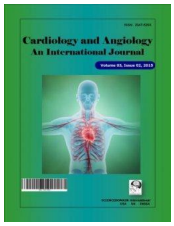

Cardiology and Angiology: An International Journal
4(2): 65-79, 2015, Article no.CA.2015.029
ISSN: 2347-520X

SCIENCEDOMAIN *international*
www.sciencedomain.org

Factors Influencing the Risk of Death among Patients with Heart Failure

Muzeyin Ahmed Berarti[1] and Ayele Taye Goshu[2*]

[1]Department of Statistics, Wolaita Sodo University, Ethiopia.
[2]Department of Statistics, School of Mathematical and Statistical Sciences, Hawassa University, Ethiopia.

Authors' contributions

This work was carried out in collaboration between both authors. Author MAB conceptualized, proposed and designed the research, design R-code and existing library routines and software packages in combination, analyzed the data and wrote the draft manuscript. Author ATG commented with dedication on the draft, participated in editing design of the study, performed the statistical analysis and approved the final manuscript. Both authors conventionally read and approved the final manuscript.

Article Information

DOI: 10.9734/CA/2015/17793
Editor(s):
(1) Francesco Pelliccia, Department of Heart and Great Vessels University La Sapienza, Rome, Italy.
Reviewers:
(1) Kuipo Yan, Department of Cardiology, Hospital of Henan College of TCM, China.
(2) Alexander Berezin, Internal medicine Department, Zaporozhye Medical University, Ukraine.
Complete Peer review History: http://sciencedomain.org/review-history/10075

Original Research Article

Received 26th March 2015
Accepted 20th April 2015
Published 7th July 2015

ABSTRACT

The purpose of this study was to investigate the impact of risk factors on the death of patients with heart failure in a cohort of patients hospitalized with heart failure disease. In this paper we used chi-square tests with the aim of studying the relationship of each factor with survival. Generalized Additive Models (GAM), particularly Generalized Additive Logistic Regression Model, was used to examine the impact of risk factors on the death of patients with heart failure out of 263 patients considered in the analysis, 18.6% patients died of heart failure. A death proportion for female was 19.6% and that of male patients was 17.5%. From the GAM analysis the predictors: age, anemia, Tuberculosis, HIV status, renal inefficiency, diabetes, hypertension and sinus were found to significantly affect the death status of a patient. Being older age, anemic, renal inefficient, TB positive, HIV positive, diabetic, hypertensive and sinus positive increase the risk of death of a heart failure patient.

Keywords: Heart failure; generalized additive model; smoothing method; splines.

**Corresponding author: E-mail: ayele_taye@yahoo.com;*

1. INTRODUCTION

Cardiovascular diseases are among the most frequent causes of death worldwide. Heart failure is an enormous medical and societal burden and a leading cause of hospitalization among cardiovascular diseases. It is a very common disease, with severe morbidity and mortality, and a frequent reason of hospitalization. Heart failure is the end stage of many cardiac and non cardiac pathological processes, from ischemic heart disease and the range of cardiomyopathies to respiratory disease and severe anemia (Ahern et al. [1]). As such, heart failure is not an underlying cause of death according to the WHO (World Health Organization) definition, but rather an intermediate cause of death with a diverse range of possible underlying causes of death. Anemia in heart failure is complex and multi factorial. Anemia resulting from a lack of sufficient Iron for synthesis of hemoglobin is by far the most frequent hematological disease of infancy and childhood (Lulu et al. [2]).

Using the historical definition by the World Health Organization, anemia is defined when Hemoglobin concentration is less than 13 g/dl for men or less than 12 g/dl for women (Oliva et al. [3]). Anemia is not a specific entity but an indication of an underlying pathologic process or disease (Lulu et al. [2]).

2. METHODS

2.1 Design of the Study

The study was a retrospective cohort study, which reviews the patient's card and information sheet. In this study secondary data was incorporated. The hospital's registry was used to retrieve data on Heart failure.

2.2 Variables in the Study

Outcome Variables: The response or outcome variable in this study is the binary response variable: Death status of patients during Hospital stay due to heart failure. This status of patient is coded as 1 if the patient died in hospital and 0 if the patient alive.

Independent Variables: The prognostic variables which were expected to be the risk factors of heart failure are categorical and continuous (see Table 1).

2.3 Statistical Methods

In this study, Chi-square analysis w as used to find out whether there is an association between each predictor variables and death status and the Generalized Additive Logistic Regression Model was used to assess the impact of various risk factors on the death in patients with heart failure[1].

In this section the flexible statistical methods (which are the extension of the traditional linear models) have been described which may be used to identify and characterize the effect of potential prognostic factors on an outcome variables. These methods are called Generalized Additive Models.

Here the logistic regression model which is among the most commonly used statistical methods in medical researches was used as specific illustration of Generalized Additive Model.

2.3.1 Generalized additive models (GAMs)

A generalized additive model is a generalized linear model with a linear predictor involving a sum of smooth functions of covariates (Hastie and Tibshirani, [4] and [5]). Generalized additive models (GAMs) follow from additive models, as generalized linear models (GLM) follow from linear models. The response may follow any exponential family distribution, or simply have a known mean variance relationship, permitting the use of a quasi-likelihood approach as described (Wood [6]). The model allow s for rather flexible specification of the dependence of the response on the covariates, but by specifying the model only in terms of smooth functions, rather than detailed parametric relationships.

To use GAMs in practice require s some extensions to GLM methods:

1. The smooth functions must be represented somehow.
2. The degree of smoothness of the functions must be made controllable, so that models with varying degrees of smoothness can be explored.
3. Some means for estimating the most appropriate degree of smoothness from data is required, if the models are to be useful for more than purely exploratory work.

[1] *R version 3.0.3 (2014-03-06) Statistical software have been used to analyze the data throughout the paper*

In general the model has the following structure

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + ... \quad (1)$$

Where, $\mu_i \equiv E(Y_i)$ and $Y_i \sim$ some exponential family distribution. Y_i is a response variable, $\mathbf{X}^*_i$ is a row of the model matrix for any strictly parametric model components, $\boldsymbol{\theta}$ is the corresponding parameter vector, and the f_j are smooth functions of the covariates, x_k.

Logistic model for binary data is one of the most widely used models in medical research. Here the dependent (outcome) variable Y_i is 0 or 1, with 1 indicating an event (like death or relapse of a disease) and 0 indicating no event. Our goal is modeling $p(y_i \mid x_{i1}, x_{i2}, ... x_{ip})$ the probability of an event given prognostic factors $x_{i1,} x_{i2}, ... x_{ip}$ .The linear logistic model assumes that the log-odds are linear:

$$\log\left(\frac{p(y_i \mid x_{i1}, x_{i2}, ... x_{ip})}{1 - p(y_i \mid x_{i1}, x_{i2}, ... x_{ip})}\right) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} \quad (2)$$

whereas the generalized additive logistic model (Hastie and Tibshirani, [7]) is:

$$\log\left(\frac{p(y_i \mid x_{i1}, x_{i2}, ... x_{ip})}{1 - p(y_i \mid x_{i1}, x_{i2}, ... x_{ip})}\right) = \beta_0 + f_1(x_{1i}) + f_2(x_{2i}) + ... + f_p(x_{pi}) \quad (3)$$

The functions $f_1, f_2, ..., f_p$ are unspecified (non-parametric) smoothing functions. The generalized additive model replaces $\sum \beta_j x_j$ with $\sum f_j(x_j)$ where f_j is an unspecified ('nonparametric') function. This function is estimated in a flexible manner using a scatter plot smoother. The estimated function $f(x_j)$ can reveal possible nonlinearities in the effect of the x_j .

2.3.2 Smoothing Methods

A spline curve is a piecewise polynomial curve, i.e., it joins two or more polynomial curves. The locations of the joins are known as "knots". In addition, there are boundary knots which could be located at or beyond the limits of the data. Smoothing splines arise as the solution to the following simple-regression problem: Find the function $\hat{f}(x)$ with two continuous derivatives that minimizes the penalized sum of squares (Wood, [8] and [6] , Fox & Weisberg, [9] and Hastie & Tibshirani, [7], Maindonald, [10]),

$$SS^*(h) = \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int_{x_{\min}}^{x_{\max}} [f''(x)]^2 dx \quad (4)$$

Table 1. Explanatory variables and their coding

Variables	Category and coding
Sex	Male= 1, Female= 0
Age at the start of treatment	Continues
Renal inefficiency	Yes= 1, No= 0
Place of residence	Rural= 0, Urban = 1
Pneumonia	Positive= 1, Negative= 0
Tuberculosis (TB)	Positive= 1, Negative= 0
Pulse rate at the start of treatment	Regular (60 − 80 bpm)=1, Irregular (≥ 80 bpm)=2
Diabetes mellitus	Positive= 1, Negative= 0
Human Immune Deficiency Virus (HIV)	Reactive= 1, Nonreactive= 0
Anemia	Anemic= 1, Non- anemic= 0
Hyper tension	Positive= 1, Negative= 0
Blood pressure at the star t of treatment	Nor mal (below1 20 /80 mm Hg)=1 High(between12 0/80 -1 79 /1 09 mm Hg)=2 Uncontrollable(above 1 80/1 10 mm Hg)=3
Sinus	Positive= 1, Negative= 0

Where, λ is a smoothing parameter, analogous to the neighborhood-width of the local polynomial estimator. Here y is a response or outcome variable, and x is a prognostic factor. The interest is to fit a smooth curve $f(x)$ that summarizes the dependence of y on x. If we were to find the curve that simply minimizes $\sum_{i=1}^{n}[y_i - f(x_i)]^2$, the result would be an interpolating curve that would not be smooth at all. In statistical work, y_i is usually measured with noise, and it is generally, more useful to smooth (x_i , y_i) data, rather than interpolating them. Notice that $\int_{x_{\min}}^{x_{\max}}[f''(x)]^2$ measures the "wiggliness" of the function f: linear f s have $\int_{x_{\min}}^{x_{\max}}[f''(x)]^2 = 0$, while non-linear f s produce values bigger than zero.

- The first term in Equation (4) is the residual sum of squares. The second term is a roughness penalty, which is large when the integrated second derivative of the regression function $f''(x)$ is large that is, when $f(x)$ is 'rough' (with rapidly changing slope). The endpoints of the integral enclose the data.
- At one extreme, when the smoothing constant is set to $\lambda = 0$ (and if all the x-values are distinct), $f(x)$ simply interpolates the data; this is similar to a local-regression estimate with $span = 1/n$. At the other extreme, if λ is very large, then $\hat{f}$ will be selected so that $\hat{f}''(x)$ is everywhere 0, which implies globally linear least-squares fit to the data (equivalent to local regression with infinite neighborhoods).

2.4 Model Selection

Choosing an appropriate model is the major issue in statistical investigations. Omitting relevant variables that are correlated with regressors causes least square s to be biased and inconsistent. Including irrelevant variables reduces the precision of least squares. So, from a purely technical point, it is important to estimate a mo de l that has all of the necessary relevant variables and no ne that are irrelevant. It is also important to u se a suitable functional form.

The mgcv-package of R Statistical software selects the degrees of freedom for each term automatically. However, it cannot automatically decide whether to drop a term all together or not. Hence the term must be removed by the investigator. The criteria for removal of a term are the following based on Wood [11]:

- If the effective degrees of freedom (edf) for the smooth term close to 1 and large p.value for parametric term.
- If the plotted confidence limit includes zero everywhere.
- If the Generalized Cross Validation (GCV) / Un-Biased Risk Estimator (UBRE) dropped when the term is dropped.

2.5 Goodness of Fit of the Model

The goodness of fit or calibration of a model mea sures how well the model describes the response variable.Assessing goodness of fit involves inves tigating how close values predicted by the model with that of observed values (Bewick and Jonathan, [12]).

After fitting the logistic regression model or once a model has been developed through the various steps in estimating the coefficients, there are several techniques involved in assessing the appropriateness, adequacy and usefulness of the model. The Pearson's Chi-square, the likelihood ratio tests (LRT), Hosmer and Lemeshow Test and the Wald tests are the most commonly used measures of goodness of fit for categorical data (Hosmer and Lemeshow, [13]). Besides these, different diagnostic plots can be used based on the model class.

The *gam.check* function of mgcv-package of R Statistical software returns four diagnostic plots for Generalized Additive Models:

1. A quintile-comparison plot of the residuals allows us to look for outliers and heavy tails.
2. Residuals versus linear predictors (simply observed y for continuous variables) helps detect non constant error variance.
3. Histogram of the residuals are good for detecting non normality
4. Response versus fitted value.

2.6 Ethical Considerations

Ethical clearance was obtained from the Hospital.

3. RESULTS AND DISCUSSION

The main objective of this study has been to assess the impact of risk factors in the death of patients with heart failure. The data of size 263 were obtained from record reviews of all inpatient heart failure patients admitted to Asella Referral Hospital from February, 2009 to March, 2012. The mean age of patients is 41.51 with standard deviation 19.784 ranging from 15 to 91.25% of patients were less than 24 years old the median age is 40 and 75% of the patients were aged below 58 years.

The output on Table 2 shows the proportions of death among patient of heart failure, frequency distribution, Chi-square, p-value and degrees of freedom with respect to each category of the categorical explanatory variables.

The results reveal that out of 263 patients considered in the analysis, 18.6% patients have died of heart failure while 81.4% were alive. A death proportion for female was 19.6% and that of male patients was 17.5%.

Hypertensive patients have higher Risk of death than any other groups. Anemia status was found significantly associated with death status of patients.

Moreover, the Table 2 shows that anemia, diabetes mellitus, HIV, hypertension, pneumonia, blood pressure, renal inefficiency, sinus and tuberculosis were found to have significant association with death status of heart failure patients. In contrast, no association was found between death status and the independent variables: sex, pulse rate and residence of patients.

Table 2. Test of association between death status and explanatory variables (Asella Referral Hospital, April 2012)

Variable	Category	Patient (N=49)				
		Alive N(%)	Dead N(%)	Total N(%)	Chi-square value (p-value)	Df
Sex	Male	99(82.5)	21 (17.5)	120 (45 .6)	0.1863(0.666)	1
	Female	115(80.4)	28 (19.6)	143 (54 .4)		
Residence	Urban	95(81.2)	22(18.8)	117(44.5)	0.0041(0.9488)	1
	Rural	119(81.5)	27(18.5)	146(55.5)		
Anemia	Anemic	45(65.2)	24(34.8)	69(26.2)	16.09(0.0000)	1
	Non-anemic	169(87.1)	25(12.9)	194(73.8)		
Diabetes	Positive	30(53.6)	26(46.4)	56(21.3)	36.27(0.000)	1
	Negative	184(88.9)	23(11.1)	207(78.7)		
HIV	Reactive	27(61.4)	17(38.6)	44(16.7)	13.9499(0.0001)	1
	Nonreactive	187(85.4)	32(14.6)	119(83.3)		
Hypertn.	Positive	41(51.9)	38(48.1)	79(30)	64.69(0.000)	1
	Negative	173(94.0)	11(6.0)	184(70)		
Pneumonia	Positive	83(73.5)	30(26.5)	113(43)	8.19(0.004)	1
	Negative	131(87.3)	19(12.7)	150(57)		
Pressure	Normal	117(90.0)	13(10.0)	130(49.4)	39.31(0.000)	1
	High	59(96.4)	6(3.6)	65(24.7)		
	Uncontrol	38(55.9)	30(44.1)	68(25.9)		
Pulse rate	Regular	111(83.5)	22(16.5)	133(50.6)	0.7751(0.3786)	1
	Irregular	103(79.2)	27(20.8)	130(49.4)		
Renal ineffi.	Yes	31(53.4)	27(46.6)	58(22.1)	38.263(0.000)	1
	No	183(89.3)	22(10.7)	205(77.9)		
TB	Positive	54(65.1)	29(34.9)	83(31.6)	21.28(0.000)	1
	Negative	160(88.9)	20(11.1)	180(68.4)		
Sinus	Positive	28(56.0)	22(44.0)	50(19)	26.21(0.000)	1
	Negative	186(87.3)	27(12.7)	213(81)		

3.1 Analysis of Generalized Additive Logistic Regression

Here Generalized Additive Logistic regression is illustrated. For each of the predictors, a smoother was fit by the *f* functions. The default spline used in the function *f* that does the smoothing is thin plate regression splines, which are slightly different from the B-splines, but are apparently preferred because they don't depend on the number of knots selected and also they generalize to smooth's of more than one variable at a time.

3.1.1 Full model

```
Data <- read.table("mom.dat", header =T)
library(mgcv)
model <-
 gam(death ~ s(age) + as.factor(pulserate)
+ as.factor(sex) + as.factor(anemia) +
as.factor(pneumonia) + as.factor(HIV) + as.
factor(TB) + as.factor(pressure) +
as.factor(renalineffeciency) + as.factor(diab
etes) + as.factor(residence) + as.factor(hyp
ertension) + as.factor(sinus), family = bino
mial(link = logit), data =Data)
summary(model)
```

Approximate significance of smooth terms:

```
edf    Ref. df   Chi.sq   p-value
s(age)  2.27   2.84   13.5    0.0033 **
Signif. codes:  0 '***'  0.001  '**' 0.01 '*'
0.05 '.' 0.1 ' ' 1
R-sq. (adj) = 0.705 Deviance explained =
69.7%
UBRE = -0.585 Scale est. = 1 n = 263
```

Graphical presentation of the data is carried out through R-code below.

```
>  library(lattice )
>  par(mfrow=c(4,4))
>  plot(model,all.terms=T,residual=T)
```

The model consists of smooth and parametric linear terms. As it can be seen from the output of the parametric model, in Table 3, anemia, Human Immune deficiency Virus (HIV), Tuberculosis (TB), renal inefficiency, Diabetes mellitus hypertension and sinus were significantly related to death of a patient. The smooth term age was significant as effective degrees o f freedom is much greater than 1 and p- value is very small.

The plot (Fig. 1) of the model shows, a semi-parametric model of death status o f Heart failure patients at follow- up Clinic of Asella Hospital, with factor for discrete variables and a smooth term for the dependence o n age. The first plot shows the smooth of age, with 95% confidence interval, while the other plots show the estimated effect, for e ach level of discrete variables.

The rug plots, along the bottom of the first plot, show the observed values of the covariate age, while the other plots show the levels (factor) of each explanatory variable. Number in y-axis caption is the effective degrees of freedom of the term being plotted for the continuous variables (age in the case of this research paper) and partial for discrete variables.

Table 3. The output of GAM when all variables are considered

	Estimate	Std. error	z value	Pr(>|z |)
(Intercept)	-8.2983	1.4530	-5.71	1.1e-08
as.factor(pulserate)2	0.0348	0.6701	0.05	0.95855
as.factor(sex)1	-0.5132	0.6796	-0.76	0.45017
as.factor(anemia)1	1.9846	0.7573	2.62	0.00878
as.factor(pneumonia)1	1.0641	0.6518	1.63	0.10255
as.factor(HIV)1	3.2934	0.9146	3.60	0.00032
as.factor(TB)1	1.7795	0.6695	2.66	0.00786
as.factor(pressure)2	-0.1340	0.8837	-0.15	0.87942
as.factor(pressure)3	-1.0965	1.1173	-0.98	0.32641
as.factor(renalineffeciency)1	2.4045	0.7279	3.30	0.00095
as.factor(diabetes)1	2.4856	0.7610	3.27	0.00109
as.factor(residence)1	0.6671	0.7177	0.93	0.35268
as.factor(hypertension)1	3.4242	1.0830	3.16	0.00157
as.factor(sinus)1	1.9232	0.7700	2.50	0.01250

The solid lines/curves represent the estimated effects, with 95% Bayesian confidence limits shown as dashed lines. If the confidence limits includes zero everywhere and the estimated straight line comfortably and fully laid confidence limit, at the point where the line passes through zero on the vertical axis, then the explanatory variable under consideration is unrelated with the response variable. The points shown on the first plot are Pearson partial residuals. For a well fitting model the partial residuals should be evenly scattered around the curve to which they relate. The plot shows anemia, Human Immune deficiency Virus (HIV), Tuberculosis (TB), renal inefficiency, diabetes mellitus, hypertension and sinus are relate d with death of patients with

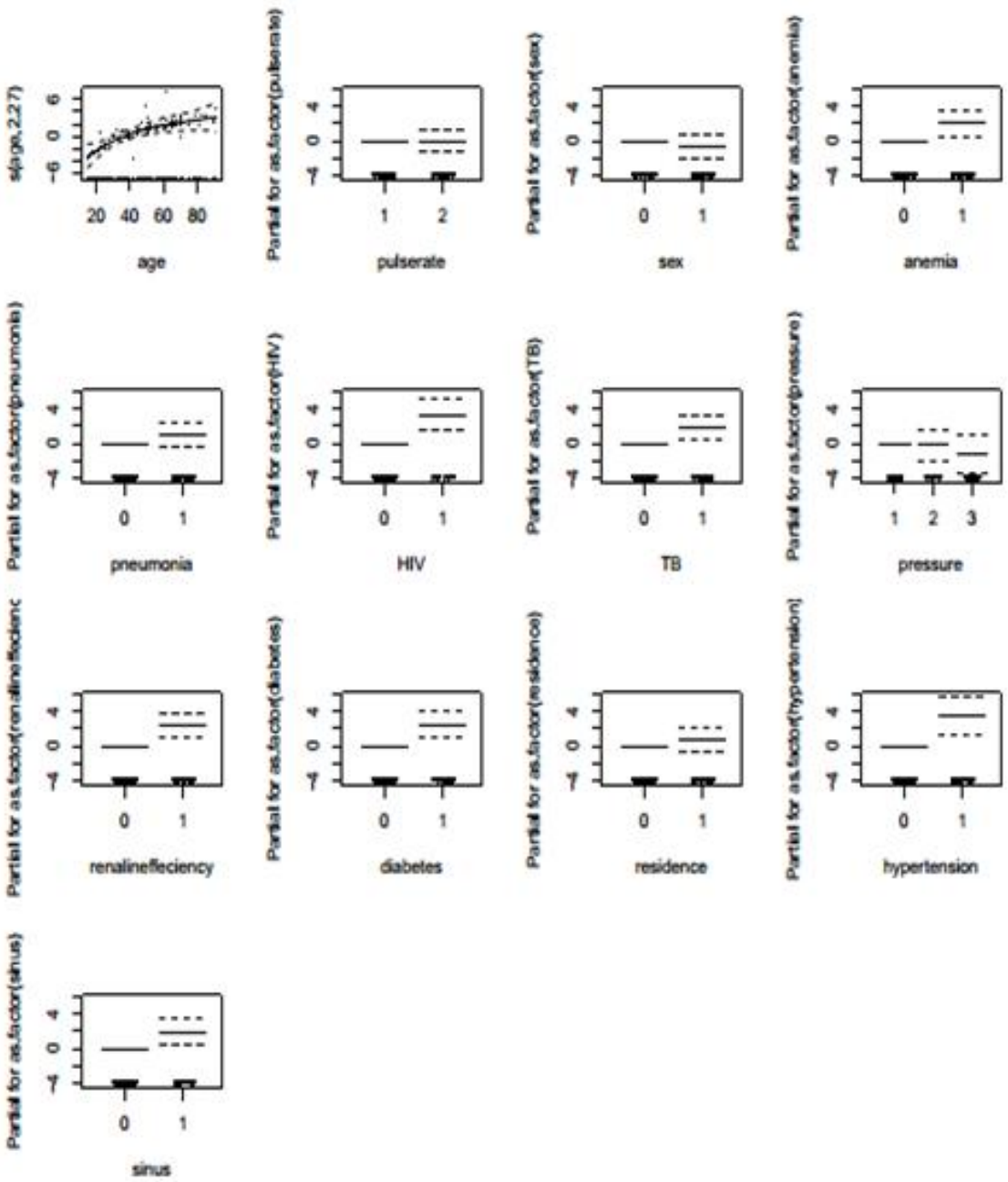

Fig. 1. Components of GAM model including all variables

Heart failure. That is, a patient with positive status of anemia, Human Immune deficiency Virus (HIV), Tuberculosis (TB), renal inefficiency, diabetes mellitus, hypertension and/or sinus was more likely had risk of death than that with respective opposite tests.

3.1.2 Variable selection

In statistical modelling, the choice of an optimal predictive model from a set of competing models is of extreme importance problem. There are a great deal of algorithms and procedures for searching the model space and selection criteria for choosing between competing models.

If all the three criteria stated under section 2.4 are satisfied, the term should be dropped (re moved). Hence, from Fig. 1, pulse rate, pressure, sex, residence and pneumonia are candidates to be dropped. Let us examine each of them dropping a term per step. It makes sense to start with the term for which the zero line is most comfortably lie within confidence band. Alternatively we can start dropping the term having largest p-value first. Accordingly pulse rate looks like the first candidate for removal.

```
Model <gam(death ~ s(age) + as.factor(sex) + as.factor(anemia) + as.factor(pneumonia) +as.factor(HIV) + as.factor(TB) + as.factor(pressure) + as.factor(renalineffeciency) +as.factor(diabetes) + as.factor(residence) + as.factor(hypertension) + as.factor(sinus), family = binomial(link =logit),data =Data)
model
```

```
Estimated degrees of freedom:
2.2744  total = 15.27437
UBRE score: -0.5924804
> pdf(file="gamplot07.pdf")
> par(mfrow=c(4,3))
>
plot(model,all.terms=T,residual=T)
> dev.off()
pdf

2
```

Un-Biased Risk Estimator (UBRE) score has decreased from -0.585 to -0.5925 supporting the decision to remove the pulse rate term. Still from Fig.2 pressure, sex, residence and pneumonia are candidates to be dropped and pressure seem the second variable to remove.

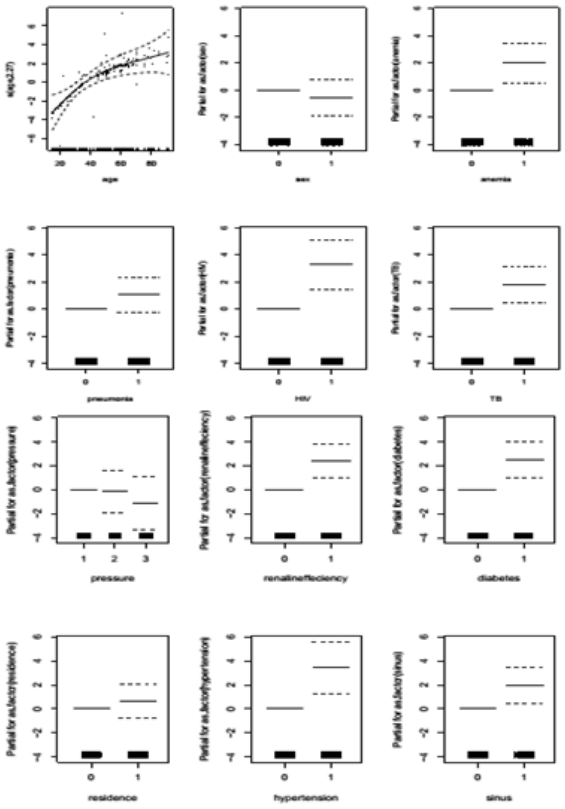

Fig. 2. Components of GAM model excluding pulse rate

```
model <- gam(death ~ s(age) + as.factor(sex) + as.factor(anemia) + as.factor(pneumonia) +
        as.factor(HIV) + as.factor(TB) + as.factor(renalineffeciency) + as.factor(diabetes) +
        as.factor(residence) + as.factor(hypertension) + as.factor(sinus), family = binomial(link =
        logit),  data = Data)
model
```

```
Estimated degrees of freedom:
2.1303  total = 13.13026
UBRE score: -0.6034817
> pdf(file="gamploto3.pdf")
> par(mfrow=c(4,4))
> plot(model,all.terms=T,residual=T)
> dev.off()
pdf
2
```

- Fig. 3 (plot for Components of GAM Model excluding pulse rate & pressure) shows sex, residence and pneumonia are candidates to be dropped and sex is the third covariate to be dropped and we were right in dropping pressure since UBRE has dropped.

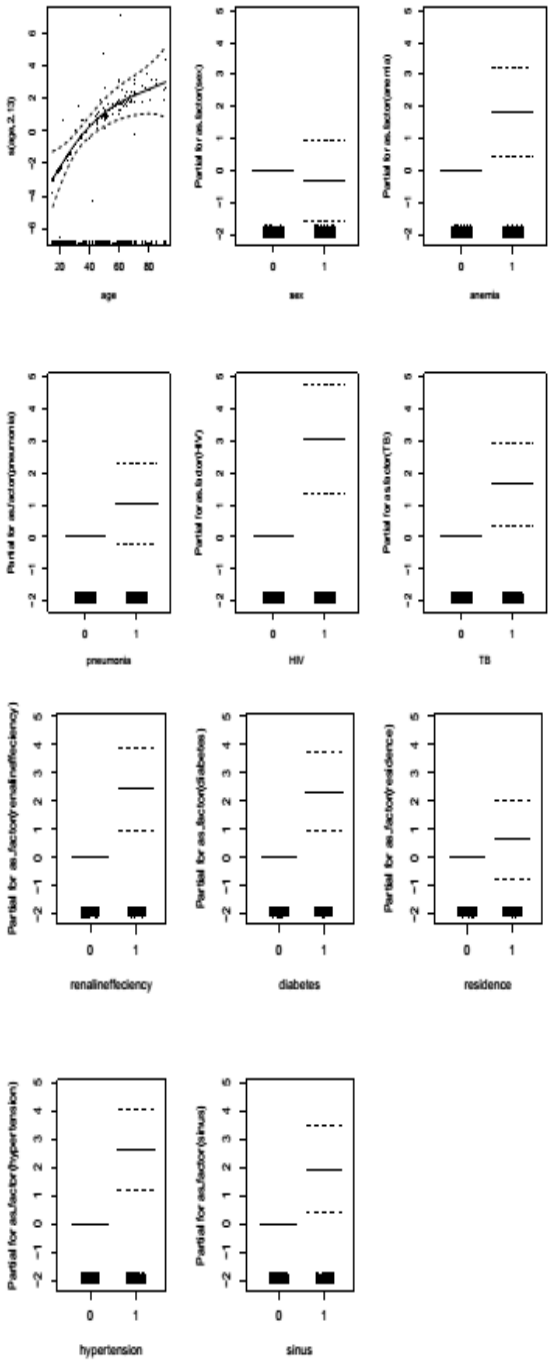

Fig. 3. Components of AM model excluding pulse rate & pressure.

```
model <- gam(death ~ s(age) + as.factor(anemia) + as.factor(pneumonia) + as.factor(HIV) +
        as.factor(TB) + as.factor(renalineffeciency) + as.factor(diabetes) + as.factor(residence) +
        as.factor(hypertension) + as.factor(sinus), family = binomial(link = logit), data = Data)
model
```

```
Estimated degrees of freedom:
2.0734  total = 12.0734
UBRE score: -0.6099116
> pdf(file="gamplot04.pdf")
> par(mfrow=c(4,3))
> plot(model,all.terms=T,residual=T)
> dev.off()
pdf
2
```

Here also UBRE is dropped and from Fig. 4 all the three conditions are satisfied for the removal of the variable sex from the model.

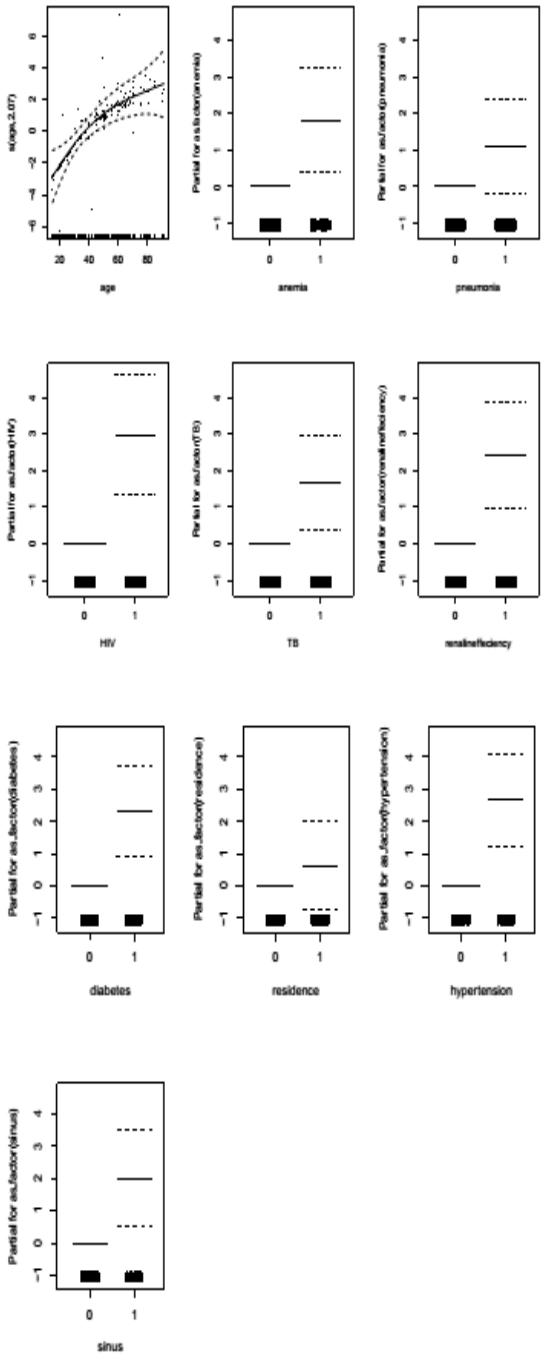

Fig. 4. Components of GAM model excluding pulse rate, pressure & sex

```
model <- gam(death ~ s(age) +
        as.factor(anemia) +
        as.factor(pneumonia) + as.factor(HIV)
        + as.factor(TB) +
        as.factor(renalineffeciency) +
        as.factor(diabetes) +
        as.factor(hypertension) +
        as.factor(sinus),  family = binomial(link
        = logit), data = Data)
model
```

UBRE is dropped and looking at Fig. 5 , all the three criteria are satisfied for removal of Residence.

```
library(mgcv)
        model1 <- gam(death ~ s(age) +
        as.factor(anemia) + as.factor(HIV) +
        as.factor(TB)                      +
        as.factor(renalineffeciency)       +
        as.factor(diabetes)                +
        as.factor(hypertension)            +
        as.factor(sinus), family = binomial(link
        = logit), data = Data)
model1
```

Considering Fig. 6, the first two criteria are satisfied for removal of pneumonia, but the third is not since UBRE increased. However, very small increases in UBRE should not prevent a term from being dropped. Thus pneumonia should be dropped. We can test the statistical significance of a term in the model by dropping it and noting the change in the deviance (Fox & Weisberg, [4]). To confirm the removal of pneumonia lets test it using analysis of deviance:

```
library(mgcv)
    model< gam(death ~ s(age) + as.factor
    (anemia) + as.factor(pneumonia) + as.facto
    r(HIV) +
     as.factor(TB) + as.factor(renalineffeciency)
     + as.factor(diabetes) + as.factor(hyperte
    nsion) +
    as.factor(sinus), family =binomial(link = logi
    t), data = Data)
    model2 <gam(death ~ s(age) + as.factor(an
    emia) + as.factor(HIV) + as.factor(TB) + as.
    factor (renalineffeciency) + as.factor(diabet
    es) + as.factor(hypertension) + as.factor(sin
    us),

    family = binomial(link = logit), data = Data)
    anova(model2, model, test = "Chisq")
```

```
Estimated degrees of freedom:
1.9467  total  = 10.94673
UBRE score: -0.61396
>  pdf(file="gamplot05.pdf")
>  par(mfrow=c(3,3))
>
plot(model,all.terms=T,residual=T)
>  dev.off()
  pdf
  2
```

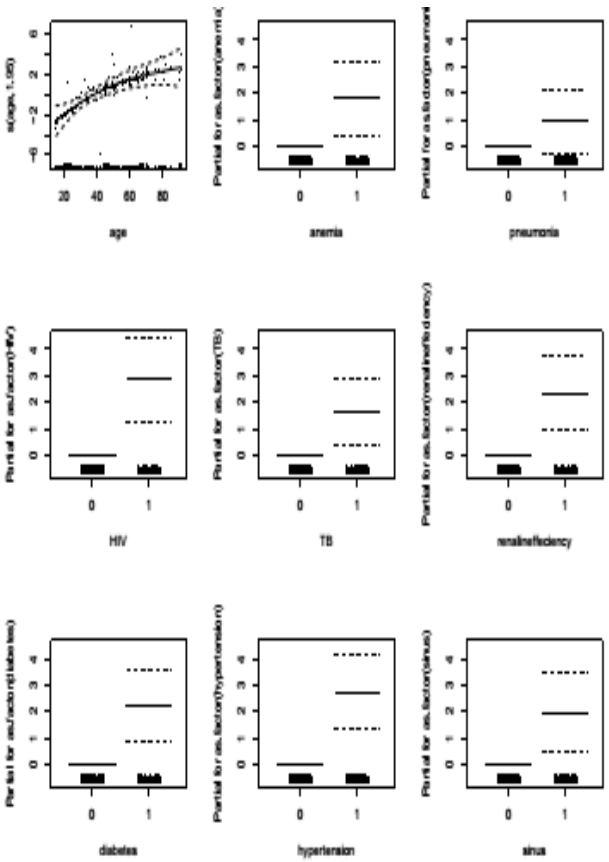

Fig. 5. Components of GAM model when pulse rate, pressure, sex & residence are excluded

```
Estimated degrees of freedom:
2  total = 1016
UBRE score: -0.6126
pdf(file = "gamplot08.pdf")
par(mfrow = c(3, 3))
plot(model1, all.terms =
T, residual = T)
dev.off()
pdf
```

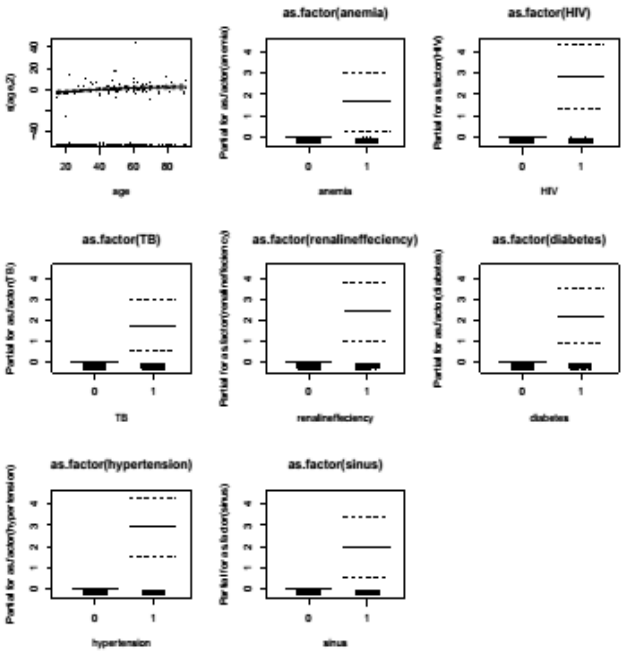

Fig. 6. Components of GAM model when pulse rate, pressure, sex, Residence & pneumonia are excluded

Analysis of Deviance Table

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1	253	81.9			
2	252	79.6	0.944	2.25	0.12

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Thus, the term pneumonia is statistically insignificant at 5% level of significance and should be dropped.

```
library(mgcv)
model <gam(death ~ s(age) + as.factor(anemia) + as.factor(HIV) + as.factor(TB) +
      as.factor(renalineffeciency) + as.factor(diabetes) + as.factor(hypertension) + as.factor(sinus),
      family = binomial(link = logit), data = Data)
summary(model)
```

	Estimate	Std. error	z value	Pr(>|z |)	
(Intercept)	-7.421	1.104	-6.72	1.8e-11	***
as.factor(anemia)1	1.664	0.691	2.41	0.01606	*
as.factor(HIV)1	2.837	0.779	3.64	0.00027	***
as.factor(TB)1	1.759	0.616	2.85	0.00431	**
as.factor(renalineffeciency)1	2.412	0.694	3.47	0.00051	***
as.factor(diabetes)1	2.201	0.664	3.32	0.00091	***
as.factor(hypertension)1	2.905	0.702	4.14	3.5e-05	***
as.factor(sinus)1	1.956	0.728	2.69	0.00723	**

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value	
s(age)	2	2.51	16.5	0.00062	***

```
Signif. codes: 0 ' *** ' 0.001 ' ** ' 0.01 ' * ' 0.05 ' . ' 0.1 ' ' 1
R- sq. (adj) = 0.694    Deviance explained = 67.6%
UBRE = -0.61257 Scale est. = 1     n = 263
```

The graph is produced as stated below.

```
> pdf(file="gamplot06.pdf")
> par(mfrow=c(3,3))
> plot(model,all.terms=T,residual=T)
> dev.off()
```

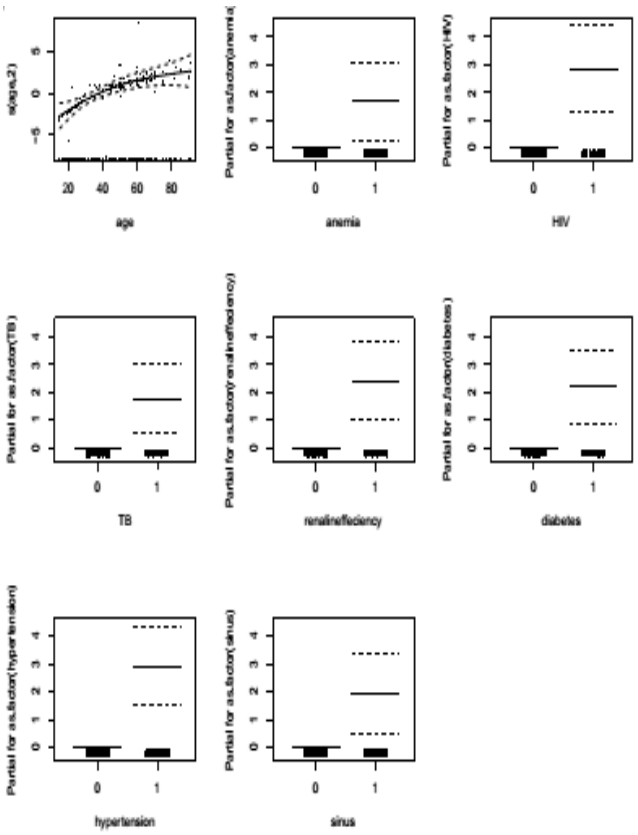

Fig. 7. Components of GAM Plot for the best fitted model: Partial-regression functions for the additive regression of overall significant variables

Both the last output and the plot (above Fig. 7) imply no further terms to delete; the model can be used for prediction. Hence, this is the best model. Therefore, the effects of the predictors age, anemia, Human Immune deficiency Virus (HIV), TB, renal inefficiency, diabetes, hypertension and sinus on the death status of the patients are found to be significant. A patient with positive status in anemia, Human Immune deficiency Virus (HIV), TB, renal inefficiency, diabetes, hypertension and/or sinus was more likely had risk of death than that with respective negative tests.

For continuous variable (age) case: The risk of death of a patient increases with increasing his/her age, after controlling all other covariates.

3.2 Goodness of Fit of the Model

After finding results, the overall adequacy of the model should be checked. There are several alternative methods to check the adequacy of the fitted model. Among many alternative methods, we used the following Diagnostic Plots to assess the model.

The normal Q-Q plot on the figure below resembles a (nearly) straight line and shows that there is no powerful outlier and influential value in the data respectively.

Residual s versus linear predictors above indicate, the standardized residuals are uncorrelated With the linear predictors; as this plot is a random scatter of points.

```
pdf(file = "modelcheck.pdf")
par(mfrow = c(2, 2))
gam.check(model)
dev.off()
```

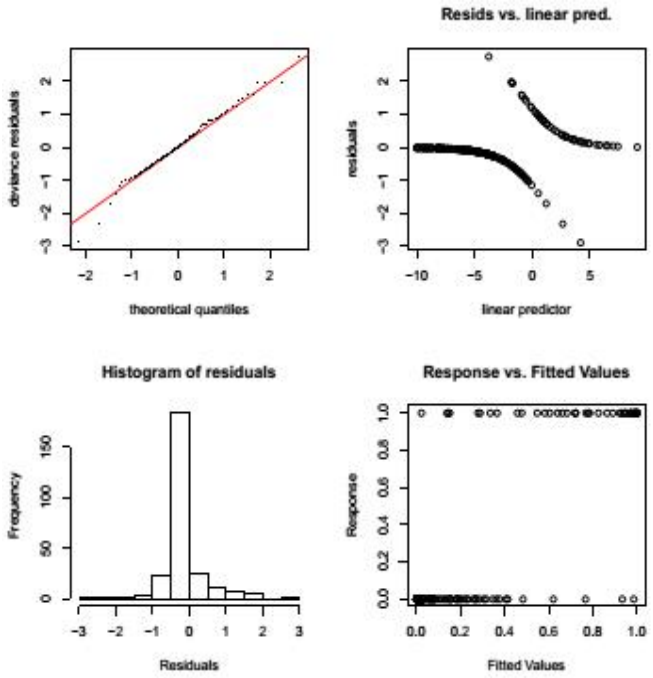

Fig. 8. Goodness of GAM Model

Histogram of the residuals approximately resemble s standard normal curve implying normality assumption is satisfied. Therefore, from the plot above, we can generalize that the model fits the data well.

3.3 Discussion

This study investigated the effect of predictor of mortality in group of hospitalized patients with heart failure. From the results, it was found that the survival of a patient is significantly related with age, anemia, TB, HIV status, hypertension, positive history of diabetics, renal inefficiency and sinus.

The findings obtained from this study were found to be comparable with similar studies in different countries. In this study renal inefficiency was related to both prevalence of anemia and death of heart failure patients. This result is highly comparable with the result obtained (Villacorta et al. [14] which found that renal dysfunction was associated with prevalence of anemia and risk of death of patients from Heart failure. According to the result of this study, hypertensive patients had higher Risk of death (48.1%) than any other groups followed by renal inefficiency (46.6%) and diabetes mellitus (46.4%), respectively. This result can best compared with the result obtained (Kosiborod et al. [15]) which found that the majority of patients had a history of hypertension (60%) and a substantial minority had diabetes (37%), and a history of renal insufficiency (20%).

Several previous studies have identified other adverse prognostic factors among patients with heart failure, including age, anemia, hypertension, positive history of diabetics, renal dysfunction and sinus. Our study suggests that Tuberculosis (TB), Human Immune deficiency Virus (HIV) status, should be added to this group of clinical variables.

4. CONCLUSIONS

The main objective of this study was to investigate effect of risk factors of death in patients with Heart failure at Asella Referral Teaching Hospital. The results revealed that out of 263 patients considered in the analysis, 18.6% patients had died of heart failure while 81.4% were alive. Hypertensive patients had higher risk of death than any other groups followed by renal inefficiency and diabetes respectively. The GAM analysis showed that the predictors: age, anemia, HIV status, renal inefficiency, diabetes, hypertension and sinus significantly affect the death status of a patient.

CONSENT

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Ahern R, Lozano R, Naghavi M, Foreman K, Gakidou E, Murray C. Improving the public health utility of global cardiovascular mortality data: The rise of ischemic heart disease. Population Health Metrics. 2011;9:8.
2. Lulu M, Birhane O, Hirut D, Shabbar J, Weber MW. Evaluation of clinical pallor in the identification and treatment of children with moderate and severe anaemia. Tropical Medicine and International Health. Addis Ababa University. 2000;11(5):805-810.
3. Oliva EN, Schey C, Hutchings AS. A review of anemia as a cardiovascular risk factor in patients with myelodysplastic syndromes. 2011;1/2:160-166.
4. Hastie T, Tibshirani R. Generalized additive models. Statistical Science. 1986;1(3):187-196.
5. Hastie T, Tibshirani R. Generalized additive models. London: Chapman and Hall; 1990.
6. Wood SN. Generalized additive models, An introduction with R. London: Chapman & Hall; 2006.
7. Hastie T, Tibshirani R. Generalized additive models for medical research. Statistical Methods in Medical Research. 1995;4:187-196.
8. Wood SN. Thin plate regression splines. Journal of Royal Statistical Society, Ser. B. 2003;65:95-114.
9. Fox J, Weisberg S. Nonparametric regression in R: An appendix to an R companion to applied regression, 2nd ed.; 2010.
10. Maindonald J. Smoothing terms in GAM models, centre for mathematics and its applications. Australian National University; 2010.
11. Wood SN. GAMs and generalized ridge regression for R. Martin Schlather University of Bayreuth, Germany. 2001;1/2.
12. Bewick L, Jonathan B. Statistics review 14: Logistic Regression; 2005.
13. Hosmer D, S. Lemeshow. Applied logistic regression. John Wiley and Sons, Inc., Newyork; 1989.
14. Villacorta H, Tello SB, Barcellos Dos Santos E, Steffen R, Wiefels C, Sales LA, Soares P, Mesquita TE. Renal dysfunction and anemia in heart failure. Arq Brascardiol. 2010;94(3):357-363.
15. Kosiborod M, Smith GL, Radford MJ, Foody JM, Krumholz HM. The prognostic importance of anemia in patients with heart failure. The American Journal of Medicine. 2003;114.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/10075