Relationship between Pretreatment Serum Albumin Levels with the Risk of Malignant Pleural Mesothelioma

Sabyasachi Mukherjee*

Department of Mathematics, NSHM Knowledge Campus, Durgapur, West-Bengal, PIN-713212, India

Abstract: *Background*: Malignant Pleural Mesothelioma (MPM) is a very rare and aggressive form of cancer. Recently it was found that pretreatment Serum Albumin (SA), the main circulating protein in blood is a significant prognostic factor for MPM patients. The objective of this present article is to show the relationship between pretreatment Serum albumin (SA) levels with the risk of MPM.

Methods: Generalized additive model (GAM), an advanced regression analysis method has been introduced here to find this mathematical relationship between the response variable (SA) and the cofactors.

Results: The main determinates of SA are identified - asbestos exposure, hemoglobin, disease diagnosis status (patients having MPM) are the factors having significant association with SA, whereas duration of asbestos exposure, duration of disease symptoms, total protein (TP), Pleural lactic dehydrogenise (PLD), pleural protein (PP), pleural glucose (PG) and C-reactive protein (CRP) are the significant continuous variables for SA. The non-parametric estimation part of this model shows Lactate dehydrogenase (LDH) and Glucose level are the significant smoothing terms. Additionally it is also found that, second and third order interactions between cofactors are highly significant for SA.

Conclusions: The findings of this present work can conclude that - serum albumin may play the role of a very significant prognostic factor for MPM disease and it has been established here through mathematical modeling. Few of the findings are already been exist in MPM research literature whereas some of the findings are completely new in the literature.

Keywords: Malignant Pleural Mesothelioma, Serum albumin, Gamma distribution, Generalized additive model, Probabilistic Modeling.

1. INTRODUCTION

Malignant pleural mesotheliomas (MPM) are very aggressive tumors, which is a disease originating from pleura, pericardium, peritoneum or tunica vaginalis and it is since the early 1960s recognized to be strongly related to asbestos exposure [1], however it may also be related to previous simian virus 40 (SV40) infection, radiation and quite possible for genetic predisposition [2, 3]. The incidence of malignant pleural mesothelioma (MPM) is extremely high in some Turkish villages where there is a low-level environmental exposure to erionite, is a naturally occurring fibrous mineral that belongs to a group of minerals called zeolites. Environmental asbestos exposure and MPM are one of the major public health problems of Turkey. Molecular mechanisms can also be implicated in the development of mesothelioma [4]. Rural living is associated with the development of mesothelioma [5-7]. Soil mixtures containing asbestos, known as 'white-soil' or 'corak' can be found in Anatolia, Turkey and 'Luto' in Greece [7, 8-11]. MPM is a fatal cancer of increasing incidence associated with asbestos exposure [12]. MPM is a malignancy that is resistant to the common tumor directed therapies, but again individual patients might respond to chemotherapy, radiotherapy or immunotherapy, and selected patients might benefit from radical surgery and multimodality treatment [13]. MM is a rare disease with an incidence rate of 1–2 per million/year [14] in the general population. In industrialized countries, the rate ranges from 1 to 5 per million/year for women and 10–30 per million/year for men [15-17]. The higher incidence rates in industrialized countries may be due to asbestos exposure [11]. Recently it is observed that, MPM are responsible for approximately 15,000–20,000 deaths annually worldwide [4]. Estimated 1000 patients have MPM in Turkey per year. The annual incidence of pleural mesothelioma was 22.4/1,000,000 in Anatolia [18].

The most of the work using this MPM dataset were diagnostic works which are based on various classifiers [23-24, 55]. Object was to classify or diagnosis the disease with minimum misclassification rate. Diagnosis usually appears when a patient visits the doctor to have symptoms checked out. Patients may be met with shortness of breath, pain in the chest or back, painful, persistent coughing or any number of other symptoms, none of which immediately alert the doctor to a diagnosis of mesothelioma [19]. Several studies were carried out about MPM epidemiology, clinics in south east of Turkey [20-22]. There are many studies on MPM disease diagnosis using artificial intelligence

*Address correspondence to this author at the Department of Mathematics, NSHM Knowledge Campus, Durgapur, West-Bengal, PIN-713212, India; Mob: +919775745309/+918918060382; E-mail- sabyasachi99@gmail.com

E-ISSN: 1929-6029/20

techniques also like, probability neural networks (PNNs), learning vector quantization (LVQ) [23], artificial immune system (AIS) and multi-layer neural network (MLNN) [24] with prognostic data. MPM is a very rare type of malignant and fatal disease with a poor prognosis.

Serum albumin (SA), the main circulating protein in blood is a prognostic factor for MPM patients. This finding is recently established by a team of Chinese researchers, the report shows that the abundant protein may offer one of the simplest ways to predict mesothelioma prognosis [25]. Human serum albumin or simply serum albumin constitutes about half of serum protein. It is produced in the liver. It is soluble and monomeric. Albumin transports hormones, fatty acids, and other compounds, buffers pH, and maintains oncotic pressure, among other functions. Albumin is synthesized in the liver as preproalbumin, which has an N-terminal peptide that is removed before the nascent protein is released from the rough endoplasmic reticulum. The product proalbumin, is in turn cleaved in the Golgi vesicles to produce the secreted albumin. The reference range for albumin concentrations in serum is approximately 35 - 50 g/L (3.5 - 5.0 g/dL), a lower-than-normal level of blood albumin may be a sign of many diseases such as liver, kidney diseases and now it is also a prognostic factor for MPM disease. It has a serum half-life of approximately 20 days. It has a molecular mass of 66.5 kDa [26].

This present article aims to explore a relationship between SA and the biochemical, demographic parameters from the dataset of MPM patients. Serum albumin (SA) is playing the role of response variable (other factors and variables are the possible cofactors) which is positive, heterogeneous and non-normally distributed continuous random variable and generally modeled through either gamma or log normal distribution. It has been also observed that few biochemical parameters are non-linearly associated with SA. So, it could be better to practice generalized additive model (GAM) in place of any other ordinary regression like multiple regression or generalized linear model (GLM)[27]. Joint GLM can also be handled this type of positive, non-normal, heterogeneous data, but still this article preferred to show the GAM application here because of the method's flexibility and efficiency in the fields of complex data analysis [28-30].

In the statistical analysis of clinical trials and observational studies, the identification and adjustment of prognostic factors are an important activity in order to get a 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 flexible statistical methods that may be used to identify and characterize the effect of potential prognostic factors on disease endpoints. These methods are called 'Generalized Additive Models' (GAM) [31-33].

The major objective of this study is to explore a relationship between SA and the other bio medical parameters of MPM patients. Many authors had used various classification techniques on this dataset for MPM disease diagnosis [23, 24], but probably, advance regression or probabilistic modeling techniques are not been used under proper modeling scheme.

2. MATERIAL AND METHODS

2.1. Material

In order to perform the research reported, the patient's hospital reports from Dicle University, Faculty of Medicine's were used in this work. One of the special characteristics of this diagnosis study is to use the real dataset taking from patient reports from this hospital [24]. Three hundred and twenty-four (324) MM patient data were diagnosed and treated. These data were investigated retrospectively and analyzed files. In the dataset, all samples have 35 features because it is more effective than other factors subsets by doctor's guidance. These features are age, gender, city, asbestos exposure, type of MM, duration of asbestos exposure, diagnosis method, keep side, cytology, duration of symptoms, dyspnoea, ache on chest, weakness, habit of cigarette, performance status, White Blood cell count (WBC), hemoglobin (HGB), platelet count (PLT), sedimentation, blood lactic dehydrogenises (LDH), Alkaline phosphatise (ALP), total protein, albumin, glucose, pleural lactic dehydrogenises, pleural protein, pleural albumin, pleural glucose, dead or not, pleural effusion, pleural thickness on tomography, pleural level of acidity (pH), C-reactive protein (CRP), class of diagnosis. Diagnostic tests of each patient were recorded. Table **1** shows the detail descriptions of variables and their descriptive statistics. This present study based on the dataset collected from UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Mesothelioma).

Table 1: Description of Variables with their Analyzed & Summarized Statistics

Variable name	Operationalization/Description	Mean	Standard deviation	Proportion of levels of Attributes
Age (Year) (x1)	Age of the patient at study time	54.74	11.02	---
Gender (F1)	Gender : (Female = 0 ; Male = 1)	---	---	0 = 41.36 %; 1= 58.64 %
City (F2)	City from where the patient belongs Nine cities are here. (Not been incorporated in model due to lack of information)	---	---	----
Asbestos exposure (F3)	Whether patients exposed to asbestos or not. No = 0 ; Yes = 1	---	---	0 = 13.58 % 1 = 86.42 %
Type of MM (F4)	Pleural =1 ; Peritoneal =2; Pericardial =3.	---	---	1 = 95.68 %; 2 = 3.40 %; 3 = 0.93 %
Duration of asbestos exposure (x2)	How many years a patient had been exposed to asbestos in his/her life	30.18	16.42	---
Diagnosis method(F5)	It refers to whether the tumor has been diagnosed before starting this study or not. No = 0 ; Yes = 1	---	---	0 = 29.63 % 1 = 70.37 %
Lung side(F6)	It refers to the lung interested by disease. Right =1 ; Left = 2 ; Both = 3	---	---	1 = 30.86 % ; 2= 62.35 % ; 3 = 6.79 %
Cytology (F7)	Cytology means the study of cells. Normal cell = 0 ; Cancerous cell = 1	---	---	0 = 71.90 % 1 = 28.10 %
Duration of symptoms (x3)	It refers to the time period, in months, in which the patients show symptoms.	5.45	4.72	---
Dyspnoea (F8)	It means shortness of breath and refers to whether a patient has difficulty breathing. No = 0 ; Yes =1	---	---	0 = 18.21 % 1 = 81.79 %
Ache on chest (F9)	It is related to the presence or absence in patient of pain in the abdomen area and in particular in chest in case of pleural mesothelioma. No = 0 ; Yes = 1	---	---	0 = 31.8 % 1 = 68.2 %
Weakness (F10)	Weakness or asthenia refers to whether or not patients feel lack of strength. No = 0 ; Yes = 1	---	---	0 = 38.9 % 1 = 61.10 %
Habit of cigarette (F11)	It is characterized by four category based on patient's habit of smoking.	---	---	1 = 56.48% ; 2 = 11.42% 3 = 16.67 % ; 4 = 15.43 %
Performance status (F12)	PS is a feature characterized by two categories and estimates whether or not patients are able to perform certain activities of daily living? No = 0 ; Yes =1	---	---	0 = 47.84 % 1 = 52.16 %
White blood (x4)	WB refers to the number of the white blood cells in one microliter.	9457	3451	---
Cell count (WBC) (x5)	WBC count can detect hidden infections and undiagnosed medical conditions.	9.55	3.34	---
Hemoglobin (HGB) (F13)	It means Hemoglobin normality test refers to the hemoglobin test that measures how much hemoglobin is in blood. Normal = 0 ; Higher than normal = 1	---	---	0 = 57.72 % 1 = 42.28 %
Platelet count (PLT) (x6)	It is a laboratory test to measure how many platelets you have in your blood.	369.7	227.6	---
Sedimentation (x7)	The sedimentation rate (sed rate) blood test measures how quickly red blood cells (erythrocytes) settle in a test tube in one hour (mm/hr).	70.69	21.75	---

(Table 1). Continued.

Variable name	Operationalization/Description	Mean	Standard deviation	Proportion of levels of Attributes
Blood lactic dehydrogenise (LDH) (x8)	LDH is a protein that helps produce energy in the body.	308.9	185.1	---
Alkaline phosphatise (ALP) (x9)	(ALP) is a protein found in all body tissues and the normal range is 44 to 147 IU/L. An ALP test may be used to detect cancers that have spread to the bones.	66.16	35.08	---
Total protein(x10)	Total protein, also known as serum total protein, is a biochemical test for measuring the total amount of protein in serum.	6.58	0.82	---
Serum Albumin (x11)	Serum albumin (SA), the main circulating protein in blood.	3.30	0.63	---
Glucose (x12)	A blood glucose test measures the amount of glucose in a sample of blood.	112.41	38.46	---
Pleural lactic dehydrogenise (PLD) (x13)	The upper limit of the normal pleural lactic dehydrogenase is 200 IU//L. A high LD indicates that pericardial fluid and while a low level indicates it is transudate.	518.5	536.3	---
Pleural protein(x14)	Normal pleural proteins count is less than 1-2 g/dL.	3.93	1.57	---
Pleural albumin(x15)	Pleural albumin is the level of albumin in the pleural fluid	2.07	0.91	---
Pleural glucose(x16)	It refers pleural fluid glucose.	48.44	27.23	---
Live or Dead (F14)	“Dead or not” refers to whether or not a patient is still alive during the study. Live = 0 ; Dead = 1	---	---	0 = 5.56 % 1 = 94.44 %
Pleural effusion (F15)	In some cases, the presentation of an “pleural effusion” signals advancement of the disease or malignant mesothelioma. No = 0 ; Yes = 1	---	---	0 = 12.96 % 1 = 87.04 %
Pleural thickness on tomography (F16)	Pleural thickness on tomography is a descriptive term given to describe any form of thickening involving either the parietal or visceral pleura. Parietal = 0 ; visceral =1	---	---	0 = 40.43 % 1 = 59.57 %
Pleural level of acidity (pH) (F17)	“Pleural level of acidity” means whether or not the pleural fluid is lower than the normal pleural fluid, which has a pH of 7.60-7.64. No =0 ; Yes = 1	---	---	0 = 47.84 % 1 = 52.16 %
C-reactive protein (CRP)(x17)	“C-reactive protein” (CRP), an acute phase reactant.	64.20	22.66	---
Class of diagnosis (F18)	Target or output value. Mesothelioma = 1; Not mesothelioma = 2	---	---	1 = 70.38 % ; 2 = 29.62 %

2.2. Methods

In this present article, an advanced regression technique namely Generalized additive model (GAM) [31, 32, 34] has been performed for finding the association between serum albumin and other parameters (biochemical, demographic and others) for MPM disease dataset. Recently, it has been established that SA is an important prognostic factor for MPM disease [25]. The factors which influenced SA the most both negatively or positively, article tries to explicit this.

Best GAM model can be selected through some model checking criteria namely R-square value, Akaike Information Criterion (AIC), Bayesian information

criterion (BIC) or Generalized Cross Validation (GCV) value and regression diagnostic plots like normal probability plot, Residuals against fitted value plot etc. [31, 32, 34]. Cofactors are significant or not judged through p-value. Approximate significance of smooth terms is also judged through p-value. For this MPM data set serum albumin (SA) is taken as response variable (Y), and age, gender, city, asbestos exposure, type of MM, duration of asbestos exposure, diagnosis method, keep side, cytology, duration of symptoms, dyspnoea, ache on chest, weakness, habit of cigarette, performance status, White Blood cell count (WBC), hemoglobin (HGB), platelet count (PLT), sedimentation, blood lactic dehydrogenises (LDH), Alkaline phosphatise (ALP), total protein, glucose, pleural lactic dehydrogenises, pleural protein, pleural albumin, pleural glucose, dead or not, pleural effusion, pleural thickness on tomography, pleural level of acidity (pH), C-reactive protein (CRP), class of diagnosis are the cofactors (X_i' s). Including class of diagnosis there are total thirty five (35) parameters, out of which eighteen (18) are categorical and seventeen (17) are continuous variables. This present MPM dataset contains total 324 numbers of patients with no missing value. [23, 24]

2.2.1. Generalized Additive Model (GAM)

GAM [31, 32, 34] is an extension of the Generalized Linear Model (GLM) [27] where the modeling of the mean functions relaxes the assumption of linearity, albeit additively 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 $j = 1,2,\ldots,p$.

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 [27]. Smoothing splines [37, 38], 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 [35-38].

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 $\sum_{j=1}^{p} f_j(x_{ij})$ [30]. 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).

The GAM regression techniques are used for this MPM disease dataset. All statistical and data analytic works, mainly GAM regression are performed in R statistical software [34].

3. RESULTS

This present section considered serum albumin (SA) as a dependent or response variable and remaining others as independent variable or cofactors. The SA is positive valued, non- normally distributed, heterogeneous (non-constant variance) continuous variable. This response variable has been modeled through gamma distributed log linked generalized additive models. The relationship between SA and the others cofactors is very complicated. The best GAM model is identified through the GCV value (Table **2**) along with the model checking criteria (Figure **1** and **2**). 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

Table 2: Results for GAM of Serum Albumin Data Analysis using Gamma Distribution with 'log' Link

Estimation of Parametric coefficients				
Covariates	**Estimate**	**Standard Error**	**t value**	**p-value**
Intercept	-4.533e+00	7.774e-01	-5.831	<0.001
Asbestos exposure(AE) (1)[#]	1.137e-01	3.885e-02	2.926	0.003
Hemoglobin (1)[#]	4.753e-02	1.721e-02	2.762	0.006
Diagnosis class (2)[##]	3.356e-02	1.864e-02	1.800	0.072
Age	9.072e-04	8.897e-04	1.020	0.308
Duration of AE	-1.823e-03	8.755e-04	-2.082	0.038
Duration of Symptoms	-3.395e-03	1.771e-03	-1.917	0.056
Total Protein (TP)	8.662e-01	1.133e-01	7.646	<0.001
Pleural lactic dehydrogenise (PLD)	4.078e-04	1.748e-04	2.332	0.020
Pleural protein (PP)	6.793e-01	1.095e-01	6.203	<0.001
Pleural albumin (PA)	-1.605e-01	1.017e-01	-1.578	0.115
Pleural glucose (PG)	6.442e-02	1.176e-02	5.477	<0.001
C-reactive protein (CRP)	2.308e-02	6.880e-03	3.355	<0.001
TP * CRP	-3.407e-03	1.052e-03	-3.239	0.01
TP * PG	-9.944e-03	1.725e-03	-5.765	<0.001
PG * CRP	-3.346e-04	1.173e-04	-2.852	0.004
TP * PP	-1.090e-01	1.472e-02	-7.408	<0.001
PP * PG	-8.709e-03	1.734e-03	-5.020	<0.001
PP * PA	3.776e-02	1.813e-02	2.083	0.038
PA * PG	4.637e-03	1.580e-03	2.936	0.003
TP * PLD	-6.339e-05	2.926e-05	-2.166	0.031
PLD * PG	-1.982e-05	7.311e-06	-2.711	0.007
TP * PG * CRP	4.812e-05	1.785e-05	2.696	0.007
TP * PP * PG	1.277e-03	2.277e-04	5.609	<0.001
PP * PA * PG	-6.955e-04	2.547e-04	-2.730	0.006
TP * PLD * PG	3.136e-06	1.143e-06	2.743	0.006
Approximate Significance of smooth terms (Non-parametric)				
Smooth Covariate	Edf	Ref. df	F value	p-value
s(LDH)	3.28	4.08	2.92	0.020
s(Glucose)	8.61	8.95	7.16	<0.001

Edf: Estimated degrees of freedom; **Ref.df:** Degrees of freedom before smoothing; **F value:** F test score.
R-sq.(adj) = 0.588 ; Deviance explained = 65.3% ; GCV = 0.0230 ; Scale estimate = 0.0196.
1[#]means at their second (higher) level & 2[##] means non MPM patients reported in the Table **1**.

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 [30]. Table **2** shows the result of the estimations of the model. For finding the true relationship between SA and the other cofactors, article has to considered second and third order interaction effects in the present model. Interaction effects is very much popular in regression and design of experiment, it means cofactors have a joint effect on response variable. In medical science data analysis also it is very much relevant, because two or three bio medical parameters

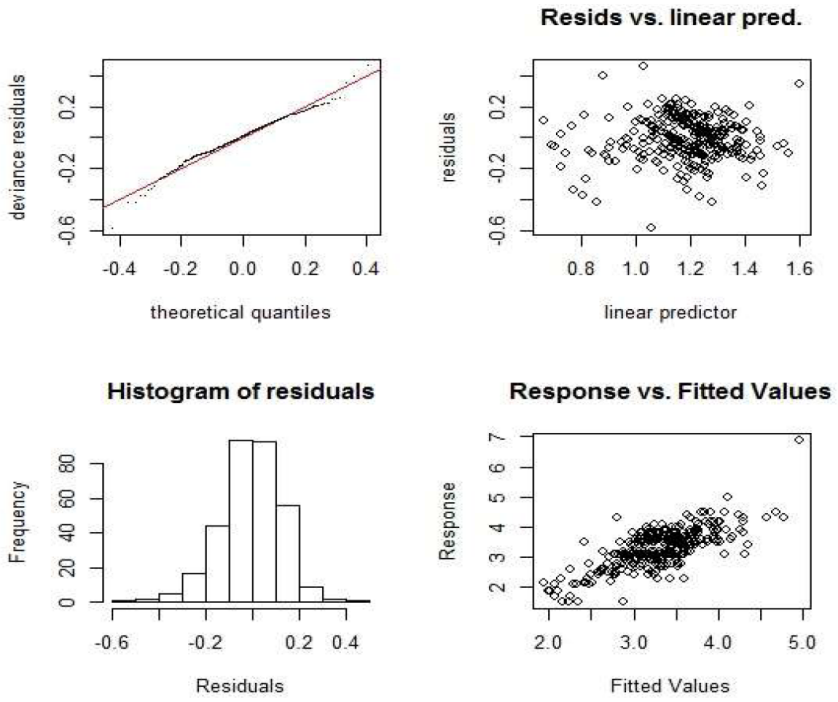

Figure 1: Model checking plots for GAM.

may have joint influence on the corresponding response variable [28, 29]. Some insignificant effects are retained 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 [40, 41]. This article considered the P-values up to approximately 10% level as significant, and more than 10% to approximately 20% as partially significant. [28-30, 40, 41].

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** & **3** shown two diagnostic plots, namely, the absolute residuals plot and the normal probability plot. In Figure **2**, displays the normal probability plot of the GAM fitted model (Table **2**), which does not show lack of fit for outliers or variables as there is not much more gap in the figure, only except in the lower and upper part of figure, which shows little deviations due to complexity of the relationship. Figure **3** the absolute residual values are plotted with respect to fitted values. 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 **4** shows the smoothness of variable LDH with 95% confidence interval, which indicates that

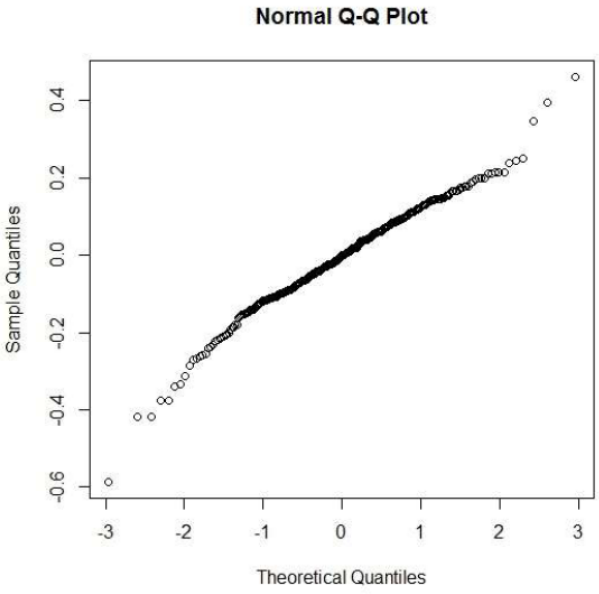

Figure 2: Normal probability plot.

after crossing a certain value of LDH the smooth curve declined. Figure **5**, shows the smoothness of variable Glucose with 95% confidence interval. This smooth curve shows non-linearity with respect to the increment of glucose value.

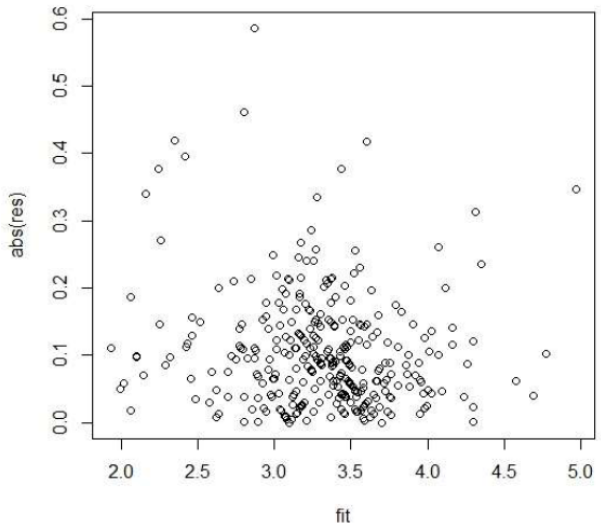

Figure 3: Absolute residual vs fitted value.

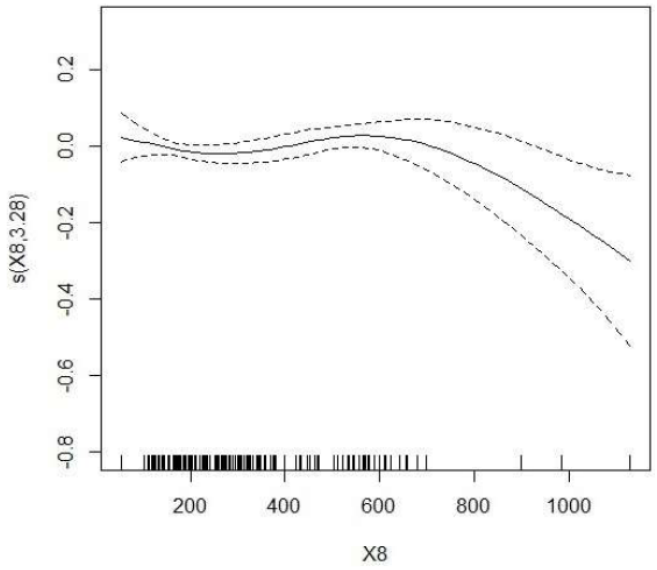

Figure 4: Plot of the smoothing term LDH.

3.1. Interpretations of Serum albumin data analysis

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

i. Serum Albumin (SA) is high positively significantly associated with the factor Asbestos exposure (Table **1**) with p-value 0.003. Those who had experienced the asbestos exposure have more SA than the other patients. This can also finds from the dataset directly using basic descriptive statistics. The average SA value for those who had the experience of asbestos exposure is 3.32, whereas for others have the average SA value 3.16.

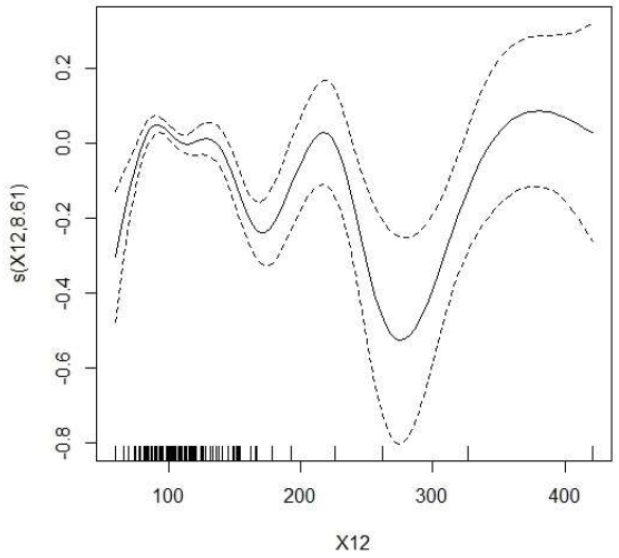

Figure 5: Plot of the smoothing term glucose.

ii. In this GAM fitted model, the factor Hemoglobin has a positive significant association with SA with p-value 0.006 which indicates that patients with higher hemoglobin than normal range having more SA value than the rest.

iii. Serum Albumin (SA) is partial positively significantly associated with the factor diagnosis class of MPM with p-value 0.07. The patient who encounters the MPM disease has a lower value of SA than others. This finding suggests that, the patients having lower level

iv. SA value getting higher chance of MPM disease. More clearly it can be concluded that non MPM patients having higher SA than MPM patients.

v. Duration of asbestos exposure has a negative significant association with SA with p-value 0.038. This indicates that if duration of asbestos exposure is increased then the SA value is decreased.

vi. SA is negatively significantly associated with the duration of symptoms with p-value 0.05, which signify that if a patient is going through a long time of MPM disease symptoms then the SA value is reduced.

vii. Total protein (TP) has a high positive significant association with SA with the p-value <0.001. It indicates that if the value of TP is increased in blood then the SA value is also increased.

viii. In this GAM fitted model, Pleural glucose (PG) has a high positive significant association with SA. With p-value <0.001, which indicates that, if PG value is increased then SA is also increased.

ix. C-reactive protein (CRP) has high positive significant association with SA with p-value <0.001. If the value of CRP is increased than SA value is also increased.

x. The joint interaction effects (TP * CRP) of total protein (TP) with C-reactive protein (CRP) is high negatively significantly associated with the SA having p-value <0.001. Though TP and CRP both are positively associated with SA, but the joint effects of these two cofactors are found to be negative. As the interaction effect (TP * CRP) is negatively associated with SA, so if both the TP and CRP increase then SA decreased.

xi. Total protein (TP) and Pleural glucose (PG) both have the positive association with SA, but the joint interaction effects (TP * PG) of TP and PG is high negatively significantly associated with the SA with the p-value <0.001. It indicates that if both the TP and PG increase then SA decreased.

xii. Total protein (TP) is positively associated with SA, which is already reported in this section. It has two more interaction effects [(TP*PP) & (TP*PLD)] with Pleural protein (PP) and with Pleural lactic dehydrogenise (PLD), which are negatively associated with SA with p-value <0.001 and 0.03 respectively. That means if both of these two cofactors TP and PP are increased then SA is decreased and this happens for TP and PLD case also, i.e. if TP and PLD are increased then SA is decreased.

xiii. In case of Pleural glucose (PG), it is positively significantly associated with SA and it has four interactions effects with the cofactors CRP, PP, Pleural albumin (PA) and PLD. Except PA other three cofactors have the positive marginal effects on SA (reported earlier), but the interaction effects (PG*CRP) of PG with CRP is high negatively associated with SA (p-value= 0.004), the interaction effects (PG*PP) of PG with PP has also very high negative association (p-value = <0.001) and the second order interaction effects (PG*PLD) of PG with PLD has high negative association with SA (p-value= 0.007). But the joint interaction effects (PG*PA) of PG with PA has the high positive association with SA (p-value = 0.03). That means if PG and CRP both are increased then SA is decreased, it also happens for PG and PP case, that is in PG and PP both are increased then SA is decreased. For PG and PLD, if they increased SA is decreased. Only in case of PG and PA, if they increased then SA is also increased.

xiv. Another second order interaction effects (TP*PLD) of total protein (TP) and Pleural lactic dehydrogenise (PLD) is reported here. If both of these two cofactors TP and PLD increased then SA is decreased with p-value 0.03.

xv. A third order interaction effects (TP*PG*CRP) of total protein (TP), pleural glucose (PG) and C-reactive protein (CRP) jointly have high positive significant association with SA (p-value=0.007). It is already been reported earlier that the second order interaction effects of (TP*CRP), (TP*PG) and (PG*CRP) are negatively associated with SA. But TP and PG along with the CRP have a very complicated relation with SA.

xvi. Apart from this, TP has two more third order interaction effects [(TP*PG*PP) & (TP*PG*PLD)] with PG and PP and with PG and PLD. The second order interaction of TP and PG is negatively associated with SA, but the third order interaction effects of TP and PG (TP*PG) with PP is high positively associated (p-value = <0.001) with SA. And in case of PLD, the third order interaction effects of TP, PG and PLD has high positive association with SA (p-value=0.006). The interrelationship between these cofactors with the response variable SA is much more complicated.

xvii. Finally, the joint third order interaction effects (PP*PA*PG) of Pleural protein (PP), Pleural albumin (PA) and Pleural glucose (PG) has high negative significant association with SA (p-value=0.006). The second order interaction effects of PP and PA (PP*PA), PA and PG (PA*PG), are positively significantly associated with SA and PP and PG (PP*PG) is negatively

significantly associated with SA already reported earlier. But PP and PA in presence of PG put negative effects on SA, which is very complicated to understand.

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

xviii. The lower part of Table **2** shows non-parametric estimation of smoothing terms namely Blood lactic dehydrogenise (LDH) and Glucose. Both of these two cofactors enter 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 the factor LDH is significant with p-value 0.02 and Glucose is highly significant with p-value <0.001.

xix. It also noticed from Table **2** that, the GAM fitted model has an Adjusted R-square value approximately 0.60 with 65% of its deviance explained. The GCV (Generalized cross validation) score is 0.0230 which is also very low compare to other models.

From Table **2**, the final selected GAM fitted gamma distributed model of the Serum Albumin (y) is shown below

$$\begin{aligned}\hat{Z} = {} & -4.55 + 0.11 * AE2 + 0.04 * Hemoglobin2 + 0.03 * Diagnosis\ class2 + 0.00^{\#} * Age \\ & - 0.0018 * Duration\ AE - 0.0033 * Duration\ of\ symptom + 0.866 * TP \\ & + 0.00^{\#} * PLD + 0.67 * PP - 0.16 * PA + 0.064 * PG + 0.023 * CRP - 0.003 \\ & * (TP \cdot CRP) - 0.009 * (TP \cdot PG) - 0.00^{\#} * (PG \cdot CRP) - 0.10 * (TP \cdot PP) \\ & - 0.008 * (PP \cdot PG) + 0.037 * (PP \cdot PA) + 0.004 * (PA \cdot PG) - 0.00^{\#} \\ & * (TP \cdot PLD) - 0.00^{\#} * (PG \cdot PLD) + 0.00^{\#} * (TP \cdot PG \cdot CRP) + 0.001 \\ & * (TP \cdot PP \cdot PG) + 0.00^{\#} * (PP \cdot PA \cdot PG) + 0.00^{\#} * (TP \cdot PLD \cdot PG) + f(LDH) \\ & + f(Glucose)\end{aligned} \quad (2)$$

[#]denotes the value of an estimate whose first three decimal places are zeros and '*f*' denotes the smoothing function.

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

4. DISCUSSION

In Yao *et al.* (2014) [25] the authors tried to establish that Serum Albumin (SA) is significant prognostic factor for MPM disease patients. MPM is a highly aggressive malignant with a very short span of median survival with approximately 9-12 months [2]. Still no such universally accepted standard therapies have been developed. Conventional medical and surgical therapies are also not completely developed with efficiency. Therefore it is very important clinical and medical science research problem to identify the risk or prognostic factors for MPM disease. These are the motivation of the present article, in which we try to find a relationship between SA and the other cofactors described in the MPM dataset. A probabilistic modeling approach has been considered here using generalized additive model commonly known as GAM, with gamma distribution and 'log' link assumptions [32-34]. For performing regression analysis a response or dependent variable is required, but in this present dataset of MPM disease has no such continuous response variable has been given. Here SA serves the purpose. Yao *et al*. (2014) showed that the pretreatment serum albumin level is an independent prognostic indicator of overall survival (OS) for MPM patients [25]. They also reported that patients with hypoalbuminaemia (albumin level ≤ 35 g/l) had been associated with significantly worse survival than those with a normal albumin level. Not only for MPM, in the field of malignant disease, SA has been shown as an independent prognostic factor in several cancers [42-45].

The report also said the prognostic role of SA in MPM is emphasized because it is a simple, inexpensive and commonly performed laboratory test. This SA is measured as a part of liver function tests, which are routinely performed for patients.

These are the reasons why we took the chance to find the determinate for SA, those who are responsible for decreasing and as well as increasing the SA. Because if a factor gives a negative impact to SA, that means decrement in SA value where as in case of positive impact of a factor means the increment in SA value. Once these risk factors are identified then it could be easy for the medical and clinical researchers to develop the standard therapies of treatments for MPM disease. This is the major objective of the present article to develop a probabilistic model using the bio medical and demographic parameters. This present work reported very important finding in terms of cofactor which gave main and interaction effects on SA.

The present result showed that the patients who have been exposed to the asbestos (name of the factor is "asbestos exposure") during their life give a significant effect to the SA. It has been well known that asbestos exposure is one of the major reasons for MPM disease. Model reported this very efficiently. Hemoglobin range which is higher than normal is also determined as a significant factor for SA model fitting.

The patients having the hemoglobin more than the normal range have the higher value of SA than others. It is also reported that the patients do not having MPM disease have a higher SA value than MPM disease patients. These findings supported many medical researches and the clinical views [25, 46-48]. Age is not a significant factor in SA modeling and also it is not significant in MPM disease. Duration of asbestos exposure is a continuous variable in this present study measured in years is the most important hallmark and the leading cause for MPM disease. [46, 47]. Here duration of asbestos exposure is negatively significantly associated with SA, which indicates the increment of duration of asbestos exposure reduced the SA value, which indirectly infers the occurrence chance of MPM. This finding is very much important because it statistically (or mathematically) proves that, duration of asbestos exposure is one of the major causes for MPM disease. Similarly duration of symptoms of disease is negatively significantly associated with SA value, which means if duration of symptoms of the disease is increased then the SA value will be decreased and Yao *et al.* (2014) shows that lower level value of SA is an important prognostic factor for MPM.

So, our present model supports the earlier finding regarding MPM disease very prominently and strongly using this GAM regression technique [30, 32-33]. Total protein, also known as serum total protein, is a biochemical test for measuring the total amount of protein in serum. The reference range for total protein is typically 6.0-8.0g/dl. Concentrations below the reference range usually reflect low albumin concentration and may refers to liver disorder and kidney disorder. Elevated total protein may indicate: inflammation or infections, such as viral hepatitis B or C, or HIV and bone marrow disorders. There is so such evidence of relationship between MPM and total protein, but the present model shows a high positive association between SA and total protein. It means if the total protein is increased in serum it will be help to increase the SA value.

In this present work PLD or pleural lactic dehydrogenase, pleural glucose and pleural protein are found to be highly positively associated with SA value, which indicates that if these pleural fluids testing measures (PLD, PG & PP) are increased then SA value should be increased.

Medical research said that a low level of pleural glucose can be link to infection or malignancy [49, 50]. That means normal level or little higher than normal level PG patient has smaller chance to get infection or malignant, here our study shows that increment in PG ensures the increment in SA. Patients or persons with standard or normal SA value have smaller chance to get MPM.

The upper limit of the normal PLD or pleural lactic dehydrogenase is 200 IU/L. A high LD indicates that pleural fluid is an exudate, while a low level indicates it is transudate. Normal PP or pleural proteins count is less than 1-2 g/dL. Pleural effusions are classified as transudates or exudates on the basis of the fluid protein level, classically, a pleural fluid protein level >30g/l is an exudate and <30g/l is a transudate, in the context of a normal serum protein level [51]. So, clinically it is established that these pleural fluid measures are very sensitive in their own level, a little deviation from their normal ranges cause various diseases including malignant. Present study shows a mathematical relationship between these pleural fluid measures with the SA, it can help to maintain the normal level of each of these biomedical parameters.

C- reactive protein (CRP) has a positive significant association with SA, which showed in this present article. It indicates that if the CRP level is increased then SA value level is also increased. Few earlier researchers found that, CRP is an acute phase reactant which has been noted to be significantly elevated in patients with metastatic disease across a variety of solid organ and hematological malignancies, including malignant pleural Mesothelioma (MPM) [52].

In a retrospective study of 115 patients with a pathologically confirmed diagnosis of MPM, elevated CRP (≥1 mg/dL) was shown to be an independent indicator of poor prognosis (HR=2.07; 95% CI: 1.23-3.46; P=0.001) [53]. As per our knowledge the mathematical relationship founds from this present article with CRP and SA is new in literature. But very interesting result founds here that CRP value along with the total protein and pleural glucose has high negative significance association with SA. Which means the joint effect of CRP and TP is negatively significantly associated with SA. That is if both of these two increase at their level jointly then it will diminished the SA value. Same result can be shown for CRP and PG case also. The conclusion is very important that individually CRP gives a positive effect on SA, but in joint interaction it gives the negative effect to SA.

Similar things happened also in case of total protein, it has a positive significant association with SA

as a main effect, but in case of joint interaction with PG, PP and PLD, together they have a negative significant association with SA, that means if they increased their level jointly then the SA value should be decreased. Decrement in SA value form it normal range may play a very important role for MPM patients [25]. The joint interaction effect of pleural glucose (PG) with pleural protein (PP) and with PLD both are high negatively significantly associated with SA, but the joint of effect of PG and pleural albumin (PA) is positively significant, which means if both of these two factor are increased at their level then SA value is also increased. In main effect the factor PG is positively significant but PA is not significant. Therefore, these joint interaction effects are very important factors in MPM disease treatment or prognosis which is new in literature of medical research from mathematical modeling perspective.

Another important finding of this work is the third order interaction effects of the factors, which is very difficult to interpret literally. These four third order interaction effects – i) TP, PG and CRP ii) TP, PG and PP iii) PP, PG and PA iv) TP, PLD and PG have been occurred in this GAM model, which obviously predicts some important relationship between SA and them, but this is too complex to interpret. The third order interaction effects of (i), (ii) and (iv) are positively significantly associated with SA whereas (iii) is negatively significantly associated with SA.

Beside these another major part is incorporation of smoothing factors in this model which help to fit the model well enough. It also gives the stable estimate of the parameters (standard error of estimates in Table **2**) and eliminates the heteroscedasticity (non-constant variance response). From this part it could be found that lactate dehydrogenase (LDH) and glucose are the significant smoothing factors which have a nonlinear relationship with SA (from Table **2** and Figure **3a** and **b**). Lactate dehydrogenase (LDH) is a protein that helps to produce energy in the body. An LDH test measures the amount of LDH in the blood and the normal value range is 105 to 333 IU/L. LDH is found in many body tissues such as the heart, liver, kidney, skeletal muscle, brain, blood cells, and lungs. High LDH were found to be prognostic indicators in mesothelioma. [54].

In our work it shows that (from Figure **4**) a high amount of LDH (more than 600 IU/L) causes the decrement in SA.

So far in our knowledge these are the most fundamental findings of this present work which has not been done before by any researcher. Now it can be verified by the medical researchers and the practitioner in clinic.

5. CONCLUSION

This current article is tried to find a relationship between serum albumin (SA) and the others cofactors based on a well-known mesothelioma pleural malignant (MPM) disease dataset (see material part). Serum albumin is treated here as a response variable with gamma distribution as an assumption. The reason behind taking SA as a response variable is that, the pretreatment serum albumin level is an independent prognostic indicator of overall survival (OS) for MPM patients [25]. We tried to model this SA 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 SA 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 SA is selected based on locating the appropriate statistical distribution. The SA 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 probabilistic model under regression framework. The current models may provide a better assistance for researchers and the medical practitioner for developing standard treatment therapies and to make decision using the individual MPM patient's risk factors. The current results have focused many

interesting conclusions. These findings may help the medical practitioners for better medical treatment. Asbestos exposure, hemoglobin, disease diagnosis status are the significant categorical variables for serum albumin, whereas duration of asbestos exposure, duration of symptoms, total protein, PLD, PP, PG and CRP are the significant continuous variables for SA. The non-parametric estimation part of this model shows LDH and Glucose level are the significant smoothing terms. Additionally it is also found in parametric estimation part that, second and third order interactions of biochemical parameters are highly significant for this SA. Most of these present findings are partially as well as completely new in MPM research literature.

Finally, taking into consideration of all relevant results found from this work- it can be conclude that, serum albumin may play a very significant prognostic factor role for MPM disease and it is not only clinical perspective but also from mathematical ground. We can predict the SA value using the fitted model presented here (equation (2)) and this probabilistic model takes MPM disease research to a strong platform.

REFERENCE

[1] Wagner JC, Sleggs CA, Marchand P. Diffuse pleural mesothelioma and asbestos exposure in the North Western Cape Province. Br J Ind Med 1960; 17: 260-71. https://doi.org/10.1136/oem.17.4.260

[2] Favoni RE, Florio T. Combined chemotherapy with cytotoxic and targeted compounds for the management of human malignant pleural mesothelioma. Trends Pharmacol Sci 2011; 32: 463-79. https://doi.org/10.1016/j.tips.2011.03.011

[3] Dogan AU, Baris YI, DoganM, Emri S, Steele I, Elmishad AG, *et al*. Genetic predisposition to fiber carcinogenesis causes a mesothelioma epidemic in Turkey. Cancer Res.2006; 66: 5063-8. https://doi.org/10.1158/0008-5472.CAN-05-4642

[4] Zervos MD, Bizekis C, Pass HI. Malignant mesothelioma 2008. Current Opinion in Pulmonary Medicine 2008; 14: 303-309. https://doi.org/10.1097/MCP.0b013e328302851d

[5] Yazicioglu S, Ilcayto R, Balci K, Sayli BS, Yorulmaz B. Pleural calcification, pleural mesotheliomas and bronchial cancers caused by tremolite dust. Thorax1980; 35: 564-569. https://doi.org/10.1136/thx.35.8.564

[6] McConnochie K, Simonato L, Mavrides P, Christofides P, Pooley FD. Mesothelioma in Cyprus: the role of tremolite. Thorax 1987; 42: 342-347. https://doi.org/10.1136/thx.42.5.342

[7] Constantopoulos SH, Theodoracopoulos P, Dascalopoulos G, Saratzis N, Sideris K. Metsovo lung outside Metsovo. Chest 1991; 99: 1158-1161. https://doi.org/10.1378/chest.99.5.1158

[8] Nishimura SL, Broaddus VC. Asbestos-induced pleural disease. CliniesIn Chest Medicine 1998; 19: 311-329. https://doi.org/10.1016/S0272-5231(05)70079-4

[9] Metintas M, Ozdemir N, Hillerdal G, Ucgun I, Metintas S, *et al*. Environmental asbestos exposure and malignant pleural mesothelioma. Respiratory Medicine 1999; 93: 349-355. https://doi.org/10.1016/S0954-6111(99)90318-9

[10] Metintas S, Metintas M, Ucgun I, Oner U. Follow-up of a Turkish cohort living in a rural area. Chest 2002; 22: 2224-2229. https://doi.org/10.1378/chest.122.6.2224

[11] Metintas M, Metintas S, Ak G, Erginel S, Alatas F, Kurt E, *et al*. Epidemiology of pleural. mesothelioma in a population with non-occupational asbestos exposure. Respirology 2008; 13: 117-121. https://doi.org/10.1111/j.1440-1843.2007.01187.x

[12] Peto J, Decarli A, Vecchia La. C, Levi F, Negri E. The European mesothelioma epidemic. British Journal of Cancer 1999; 79: 666-672. https://doi.org/10.1038/sj.bjc.6690105

[13] Burgers JA, Damhuis RA. Prognostic factors in malignant mesothelioma. Lung Cancer 2004; 45: S49-54. https://doi.org/10.1016/j.lungcan.2004.04.012

[14] McDonald JC, McDonald AD. The epidemiology of mesothelioma in historical context. European Respiratory Journal 1996; 9: 1932-1942. https://doi.org/10.1183/09031936.96.09091932

[15] Spirtas R, Beebe GW, Connelly RR, Wright WE, Peters JM, *et al*. Recent trends in mesothelioma incidence in the United States. American Journal of Industrial Medicine 1986; 9: 397-407. https://doi.org/10.1002/ajim.4700090502

[16] Peto J, Hodgson JT, Matthews K, Jones JR. Continuing increase in mesothelioma mortality in Britain. Lancet 1995; 345: 535-539. https://doi.org/10.1016/S0140-6736(95)90462-X

[17] Leigh J, Corvalan CF, Grimwood A, Berry G, Ferguson DA, *et al*. The incidence of malignant mesothelioma in Australia 1982-1988. American Journal of Industrial Medicine 1991; 20: 643-655. https://doi.org/10.1002/ajim.4700200507

[18] National Mesothelioma committee. http://www.mesothelioma-tr.org (accessed November 10, 2014).

[19] Mesothelioma News (accepted: 29.06.11) http://www.meso-theliomanews.com/medical/mesothelioma-diagnosis/pleural mesothelioma

[20] Tanrikulu AC, Senyigit A, Dagli CE, Babayigit C, Abakay A. Environmental malignant pleural mesothelioma in Southeast Turkey. Saudi Medical Journal 2006; 27(10): 1605-1607.

[21] Senyiğit A, Bayram H, Babayiğit C, Topcu F, Nazaroğlu H, Bilici A, *et al*. Malignant pleural mesothelioma caused by environmental exposure to asbestos in the Southeast of Turkey: CT findings in 117 patients. Respiration 2000; 67(6): 615-622. https://doi.org/10.1159/000056290

[22] Senyiğit A, Babayiğit C, Gökirmak M, Topçu F, Asan E, Coşkunsel M, *et al*. Incidence of malignant pleural mesothelioma due to environmental asbestos fiber exposure in the southeast of Turkey. Respiration 2000; 67(6): 610-614. https://doi.org/10.1159/000056289

[23] OrhanEr, Tanrikulu AC, Abakay A, Temurtas F. An approach based on probabilistic neural network for diagnosis of Mesothelioma's disease. Computers & Electrical Engineering 2011; 38: 75-81. https://doi.org/10.1016/j.compeleceng.2011.09.001

[24] OrhanEr, Tanrikulu AC, Abakay A. Use of artificial intelligence techniques for diagnosis of malignant pleural mesothelioma. Dicle Medical Journal 2015; 42(1): 5-11. https://doi.org/10.5798/diclemedj.0921.2015.01.0520

[25] Yao ZH, Wan YY, Liu QH, Lin DJ, *et al*. Serum albumin as a significant prognostic factor in patients with malignant pleural mesothelioma. Tumor Biology 2014; 35(7): 6839-6845. https://doi.org/10.1007/s13277-014-1938-5

[26] Human serum albumin (From Wikipedia, the free encyclopedia) https://en.wikipedia.org/wiki/Human_serum_albumin

[27] R.H. Myers., D.C. Montgomery., G.G.Vining, Generalized Linear Models with Applications in Engineering and the Sciences. New York: John Wiley & Sons; 2002.

[28] 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.

[29] Das RN, Mukherjee S, Panda RN. Association between Body Mass Index and Cardiac Parameters of Worcester Heart Attack Study.BAOJ Cell Mol Cardio 2016; 2: 006.

[30] 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. https://doi.org/10.5530/jcdr.2017.4.31

[31] D. Ruppert., M.P.Wand., R.J.Carroll, Semi parametric Regression, first ed. Cambridge University Press New York; 2003. https://doi.org/10.1017/CBO9780511755453

[32] T.Hastie., R.Tibshirani, Generalized additive models. John Wiley & Sons, Inc.; 1990.

[33] Hastie T, Tibshirani R. Generalized additive models for medical research. Statistical Methods in Medical Research 1995; 4: 187-196. https://doi.org/10.1177/096228029500400302

[34] SN.Wood, Generalized Additive Models: An Introduction with R. London: Chapman and Hall; 2006. https://doi.org/10.1201/9781420010404

[35] Currie ID, Durban M, Eilers PH. Generalized linear array models with applications to multidimensional smoothing. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2006; 68(2): 259-80. https://doi.org/10.1111/j.1467-9868.2006.00543.x

[36] P.J.Green., B.W. Silverma, Nonparametric regression and generalized linear models: a roughness penalty approach. CRC Press; 1993. https://doi.org/10.1201/b15710

[37] Ruppert D. Selecting the number of knots for penalized splines. Journal of computational and graphical statistics 2002; 11(4): 735-57. https://doi.org/10.1198/106186002853

[38] Eilers PH, Marx BD. Flexible smoothing with B-splines and penalties. Statistical science 1996; 1: 89-102. https://doi.org/10.1214/ss/1038425655

[39] S.Chatterjee., A.S.Hadi, Regression Analysis by Example, fifth ed. John Wiley & Sons, New Jersey; 2006. https://doi.org/10.1002/0470055464

[40] Nelder JA, Lee Y. Generalized linear models for the analysis of Taguchi-type experiments. Applied Stochastic Models and Data Analysis 1991; 7: 107-120. https://doi.org/10.1002/asm.3150070110

[41] Y.Lee., J.A.Nelder., Y.Pawitan, Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood. Chapman & Hall, London (2006). https://doi.org/10.1201/9781420011340

[42] Espinosa E, Feliu J, Zamora P, Gonzalez Baron M, Sanchez JJ, Ordonez A, *et al*. Serum albumin and other prognostic factors related to response and survival in patients with advanced non-small cell lung cancer. Lung Cancer 1995; 12: 67-76. https://doi.org/10.1016/0169-5002(95)00407-R

[43] Lu HJ, Chen KW, Tzeng CH, Liu JH, Chiou TJ, Yen CC, *et al*. Evaluation of prognosis for carcinoma of unknown origin in elderly patients. Oncology 2012; 83: 24-30. https://doi.org/10.1159/000337983

[44] Wang CY, Hsieh MJ, Chiu YC, Li SH, Huang HW, Fang FM, *et al*. Higher serum C-reactive protein concentration and hypoalbuminemia are poor prognostic indicators in patients with esophageal cancer undergoing radiotherapy. RadiotherOncol 2009; 92: 270-5. https://doi.org/10.1016/j.radonc.2009.01.002

[45] Lambert JW, Ingham M, Gibbs BB, Given RW, Lance RS, Riggs SB. Using preoperative albumin levels as a surrogate marker for outcomes after radical cystectomy for bladder cancer. Urology 2013; 81: 587-92. https://doi.org/10.1016/j.urology.2012.10.055

[46] Spirtas R, Heineman EF, Bernstein L, *et al* Malignant mesothelioma: attributable risk of asbestosexposure. Occupational and Environmental Medicine 1994; 51: 804-811. https://doi.org/10.1136/oem.51.12.804

[47] Whitwell F, Rawcliffe RM Diffuse malignant pleural mesothelioma and asbestos exposure Thorax 1971; 26: 6-22. https://doi.org/10.1136/thx.26.1.6

[48] Berardi R, Fiordoliva I, De Lisa M, Ballatore Z, *et al* Clinical and pathologic predictors of clinical outcome of malignant pleural mesothelioma. Tumori 2016; 102(2): 190-5. https://doi.org/10.5301/tj.5000418

[49] Momenin N, Colletti PM, Kaptein EM. Low pleural fluid-to-serum glucose gradient indicatespleuroperitoneal communication in peritoneal dialysis patients: presentation of two cases and a review of the literature. Nephrol Dial Transplant 2012; 27: 1212-1219. https://doi.org/10.1093/ndt/gfr393

[50] Limthongkul S. The pathogenesis of low pleural fluid glucose in acidotic malignant pleural effusions. J Med Assoc Thai. 1989; 72(9): 492-7.

[51] Khaleeq G, Musani AI. Emerging paradigms in the management of malignant pleural effusions. Respir Med 2008; 102(7): 939-48. https://doi.org/10.1016/j.rmed.2008.01.022

[52] Nojiri S1, Gemba K, Aoe K, Kato K, Yamaguchi T, Sato T, Kubota K, Kishimoto T. Survival and prognostic factors in malignant pleural mesothelioma: a retrospective study of 314 patients in the west part of Japan. Jpn J ClinOncol. 2011; 41(1): 32-9. https://doi.org/10.1093/jjco/hyq159

[53] Ghanim B, Hoda MA, Winter MP, Berger LW. Pretreatment Serum C-Reactive Protein Levels Predict Benefit From Multimodality Treatment Including Radical Surgery in Malignant Pleural Mesothelioma. Annals of surgery 2012; 256(2): 357-62. https://doi.org/10.1097/SLA.0b013e3182602af4

[54] Zhuo Y, Lin L, Wei S, Zhang M. Pretreatment elevated serum lactate dehydrogenase as a significant prognostic factor in malignant mesothelioma- A meta-analysis. Medicine (Baltimore) 2016; 95(52): e5706. https://doi.org/10.1097/MD.0000000000005706

[55] Mukherjee S. Malignant Mesothelioma Disease Diagnosis using Data Mining Techniques. Applied Artificial Intelligence 2018; 32(3): 293-308. https://doi.org/10.1080/08839514.2018.1451216

Received on 20-11-2020 Accepted on 17-12-2020 Published on 31-12-2020

https://doi.org/10.6000/1929-6029.2020.09.08