants of Liver Patients. JOP. Journal of the Pancreas. 2018; 19 (1): 18-23.

[42] Jr Collins J. W, David R. J, Handler A, Wall S, Andes S. Very Low Birth weight in African American Infants: The Role of Maternal Exposure to Interpersonal Racial Discrimination, American Journal of Public Health. 2004; 94 (12): 2132–2138.

[43] Alexander G. R, Tompkins M. E, Altekruse J. M, Hornung C. A. Racial differences in the relation of birth weight and gestational age to neonatal mortality. Public Health Reports. 1985; 100 (5): 539–547.

[44] Stevenson D, Verter J, Fanaroff A, Oh W, Ehrenkranz R, Shankaran S, Donovan E, Wright L, Lemons J, Tyson J, Korones S, Bauer C, Stoll B, Papile L, Sex differences in outcomes of very low birthweight infants: the newborn male disadvantage, Archives of disease in childhood. Fetal and neonatal edition. 2000; 83 (3): 182–185.

[45] Magee BD, Role of multiple births in very low birth weight and infant mortality, Journal of Reproductive Medicine 2004; 49 (10): 812-816.

[46] Impact of multiple births on low birth weight--Massachusetts, 1989-1996. Centers for Disease Control and Prevention (CDC), MMWR Morbidity and Mortality Weekly Report. 1999; 48 (14): 289-292.

[47] Aly H, Massaro A, Acun C, Ozen M. Pneumothorax in the newborn: clinical presentation, risk factors and outcomes, The Journal of Maternal-Fetal & Neonatal Medicine. 2014; 27 (4): 402-416.

[48] Terzic S, Heljic S, Panic J, Sadikovic M, Maksic H. Pneumothorax in premature infants with respiratory distress syndrome: focus on risk factors. Journal of Pediatric and Neonatal Individualized Medicine. 2016; 5 (1): e050124.

[49] Tin W, Gupta S. Optimum oxygen therapy in preterm babies. Archives of disease in childhood Fetal and neonatal edition. 2007 Mar; 92 (2): F143–F147.

[50] Cummings JJ, Polin RA, COMMITTEE ON FETUS AND NEWBORN. Oxygen Targeting in Extremely Low Birth Weight Infants. Pediatrics. 2016; 138 (2): e20161576.

[51] Rahi E, Baneshi M. R, Mirkamandar E, Maghsoudi S. H, Rastegari A. A Comparison between APGAR Scores and Birth Weight in Infants of Addicted and Non-Addicted Mothers, Addiction & Health 2011; 3 (1): 61-67.

[52] Hegyi T, Carbone T, Anwar M, Ostfeld B, Hiatt M, Koons A, Pinto-Martin J, Paneth N. The Apgar Score and Its Components in the Preterm Infant. Pediatrics, 1998; 101 (1): 77-81.

[53] Behnke M, Carter RL, Hardt NS, Eyler FD, Cruz AC, Resnick MB. The relationship of Apgar scores, gestational age, and birth weight to survival of low-birth weight infants. American journal of perinatology 1987; 4 (2): 121-4.

[54] Gera T, Ramji S. Early Predictors of Mortality in Very Low Birth Weight Neonates, Indian Pediatrics 2001; 38: 596-602.

[55] Wilejto M, Steele M, Jadavji T. Dropping platelet counts in the neonatal intensive care unit – an unsuspected cause for thrombocytopenia in a neonate. Pediatric Child Health. 2011; 16 (9): 557–558.

[56] Lesiński J. Relationship between length of gestation, birth weight and certain other factors: A statistical study. Bulletin of the World Health Organization 1962; 26 (2): 183–191.

[57] Ozdemir O, Tunay Z. O, Acar D. E, Erol M. K, Sener E, Acar U. The relationship of birth weight, gestational age, and postmenstrual age with ocular biometry parameters in premature infants. Arquivos Brasileiros De Oftalmologiavol. 2015; 78 (3): 146-149.