Quest Journals
Journal of Research in Applied Mathematics
Volume 3 ~ Issue 5 (2017) pp: 01-10
ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

Generalized Additive and Generalized Linear Modeling for Children Diseases

Neha Jain, Roohi Gupta and Rahul Gupta*
Rahul Gupta Department of Statistics, University Of Jammu, Jammu, J and K, India

Received 03 Feb, 2017; **A**ccepted 18 Feb, 2017 © The author(s) 2017. **P**ublished with open access at **www.questjournals.org**

ABSTRACT: *This paper is necessarily restricted to application of Generalised Linear Models(GLM) and Generalised Additive Models(GAM), and is intended to provide readers with some measure of the power of these mathematical tools for modeling Health/Illness data systems. We are all aware that illness, in general and children illness, in particular is amongst the most serious socio-economic and demographic problems in developing countries, and they have great impact on future development. In this paper we focus on some frequently occurring diseases among children under fourteen years of age, using data collected from various hospitals of Jammu district from 2011 to 2016.The success of any policy or health care intervention depends on a correct understanding of the socio economic environmental and cultural factors that determine the occurrence of diseases and deaths. Until recently, any morbidity information available was derived from clinics and hospitals. Information on the incidence of diseases, obtained from hospitals represents only a small proportion of the illness, because many cases do not seek medical attention .Thus, the hospital records may not be appropriate from estimating the incidence of diseases from programme developments. The use of DHS data in the understanding of the childhood morbidity has expanded rapidly in recent years. However, few attempts have been made to address explicitly the problems of non linear effects on metric covariates in the interpretation of results .This study shows how the GAM model can be adapted to extent the analysis of GLM to provide an explanation of non linear relationship of the covariate. Incorporation of non linear terms in the model improves the estimates in the terms of goodness of fit. The GLM model is explicitly specified by giving symbolic description of the linear predictor and a description of the error distribution and the GAM model is fit using the local scoring algorithm, which iteratively fits weighted additive models by back fitting. The back fitting algorithm is a Gauss-Seidel method of fitting additive models by the iteratively smoothing partial residuals. The algorithm separates the parametric from the non parametric parts of the fit, and fits the parametric part using weighted linear least squares within the back fitting algorithm.*

Keywords: *Generlised additive model, Generalised linear model, weighted linear least squares*

I. INTRODUCTION

Generalized additive model (GAM) is a generalized linear model in which the linear predictor depends linearly on unknown smooth functions of some predictor variables, and interest focuses on inference about these smooth functions. GAMs were originally developed by Trevor Hastie and Robert Tibshirani to blend properties of generalized linear models with additive models. Generalized linear model (GLM) is a flexible generalization of ordinary linear regression that allows for response variables that have error distribution models other than a normal distribution. The GLM generalizes linear regression by allowing the linear model to be related to the response variable via a link function and by allowing the magnitude of the variance of each measurement to be a function of its predicted value.Generalized linear models were formulated by John Nelder and Robert Wedderburn as a way of unifying various other statistical models, including linear regression , logistic regression and Poisson regression . They proposed an iteratively reweighted least squares method for maximum likelihood estimation of the model parameters. Maximum-likelihood estimation remains popular and is the default method on many statistical computing packages. Other approaches, including Bayesian approaches and least squares fits to variance stabilized responses, have been developed. Significant statistical development in the last three decades has been the advances in regression analysis provided by generalized additive models (GAM) and generalized linear models (GLM).These three alphabet acronyms translate into a great scope for application in many areas of applied scientific research. Based on developments by Cox and Snell[1] in the late sixties, the first seminal publications, also providing the link with practice (through software availability), were

those of Nelder and Wedderburn[2] and Hastie and Tibshirani[3]. Since their development, both approaches have been extensively applied in medical and health related research, as evidenced by the growing number of published papers incorporating these modern regression tools.

mathematical extensions of linear models that do not force data into unnatural scales, and thereby allow for non-linearity and non-constant variance structures in the data (Hastie and Tibshirani, [3]). They are based on an assumed relationship between the mean of the response variable and the linear combination of the explanatory variables. Data may be assumed to be from several families of probability distributions, including the normal, binomial, Poisson, negative binomial, or gamma distribution, many of which better fit the non-normal error structures of most ecological data. Thus, GLMs are more flexible and better suited for analyzing relationships, which can be poorly represented by classical Gaussian distributions (see Austin[4]). GAMs (Hastie and Tibshirani[3]) are semi-parametric extensions of GLMs; the only underlying assumption made is that the functions are additive and that the components are smooth. A GAM, like a GLM, uses a link function to establish a relationship between the mean of the response variable and a 'smoothed' function of the explanatory variable(s). The strength of GAMs is their ability to deal with highly non-linear and non-monotonic relationships between the response and the set of explanatory variables. GAMs are sometimes referred to as data- rather than model driven. This is because the data determine the nature of the relationship between the response and the set of explanatory variables rather than assuming some form of parametric relationship (Yee and Mitchell [5]. Like GLMs, the ability of this tool to handle non-linear data structures can aid in the development of models that better represent the underlying data, and hence increase our understanding of real life systems. Few syntheses of GLMs and GAMs have been made since the first papers encouraged their use in environmental studies (Austin and Cunningham[6] and Nicholls[7]).

This work is necessarily restricted to application of GLMs and GAMs, and is intended to provide readers with some measure of the power of these statistical tools for modeling Health/Illness data systems. We are all aware that illness, in general and children illness, in particular is amongst the most serious socio-economic and demographic problems in developing countries, and they have great impact on future development. Demographic and health surveys are designed to collect data on health and nutrition of children and mother as well as on fertility and family planning. The discovery of some vaccination, during the last decade, has reduced morbidity and mortality in most cases. Despite this, some diseases are still the major cause of death in childhood .In this paper we focus on some frequently occurring diseases among children under fourteen years of age, using data collected from various hospitals of Jammu district(J and K State, India) from 2011 to 2016.The success of any policy or health care intervention depends on a correct understanding of the socio economic environmental and cultural factors that determine the occurrence of diseases and deaths. Until recently, any morbidity information available was derived from clinics and hospitals. Information on the incidence of diseases, obtained from hospitals represents only a small proportion of the illness, because many cases do not seek medical attention .Thus, the hospital records may not be appropriate from estimating the incidence of diseases from program developments. The use of DHS data in the understanding of the childhood morbidity has expanded rapidly in recent years.However, few attempts have been made to address explicitly the problems of non linear effects on metric covariates in the interpretation of results .This study shows how the GAM model can be adapted to extent the analysis of GLM to provide an explanation of non linear relationship of the covariate. Incorporation of non linear terms in the model improves the estimates in the terms of goodness of fit. The GLM model is explicitly specified by giving symbolic description of the linear predictor and a description of the error distribution and the GAM model is fit using the local scoring algorithm, which iteratively fits weighted additive models by back fitting. The back fitting algorithm is a Gauss-Seidel method of fitting additive models by the iteratively smoothing partial residuals. The algorithm separates the parametric from the non parametric parts of the fit, and fits the parametric part using weighted linear least squares within the back fitting algorithm.The rest of the paper is organized as follows. Section II proposes model descriptions and estimation procedure applied based on Generalized Additive Models (GAM). Section III presents the outcomes obtained and compares the result based on GLM and GAM. Finally, Section IV summarizes and concludes.

II. DESCRIPTION OF MODEL AND SIGNIFICANCE

To extend the additive model to a wide range of distribution families, Hastie and Tibshirani [3] proposed generalized additive models. These models assume that the mean of the dependant variable depends in additive predictor through a non linear link function. Generalized additive models permit the response probability distribution to be any member of the exponential family of distribution. Many widely used statistical models belong to this general class , including additive models from Gaussian data , non parametric logistic models for binary data and non parametric log-linear models for Poisson data.In GLM, the dependent variable values are predicted from a linear combination of predictor variables, which are "connected" to the dependent variable via a link function .Let Y be a response random variable and $X_1, \ldots\ldots, X_p$ be a set of predictor variables.

In generalized linear model a response variable Y can be viewed as a method for estimating for the value of Y depends on the value of $X_1, \ldots\ldots, X_p$.
The generalized linear model is assumed to be
$E(Y) = f(X_1, \ldots\ldots, X_p) = g(\beta_0 + \beta_1 X_1 + \cdots + \beta_P X_P)$, where g(.) is known as link function .
Given a sample of values for Y and X, estimates of $\beta_0, \beta_1, \ldots\ldots, \beta_P$ are often obtained by the least squares method or maximum likelihood method.The additive model generalizes the linear model by modeling expected value of Y as

$$E(Y) = f(X_1, \ldots\ldots, X_p) = S_0 + S_1(X_1) + \cdots + S_P(X_P)$$

where $S_i(X_i)$, i=1,----,p are smooth functions .

The usual linear function covariate $\beta_j X_j$ is replaced with $S_i(X)$, an unspecified smooth function. These functions are not given a parametric form but instead are estimated in a non parametric fashion. In addition, the additive models require specification of the smooth functioning using as a scatter plot smoother such as Loess (a locally weighted regression smoother), running mean or a smooth spline. The scatter plot smoother used in this application of the additive model is the cubic β-spline. The degree of smoothing in a scatter plot smoother, for example in a Loess, is controlled by the span, which is the proportion of points contained in each neighborhood (the set of X values within a defined distance to X_j). The resulting 'smooths'characterizes the trend of the response variable as a function of the predictor variables.

The algorithm for generalized additive models is a little more complicated. Generalized additive models (GAM) extend generalized linear models in the same manner as additive models extend linear regression models, that is ,by replacing the linear form $\alpha + \sum_j X_j(\beta_j)$ with the additive form $\alpha + \sum_j S_j(\beta_j)$.

The fitting of the GAM is an iterative looping process involving the scatter plot smooth,the back fitting algorithm, and the local scoring algorithm, a generalization of the Fisher scoring procedure in a GLM. Each iterations of the local scoring algorithm produces a new working response and weights that are directed back to the backfitting algorithm which produces a new additive predictor using the scatterplot smoother . The back fitting and local scoring algorithms consider the estimation of the smoothing term S_k the additive model. Many ways are available to approach the formulation and estimation of additive models. The back fitted algorithm is a general algorithm that can fit an additive model using any regression-type smoothers.
Define the jth set of partial residuals as

$$R_j = Y - S_0 - \sum_{k \neq j} S_k X_k$$

The partial residuals removes the effects of all the other variables from j ; therefore they can be used to model of effects against X_j. This is the foundation for the back fitting algorithm , providing a way for estimating each smoothing function $S_j(.)$ given estimates $\{ S_i(.), i \neq j\}$; for all the others . The back fitting algorithm iterative ,starting with initial functions $S_0, \ldots., S_P$ and iteration cycling through the partial residuals , fitting the individual smoothing components to its partial residuals .iteration proceeds until the individual components do not change . The algorithm so far described fits just additive models.

In the same way, estimation of the additive terms for generalized additive models is accomplished by replacing the weighted linear regression for the adjusted dependent variable by the weighted back fitting algorithm, essentially fitting a weighted additive model. The algorithm used in the case is called the local scoring algorithm .it is also an iterative algorithm and starts with initial estimates of $S_0, \ldots., S_P$. During iteration, an adjusted dependent variable and a set weight are computed, and then the smoothing components are estimated using a weighted back fitting algorithm. The scoring algorithm stops when the deviance of the estimates ceases to decrease.

Overall, then the estimating procedure for generalized models consists of two loops. Inside each step of the local scoring algorithm (outer loop), a weighted back fitting algorithm (inner loop) is used until convergence. Then, based on the estimates from this weighted back fitting algorithm, a new set of weights is calculated and the next iteration of the scprong algorithm starts. Any non- parametric smoothing method can be used to obtain $s_j(x)$. The GAM procedure implements the β- spline and local regression methods for univariate smoothing components and the thin-plate smoothing spline for bivariate smoothing components.
A unique aspect of generalized additive models is the non- parametric functions of the predictor variables. Hastie and Tibshirani[3] discuss various general scatter plot 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. Onse such scatter plot smoother is the cubic smoothing splines smoother, which generally produces a smooth generalization of the relationship between the two variables in the scatter plot. Computational details regarding this smoother can be found in Hastie and Tibshirani[3].

A step –wise GAM is performed to determine the best fitting model based on the criteria of the lowest (Akaike Information Criterion) test statistic which is a function and the effective member of parameters being estimated. The AIC in the step –wise GAM (Hastie[8] is calculated as

$$AIC=D+2df\varphi$$

where D= Deviance (residual sum of squares), df= effective degrees of freedom, and φ = dipersion parameter(variance).

The model with the lowest AIC is considered to have the best number of parameters to include in the final model. The deviance estimated in the model, analogous to the residual sum of squares, is a measure of the fit of the model.a pseudo coefficient of determination R^2, is estimated as 1.0 minus the ratio of the deviance of the model to the deviance of the null model.

Bayesian information criterion (BIC) or Schwarz criterion (also SBC, SBIC) is an alternative criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is based, in part, on the likelihood function and it is closely related to the Akaike information criterion (AIC). When fitting models, it is possible to increase the likelihood by adding parameters, but doing so may result in overfitting. Both BIC and AIC resolve this problem by introducing a penalty term for the number of parameters in the model; the penalty term is larger in BIC than in AIC. The BIC was developed by Gideon E. Schwarz and published in a 1978 paper, where he gave a Bayesian argument for adopting it.

The BIC is defined as

$$BIC = -2 \ln\hat{L} + K \ln(n),$$ where x = the observed data; Θ = the parameters of the model; n = the number of data points in x, the number of observations, or equivalently, the sample size; k = the number of free parameters to be estimated. If the model under consideration is a linear regression, k is the number of regressors, including the intercept; $\hat{L}$ = the maximized value of the likelihood function of the model M , i.e. $\hat{L} = p(\frac{x}{\hat{\theta}}, M)$, where $\hat{\theta}$ are the parameter values that maximize the likelihood function.

III. MODELLING AND DATA ANALYSIS FOR JAMMU DISTRICT

It is believed that the children disease cause degradation in the nutritional state and that successive episode may compromise physical development of infants, leading to malnutrition. However, the risk that under nourished children are more likely to develop diseases is as yet inconclusive. Some diseases affects mainly children in their first year of life but especially at weaning age. During this period a higher mortality rate is observed and the nutritional consequences are more serious. In this study data related to Children affected by diseases like Acute Gastroenteritis(AGE), Thallesemia, Bronchitis , Seizure and Anemia was collected and analysed for providing the best model in Jammu District, constituting its eight blocks namely Akhnoor, Khour, Bhalwal, R S Pura, Satwari, Jammu ,Kot Bhalwal and Marh.Diseases situation in each block is not same. Division is one of the most independent variable for this study.The following tables shows an overall scenario of these diseases in Jammu District children by blocks.

Table 1: Total Number And Percentage Of Acute Gastroenteritis(Age) In Jammu District By Blocks.

			HAD AGE	NO AGE	TOTAL
JAMMU DISTRICT	AKHNOOR	COUNT(%)	412(31.69%)	888(68.30%)	1300(100%)
	KHOUR	COUNT(%)	104(20.55%)	402(79.44)	506(100%)
	BHALWAL	COUNT(%)	87(25.51%)	254(74.48%)	341(100%)
	SATWARI	COUNT(%)	206(18.10%)	932(81.89%)	1138(100%)
	R S PURA	COUNT(%)	446(32.08%)	944(67.91%)	1390(100%)
	JAMMU	COUNT(%)	151(20.13%)	599(79.86%)	750(100%)
	DANSAL	COUNT(%)	258(29.35%)	621(70.64%)	879(100%)
	MARH	COUNT(%)	336(31.81%)	720(68.72%)	1056(100%)
TOTAL			2213(29.275)	5347(70.72%)	7560(100%)

From Table 1, we see that Akhnoor and Marh blocks are more affected area than other six blocks in Jammu District. Satwari and Jammu blocks are less affected area with AGE as compared to other divisions. Again, percentage of occurring AGE in rural area is higher than in urban area.

Table 2: Total Number And Percentage Of Thallesemia In Jammu District By Blocks.

			HAD Thallesemia	NO Thallesemia	Total
JAMMU DISTRICT	AKHNOOR	COUNT(%)	127(14.03%)	1173(85.96%)	1300(100%)
	KHOUR	COUNT(%)	71(17%)	435(82.99%)	506(100%)
	BHALWAL	COUNT(%)	58(17.75%)	283(82.24%)	341(100%)
	SATWARI	COUNT(%)	202(7.84%)	936(92.15%)	1138(100%)
	R S PURA	COUNT(%)	109(8.53%)	1281(91.46%)	1390(100%)
	JAMMU	COUNT(%)	64(10.12%)	686(89.81%)	750(100%)
	DANSAL	COUNT(%)	89(11.45%)	790(88.54%)	879(100%)
	MARH	COUNT(%)	121(9.76%)	935(90.23%)	1056(100%)
TOTAL			1003(13%))	6557(86.73%)	7560(100%)

From Table 2, we can see that Bhalwal and Akhnoor are highly affeted areas of Thallesemia than other blocks. Satwari is least affected amongst the other blocks. The probability of occurring Thallesemia for rural and urban area has no significant difference.

Table 3 Total Number And Percentage Of Bronchitis In Jammu District By Blocks.

JAMMU DISTRICT			HAD Bronchitis	NO Bronchitis	TOTAL
	AKHNOOR	COUNT(%)	80(6.15%)	1220(93.84%)	1300(100%)
	KHOUR	COUNT(%)	31(6.12%)	475(93.87)	506(100%)
	BHALWAL	COUNT(%)	35(10.26%)	306 (89.73%)	341(100%)
	SATWARI	COUNT(%)	71(6.27%)	1067(93.76%)	1138(100%)
	R S PURA	COUNT(%)	62(4.46%)	1328(95.53%)	1390(100%)
	JAMMU	COUNT(%)	26(3.46%)	724(96.53%)	750(100%)
	DANSAL	COUNT(%)	46(5.23%)	833(94.76%)	879(100%)
	MARH	COUNT(%)	44(4.16%)	1012(95.83%)	1056(100%)
TOTAL			433(5.72%)	7127(94.72%)	7560(100%)

From Table 3, we see that the maximum number of cases of Bronchitis came from Bhalwal while Jammu block is the least affected area of Bronchitis. It is more common in rural area than in urban areas.

Table 4 Total Number And Percentage Of Seizure In Jammu District By Blocks.

JAMMU DISTRICT			HAD Seizure	NO Seizure	TOTAL
	AKHNOOR	COUNT(%)	107(8.23%)	1193(91.76%)	1300(100%)
	KHOUR	COUNT(%)	59(11.66%)	447(88.33%)	506(100%)
	BHALWAL	COUNT(%)	38(11.14%)	293 (88.59%)	341(100%)
	SATWARI	COUNT(%)	96(8.43%)	1042(91.56%)	1138(100%)
	R S PURA	COUNT(%)	94(6.76%)	1286(93.23%)	1390(100%)
	JAMMU	COUNT(%)	58(7.73%)	692(92.26%)	750(100%)
	DANSAL	COUNT(%)	66(7.50%)	813(92.44%)	879(100%)
	MARH	COUNT(%)	69(6.53%)	987(93.46%)	1056(100%)
TOTAL			669(8.54%)	6891(91.15%)	7560(100%)

From Table 4, Khour and Bhalwal are highly affected from Seizure than other blocks. Marh and R S Pura had least impact of Seizure amongst the rest of the blocks.

Table 5 Total Number And Percentage Of Anaemia In Jammu District By Blocks.

JAMMU DISTRICT			HAD Anaemia	NO Anaemia	TOTAL
	AKHNOOR	COUNT(%)	74(5.69%)	1226(94.30%)	1300(100%)
	KHOUR	COUNT(%)	41(8.10%)	465(91.81%)	506(100%)
	BHALWAL	COUNT(%)	23(6.74%)	318 (93.25%)	341(100%)
	SATWARI	COUNT(%)	63(5.53%)	1075(94.46%)	1138(100%)
	R S PURA	COUNT(%)	79(5.68%)	1311(94.31%)	1390(100%)
	JAMMU	COUNT(%)	44(5.86%)	706(94.13%)	750(100%)
	DANSAL	COUNT(%)	55(2.51%)	824(93.74%)	879(100%)
	MARH	COUNT(%)	56(5.30%)	1000(94.69%)	1056(100%)
TOTAL			465(6.15%)	7095(93.54%)	7560(100%)

From Table 5, Anaemia is highest in Khour block and least in Dansal block.

Analyzing the above tables we see that the children in rural areas of Jammu District are more prone to diseases than that of urban areas. This may be due to poor hygiene, malnutrition, lack of awareness in mother etc. To get an overall scenario of these diseases with different covariates we explore these by modeling.In this study, there different models are used for analyzing occurrence of these diseases in Jammu district of Jammu and Kashmir. Model 1 is a generalized linear model where we consider sex, residence, division and season with the diseases. In model 2, we added one more independent variable child age with model 1. Model 1 and Model 2 are computed using Poisson distribution . In model 3 we use ordinal logistic distribution.

Table 6A comparison of Different Models Of The Bronchitis Disease In Children Less Than 14 Years Old In Jammu District

	MODEL 1	MODEL 2	MODEL 3
INTERCEPT	**-3.139**	**-3.299**	**3.483**
SEX			
MALE	**-0.128**	**-0.130**	**0.238**
FEMALE	-	-	-
RESIDENCE			
URBAN	**-0.749**	**-0.734**	**-1.139**
RURAL	-	-	-
SEASON			
SUMMER	**1.116**	**1.115**	**1.169**

WINTER	-	-	-
ZONE			
NORTH	**-0.300**	**-0.287**	**0.446**
EAST	**-1.622**	**-1.614**	**-1.467**
SOUTH	**0.834**	**-0.820**	**-0.536**
CENTRAL	**0.283**	**0.283**	**0.728**
WEST	-	-	-
CHILD AGE			
0-5 AGE GRP	-	**0.344**	**0.367**
6-10 AGE GRP		**0.092**	**0.097**
11-14 AGE GRP		-	-
AIC	**5.241E3**	**5.259E3**	**5.175E3**
BIC	**5.301E3**	**5.417E3**	**5.334E3**

In this analysis, we see that probability of occurring Bronchitis in summer season is more than in winter season. The probability of occurring Bronchitis in rural and urban areas has no significant difference. We also see that occurring Bronchitis in south and central zone of Jammu district is higher than rest of the zones. we see that AIC for model 1 is greater than AIC for model 3 which means Model 3 interprets the data quite well and generalized additive model fits well and explain more information than generalized linear models.

Table 7: A Comparison Of Different Models Of The Seizure Disease In Children Less Than 14 Years Old In Jammu District

	MODEL 1	**MODEL 2**	**MODEL 3**
INTERCEPT	**-2.115**	**-2.133**	**2.006**
SEX			
MALE	**-0.026**	**-0.028**	**-0.031**
FEMALE	-	-	-
RESIDENCE			
URBAN	**-0.381**	**-0.366**	**-0.400**
RURAL	-	-	-
SEASON			
SUMMER	**-0.366**	**-0.370**	**-0.412**
WINTER	-	-	-
ZONE			
NORTH	**0.252**	**0.272**	**0.306**
EAST	**-1.430**	**-1.412**	**-1.499**
SOUTH	**0.244**	**0.263**	**0.296**
CENTRAL	**0.209**	**0.213**	**0.230**
WEST	-	-	-
CHILD AGE			
0-5 AGE GRP	-	**-0.005**	**-0.005**
6-10 AGE GRP		**0.100**	**0.112**
11-14 AGE GRP		-	-
AIC	**8.332E3**	**8.345E3**	**8.213E3**
BIC	**8.392E3**	**8.503E3**	**8.372E3**

In this analysis, we see that probability of occurring Seizure in summer season and winter season has no significance difference. The probability of occurring Seizure in rural and urban areas has no significant difference. We also see that occurring Seizure in south and north zone of Jammu district is higher than rest of the zones. we see that AIC for model 1 is greater than AIC for model 3 which means Model 3 interprets the data quite well and generalized additive model fits well and explain more information than generalized linear models.

Table 8 A Comparison Of Different Models Of The Age Disease In Children Less Than 14 Years Old In Jammu District

	MODEL 1	**MODEL 2**	**MODEL 3**
INTERCEPT	**-1.051**	**-1.047**	**0.574**
SEX			
MALE	**-0.258**	**-0.257**	**-0.383**
FEMALE	-	-	-
RESIDENCE			
URBAN			
RURAL	**-0.022**	**-0.028**	**-0.050**
	-	-	-
SEASON			
SUMMER	**-0.296**	**-0.293**	**-0.435**
WINTER	-	-	-
ZONE			
NORTH	**0.032**	**0.034**	**0.039**
EAST	**-1.076**	**-1.084**	**-1.314**
SOUTH	**0.489**	**0.482**	**0.721**
CENTRAL	**0.219**	**0.217**	**0.307**
WEST	-	-	-

CHILD AGE			
0-5 AGE GRP	-	-0.019	-0.030
6-10 AGE GRP		-0.029	-0.045
11-14 AGE GRP		-	-
AIC	1.774E4	1.775E4	1.622E4
BIC	1.780E4	1.791E4	1.638E4

In this analysis, we see the occurrence of AGE is higher in east and south zone as compared to other zones. The probability of occurring AGE for rural and urban areas has no significance differences.We see that Residual Degrees of Freedom and Residual Deviance for smooth analysis is less that without smooth analysis and AIC for model 1 is greater that AIC of model 3 which means model 3 interprets the data quite well and generalized additive model fits well and explain more information than generalized linear models.

Table 9: A Comparison Of Different Models Of The Thallesemia Disease In Children Less Than 14 Years Old In Jammu District

	MODEL 1	MODEL 2	MODEL 3
INTERCEPT	-3.022	-3.030	3.032
SEX			
MALE	0.186	0.186	0.215
FEMALE	-	-	-
RESIDENCE			
URBAN	0.981	0.977	1.115
RURAL	-	-	-
SEASON			
SUMMER	0.291	0.291	0.351
WINTER	-	-	-
ZONE			
NORTH	0.244	0.241	0.266
EAST	2.018	2.020	2.635
SOUTH	0.284	0.279	0.310
CENTRAL	-0.891	-0.891	-1.012
WEST	-	-	-
CHILD AGE			
0-5 AGE GRP	-	0.057	0.070
6-10 AGE GRP		0.013	0.017
11-14 AGE GRP		-	-
AIC	1.041E4	1.043E4	1.007E4
BIC	1.047E4	1.059E4	1.023E4

Table 10: A Comparison Of Different Models Of The Anaemia Disease In Children Less Than 14 Years Old In Jammu District

	MODEL 1	MODEL 2	MODEL 3
INTERCEPT	-5.962	-5.890	7.592
SEX			
MALE	0.524	0.508	0.177
FEMALE	-	-	-
RESIDENCE	-0.598		
URBAN	-	-0.616	0.982
RURAL		-	-
SEASON			
SUMMER	1.405	1.400	1.370
WINTER	-	-	-
ZONE			
NORTH	-1.305	-1.299	0.416
EAST	-0.179	-0.194	-18.886
SOUTH	-0.099	-0.097	0.099
CENTRAL	0.032	-0.054	-0.348
WEST	-	-	-
CHILD AGE			
0-5 AGE GRP	-	0.812	0.818
6-10 AGE GRP		0.783	0.788
11-14 AGE GRP		-	-
AIC	1.058E3	1.065E3	1.065E3
BIC	1.118E3	1.224E3	1.224E3

We also estimated a logistic GAM with smoothing applied to the major of child age. At this stage , we could either conduct a series of likelihood ratio test or plot the non parametric estimate and inspect that for non linearity.

Visual inspection of the plot may be enough to understand which terms are non linearly related and non parametric estimate. The visual test is quite clear that child age is non linearly related .

Figure 1: Generalized additive model for Anemia disease in children 0-14 age group in Jammu district as a function of child age.

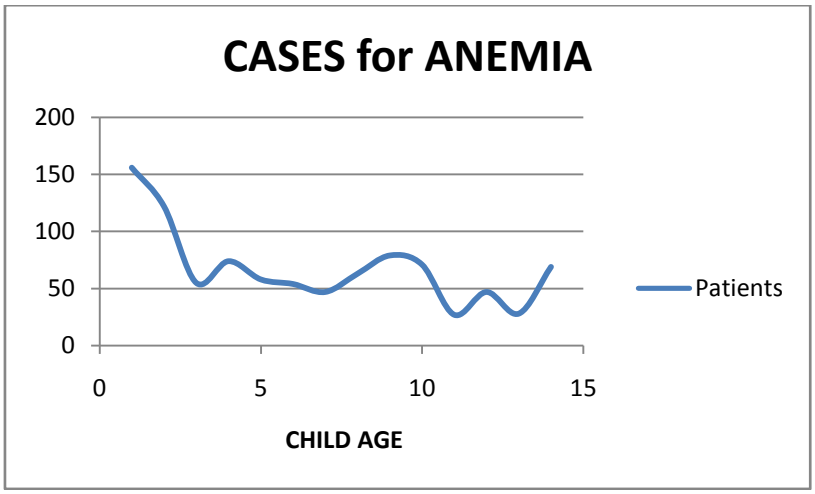


Figure 2: Generalized additive model for Seizure disease in children 0-14 age group in Jammu district as a function of child age.

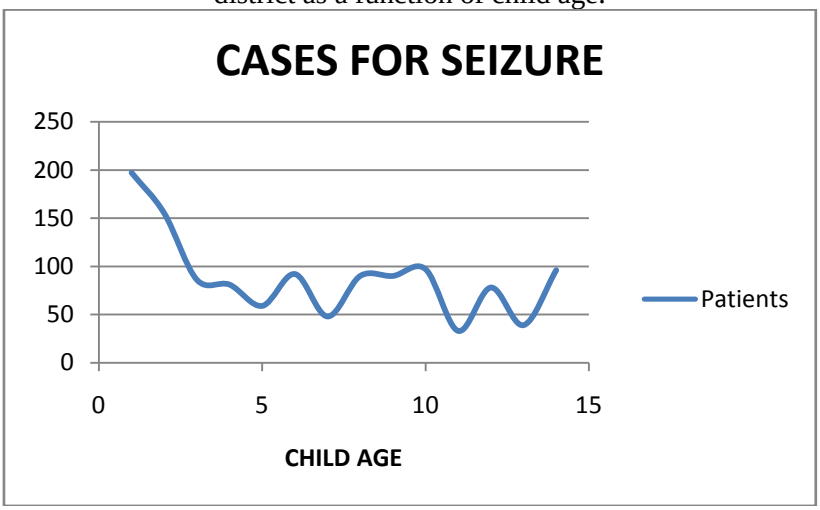


Figure 3: Generalized additive model for Bronchitis disease in children 0-14 age group in Jammu district as a function of child age.

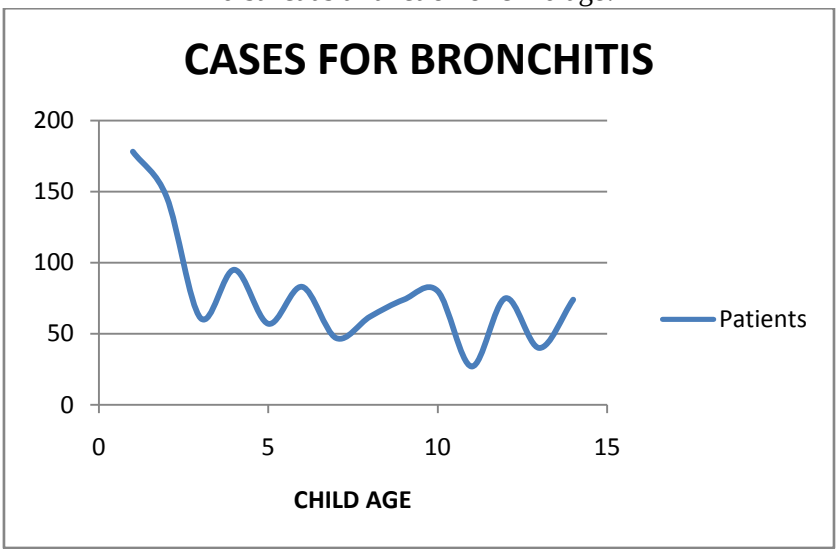


Figure 4: Generalized additive model for Thallesemia disease in children 0-14 age group in Jammu district as a function of child age.

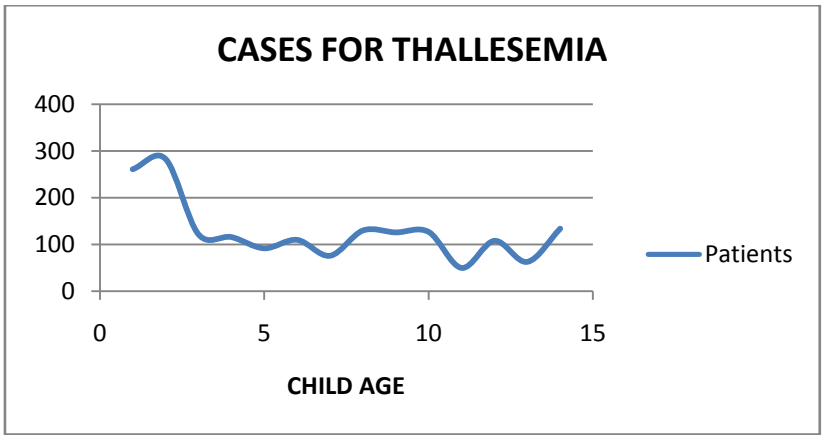
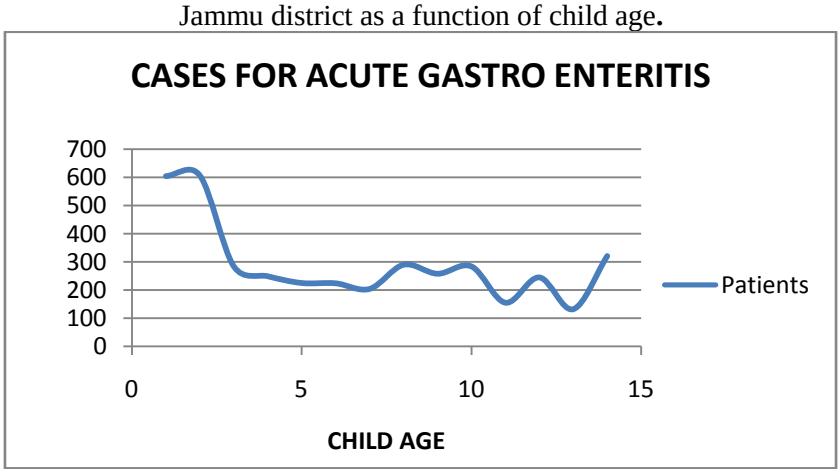


Figure 5: Generalized additive model for Acute Gastro Enteritis disease in children 0-14 age group in Jammu district as a function of child age**.**



Generalized additive models are very flexible, and can provide an excellent fit in the presence of non linear relationships and significant noise in the predictor variables. However, note that because of this flexibility, you must be extra cautious not to over-fit the data, i.e., apply an overly complex model(with many degrees of freedom) to data so as to produce a good fit that likely will not replicate in subsequent validation studies. In other words, evaluate whether the added complexity (generality) of generalized additive models (regression smoothers) is necessary in order to obtain a satisfactory fit to the data. Often, this is not the case, and given a comparable fit of the models, the simpler generalized linear model is preferable to the more complex generalized additive model.

IV. OBSERVATIONS AND CONCLUSION

Children affected by diseases like Acute Gastroenteritis(AGE), Bronchitis, Anemia, Seizure and Thallesemia remains a leading cause of childhood morbidity in developing countries like India. These diseases are major cause of illness in young Children and its prevalence is higher at low aged child particularly due to immature immune system, genetic reasons, neighborhood deprivation and exposure to environmental pollution. Children of rural areas are more susceptible to AGE, Bronchitis and Anemia diseases than Children of urban areas because of unhygienic living conditions, lack of good drinking water facilities, bad toilet facilities, nutritional deficiencies etc. The Generalised Linear Model (GLM) and Generalised Additive Model(GAM) ,particularly, by assuming ordinal logistic distribution(in case of local settings) are good diagnostic techniques for studying the status of Children's diseases in any area and helps in forming government policies for mitigating health problems of our society to create conducive atmosphere for further sustainable development.

REFERENCES

[1]. Cox, D. R. and Snell, E. J. (1968). A general definition of residuals (with discussion). J. Roy. Statist. Soc. B, 30, 248-275
[2]. Nelder, J.A. and Wedderburn, R. W. M. (1992).Generalised linear models. J.R. Statist. Soc. A 135: 370-84.
[3]. Hastie, T.J. and Tibshirani, R.J. (1990). Generalized Additive Models, New York: Chapman and Hall.
[4]. Austin, M.P.(1987). Models for the analysis of species response to environmental gradients. Vegetatio,69, 35-45.
[5]. Yee, Thomas W. and Mitchel, Neil D.(1991). Generalised additive models in plant ecology. Jour. of Vegetation Science 2: 587-602.

[6]. Austin, M.P. and Cunningham,R.B.(1981). Observational analysis of environmental gradients. Proc.Ecol.Soc.Aust.11,109-119.
[7]. Nicholis,A.O.(1989). How to make a biological survey go further with generalized linear models. Boil.Conserve.50,51-75.
[8]. Hastie, T.J. (1992). Generalized additive models In: Statistical models in S. Chambers, J.M. and T.J. Hastie (eds).
[9]. Wadsworth and Brooks, Pacific Grove.