AJHS
Asian Journal of Home Science
Volume **13** | Issue 1 | June, 2018 | 68-74

DOI: 10.15740/HAS/AJHS/13.1/68-74
e ISSN-0976-8351 ■ Visit us: *www.researchjournal.co.in*

Research Paper

An application of generalized linear model in survival analysis

■ Yasmeena Ismail, S.A Mir, M.A Bhat **and** Nageena Nazir

Received: 17.12.2017; **Revised:** 19.03.2018; **Accepted:** 05.04.2018

■ **ABSTRACT :** Diabetes is a chronic, often debilitating and sometimes fatal disease, in which the body either cannot produce insulin or cannot properly use the insulin it produces. Type 1 diabetes occurs when the immune system mistakenly attacks and kills the beta cells of the pancreas. Type 2 diabetes occurs when the body can't properly use the insulin that is released (called insulin insensitivity) or does not make enough insulin. Diabetic nephropathy, also known as Kimmelstiel Wilson syndrome or nodular diabetic glomerulosclerosis or intercapillary glomerulonephritis, is a clinical syndrome characterized by albuminuria (>300 mg/day or >200 mcg/min), permanent and irreversible decrease in glomerular filtration rate (GFR), the rate of rise in serum creatinine (SrCr). According to the WHO, 31.7 million people were affected by diabetes mellitus (DM) in India in the year 2000. This figure is estimated to rise to 79.4 million by 2030, the largest number in any nation in the world. In this paper, survival analysis will be done of the type 2 diabetic nephropathy patients through generalized linear model. Most appropriate distribution for duration of diabetes is selected through Bayesian information criterion value. Then two generalized linear models are fitted by taking the duration of diabetes as response variable and the predictors as SrCr, number of successes; GFR, number of successes, respectively. These covariates are linked with the response variable using different link functions. At the last, survival function under different links will be compared.

See end of the paper for authors' affiliations

Nageena Nazir
S.K. University of Agricultural Sciences and Technology of Kashmir, Shalimar, Srinagar (J&K) India
Email : nazir.nageena@gmail.com

■ **KEY WORDS:** Generalized linear model, Link function, Bayesian information criterion, Survival function, Diabetic nephropathy, GFR

■ **HOW TO CITE THIS PAPER :** Ismail, Yasmeena, Mir, S.A., Bhat, M.A and Nazir, Nageena (2018). An application of generalized linear model in survival analysis. *Asian J. Home Sci.,* **13** (1) : 68-74, **DOI: 10.15740/HAS/AJHS/13.1/68-74.** Copyright@ 2018: Hind Agri-Horticultural Society.

Linear regression attempts to model the relationship between two variables, where one is the dependent or response variable and other is the independent or predictor variable. Generalized linear models are an extension of classical linear models introduced by Nelder and Weddeburn 1972 (McCullagh and Nelder, 1989). They showed that regression and analysis of variance methods could be applied to any response variable whose distribution belongs to the exponential family (Stroup and Kachman, 1994). In a generalized linear model three elements are involved. We have already looked at two of them, the probability distribution, the linear structure and the third is the link function. Generalized linear models include as special cases, linear regression and analysis of variance models, logit and probit models for quantal responses, log linear models and multinomial response models for counts and some commonly used models for survival data. GLM have application in disciplines as widely varied as agriculture, demography, ecology, economics, education, engineering, environmental studies and pollution, geography, geology, history, medicine, political science, psychology and sociology.

Survival analysis is the name for a collection of statistical techniques used to describe and quantify time to event data. In survival analysis we use the term 'failure' to define the occurrence of the event of interest. The term 'survival time' specifies the length of time taken for failure to occur. Failure time data or survival data are frequently encountered in biomedical studies, engineering, and reliability research. In medical studies, clinical endpoints for assessment of efficacy and safety of a promising therapy usually include occurrence of some predefined events such as deaths, the onset of a specific disease, the response to a new chemotherapy in treatment of some advanced cancer, the eradication of an infection caused by a certain micro-organism, or serious adverse events. The statistical analysis of survival data has been well developed in the literature. The estimation of the survival distribution can be done by the Kaplan-Meier product limit estimator, which can also be viewed as a kind of nonparametric maximum likelihood estimator. Several survival distributions are proposed and an appropriate distribution is identified by the various information criterions like AIC, BIC and AICC. AIC stands for Akaike's Information Criterion (Akaike, 1973). AIC is aimed at finding the best approximating model to the unknown true data generating process. It could be argued that a good model selection criterion should work even if the user tries a "bad" (e.g., over parameterized) model: if the model is bad, the criterion should be able to detect this. In this regard, AIC fails. In order to remove this deficiency, (Hurvich and Tsai, 1989) introduced a corrected version, AIC_C which refers to Finite Sample Corrected AIC. BIC stands for Bayesian information criterion unlike Akaike Information Criteria, BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models (Schwarz, 1978 and Kass and Raftery, 1995). Models that minimize the Bayesian Information Criteria are selected. From a Bayesian perspective, BIC is designed to find the most probable model given the data.

Diabetes is a chronic, often debilitating and sometimes fatal disease, in which the body either cannot produce insulin or cannot properly use the insulin it produces. Type 1 diabetes occurs when the immune system mistakenly attacks and kills the beta cells of the pancreas. Type 2 diabetes mellitus is a lifelong (chronic) disease in which the body becomes resistant to the normal effects of insulin and/or gradually loses the capacity to produce enough insulin in the pancreas. Onset is usually after 40 years of age but can occur at any age. Diabetic nephropathy, also known as Kimmelstiel Wilson syndrome or nodular diabetic glomerulosclerosis or intercapillary glomerulonephritis, is a clinical syndrome characterized by albuminuria (>300 mg/day or >200 mcg/min), permanent and irreversible decrease in glomerular filtration rate (GFR), the rate of rise in serum creatinine (SrCr). Throughout the world the number of the people developing type 2-DM is increased dramatically. According to the WHO, 31.7 million people were affected by diabetes mellitus (DM) in India in the year 2000. This figure is estimated to rise to 79.4 million by 2030, the largest number in any nation in the world.

Hakulinen and Tenkanen estimated the relative survival rates of lung cancer patients by assuming a Binomial distribution and applying generalized linear model approach with log-log link (Hakulinen and Tenkanen, 1987). Karem applied general and generalized linear models for determining which combination of effects allows for the optimal prediction of survival for lung cancer patients. They showed that a full effects generalized linear model outperforms the general linear model (Karem, 2006). Yuan, Hong and Shyr also studied the survival patterns of lung cancer patients by applying Cox proportional hazard models (Yuan *et al.*, 2007). Akram, Ullah and Taj investigated the survival pattern of cancer patients using the non-parametric and parametric modeling strategies. They applied Kaplan-Meier method and Weibull model based pn Anderson-Darling test to the real life time data of cancer patients (Akram *et al.*, 2007). Gurprit Grover, A Sabharwal and J Mittal estimated the survival functions of type 2 diabetic patients with renal complication. They also compared the estimated survival functions under the log and reciprocal links with Kaplan Meier (KM) estimates graphically (Grover *et al.*, 2013).

In this paper, data of type 2 diabetic patients was collected from SKIMS, Srinagar J&K (data base of Dr. Lal path's lab). The dataset consists of 53 Diabetic Nephropathy patients. Aim of this study is to obtain the survival function with the help of the data on type 2 diabetic patient. we first fit four different distributions separately on two models. And then choose the models with minimum AIC, BIC and AIC$_C$. Gamma distribution comes out to be the best distribution for the first model and Inverse Gaussian for the second model based on the values of AIC, BIC, AIC$_C$. The two Generalized linear regression analysis are performed by considering log duration as response variable, SrCr and number of success as independent variables for the first model and considering duration as response variable, GFR and number of success as independent variable for second model. These responses are linked with the independent variables by two link functions. And based on the estimates of both the models we will find the survival function by Kaplan Meier approach. This work is an extension of Grover *et al.* (2013) paper where they estimated the survival function based on the first model of this paper we have added up the second model which is based on GFR which takes into account the ages and gender of the patients under consideration.

■ RESEARCH METHODS

Generalized linear model is defined in terms of a set of independent random variables $Y_1,\ldots,Y_N$ each with a distribution from the exponential family . The Poisson, Normal, Binomial, Gamma, Inverse Gaussian distributions are some of the members of this family. The distribution of each Y_i has the canonical form and depends on a single parameter θ_i , thus

$$\mathbf{f(y_i;\theta_i) = \exp\left[y_i b_i(\theta_i) + c_i(\theta_i) + d_i(y_i)\right]} \quad (1)$$

The parameter θ_i are typically not of direct interest. Suppose that E $(Y_i) = \mu_i$ where μ_i is some function of θ_i. For a generalized linear model there is a transformation of μ_i such that

$$\mathbf{g(\mu_i) = x_i^T\beta} \quad (2)$$

In this equation *g* is a monotone, differentiable function called the link function. The most common link function are identity, log, reciprocal, power, cumulative logit. Log and identity link functions are used with all the distributions. Then an appropriate distribution is selected from the following four members of the exponential family of distributions Gamma, Inverse Gaussian, Normal and Multinomial. The important properties of these distributions are the gamma distribution is the most popular model for analyzing the skewed data. It is suitable for modeling data with different types of hazard rate function: increasing, decreasing, in the form of bathtub and unimodal. This characteristic is useful for estimating individual hazard rate functions and both relative hazards and relative times (Cox and Mann, 2008). Cox *et al.* (2007) presented a parametric survival analysis and taxonomy of the gamma hazard rate function. The hazard rate function of the Inverse Gaussian distribution has ∩-shape like log-normal, generalized Weibul and Log-logistic distributions, *i.e.* the hazard rate of Inverse Gaussian distribution is unimodal which increases from 0 to its maximum value and then decreases asymptotically to a constant. This is the reason Inverse Gaussian distribution is used often in reliability and survival analysis. The hazard function of lognormal could be increasing and then decreasing with time *i.e.*,non monotonic (Cox *et al.*, 2007). The probability distribution function likelihood function and the survival function of above distributions are:

Gamma distribution :

$$\mathbf{f(t\,|\,\lambda,\gamma) = \frac{\lambda^{\gamma}}{\Gamma\gamma} t^{\gamma-1} e^{-\lambda t}; \lambda > 0, \gamma > 0 \;\&\; t > 0} \quad (3)$$

$$\mathbf{S(t) = [1- I\ (\lambda t,\ \gamma)]} \quad (4)$$

where I $(\lambda t, \gamma)$ is the incomplete Gamma function defined as,

$$\mathbf{I}(\lambda t, \gamma) = \mathbf{\frac{1}{\Gamma\gamma} \int_0^{\lambda t} u^{\gamma-1} e^{-u} du} \quad (5)$$

$$\mathbf{f(t\,|\,\lambda,\gamma) = \frac{\lambda^{\gamma}}{\Gamma\gamma} t^{\gamma-1} e^{-\lambda t}; \lambda > 0, \gamma > 0 \;\&\; t > 0} \quad (6)$$

Inverse Gaussian Distribution :

$$\mathbf{f(t\,|\,\mu,\lambda) = \sqrt{\frac{\lambda}{2\lambda t^3}} \exp\left[\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right]; t > 0, \mu > 0 \;\&\; \lambda > 0} \quad (7)$$

$$\mathbf{S(t) = 1 - \sqrt{\frac{\lambda}{2\pi}} \int_0^t \left(\frac{1}{x}\right)^{3/2} \exp\left[\frac{-\lambda X}{2\mu^2}\left(1-\frac{\mu}{X}\right)^2\right] dx} \quad (8)$$

$$\mathbf{L = \left(\frac{\lambda}{2\pi}\right)^{r/2} \frac{1}{\prod_{i=1}^{n}\left(\sqrt{t_i^3}\right)^{\delta_i}} \exp\left[\frac{-\lambda \sum_{i=1}^{n} \delta_i (t_i - \mu)^2}{2\mu^2 t_i}\right] \prod_{i=1}^{n} [S(T_i)]^{(1=\delta_i)}} \quad (9)$$

Normal Distribution :

$$f(t\mid\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left[\frac{t-\mu}{\sigma}\right]^2\right) \qquad (10)$$

$$L=-n\log\sigma-\frac{n}{2}\log2\pi-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(t_i-\mu)^2 \qquad (11)$$

$$S(t)=1-\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t}e^{-(t-\mu)^2/(2\sigma^2)}dt=1-\frac{1}{2}\left[1+\operatorname{erf}\left(\frac{t-\mu}{\sigma\sqrt{2}}\right)\right]$$

(12)

Multinomial Distribution :

$$f(t_1,t_2,\dots,t_m\mid p_1,p_2,\dots,p_m)=\frac{n!}{\prod t_i!}\prod p_i^{t_i}=\left({}^{n}c_{t_1,t_2,\dots,t_m}\right)p_1^{t_1}p_2^{t_2}\dots p_m^{t_m}$$

$$L=\log n!-\sum_{i=1}^{m}\log t_i!+\sum_{i=1}^{m}t_i\log p_i+\lambda\left(1-\sum_{i}^{m}p_i\right)$$

$$P(A_1\dots A_m\mid\sum_{1}^{m}T_m=N)=\frac{P(A_1\dots A_m)}{P(\sum_{1}^{m}T_m=N)}P(\sum_{1}^{m}T_i=N\mid A_1\dots A_m)$$

Once all the above distribution are fitted to both the models an appropriate model is selected by comparing the values of AIC, BIC, $AIC_{C.}$ The idea of AIC is to select the model that minimizes the negative likelihood penalized by the number of parameters as specified in

AIC = –2 log p (L) + 2p

where, -2log p (L) is the deviance term and L refers to the likelihood under the fitted model and p is the number of parameters in the model. As already mentioned that if the model is bad, the particular criterion should be able to detect this, thus AIC fails a corrected version of AIC is introduced *i.e.*, AIC_C which refers to Finite Sample Corrected AIC. AIC_C, defined by

$$AIC_C(p,q)=-2\log[\text{likelihood}(p,q)]+2(p+q+1)\ \frac{n}{n-p-q-2} \qquad (13)$$

In AIC_C, we take penalty term for AIC, which we can consider to be 2(p+q+1) and multiply it by the correction factor $\frac{n}{n-p-q-r}$, And the third criterion is BIC, defined as:

$$BIC = -2\log p(L) + p\log(n) \qquad (14)$$

BIC differs from AIC only in the second term which now depends on sample size n. The model with the lowest AIC, BIC, AIC_C values is preferred.

Application :

Study of 53 diabetic nephropathy patients was carried out from the nephrology department of SKIMS, J&K (pathological tests were done in Dr Lal's Path Lab). The data regarding the age at which the diabetes was diagnosed, gender, protein albumin, 24 hour urine collection, SrCr values, Fasting Blood Glucose (FBG), Diastolic Blood Pressure(DBP) and Systolic Blood Pressure(SBP). GFR was also calculated by CKD-EPI creatinine equation (2009) expressed

$$eGFR = 141 \text{ x } \min(S_{Cr}\ k, 1)^a \text{ x } \max(S_{Cr}/k, 1)^{-1.209} \text{ x } 0.993^{Age} \text{ x } 1.018\ [\text{iffemale}] \text{ x } 1.159\ [\text{ifblack}] \qquad (15)$$

GFR is an important marker for the development of diabetic nephropathy. On the basis of all this information the above two discussed models are fitted and their AIC, BIC, AIC_C values are calculated. Table 1 gives the AIC, BIC, AIC_C values of the distributions fitted for the Generalized Linear Model

$$\text{Log}(Dur)_i = \beta_0 + \beta_1(SrCr)_i + \beta_2(\text{no. of success})_i \qquad (16)$$

where,

Dur = Duration of diabetes

SrCr = Serum creatinine

No. of success = The number of times SrCr exceeds its normal range (1.4mg/ml)

■ RESEARCH FINDINGS AND DISCUSSION

It is clear from the Table 1 that gamma distribution has the minimum value for AIC, BIC, AIC_C thus gamma distribution is the most appropriate distribution for the duration of diabetes of model (16).

Now, gamma generalized linear model with log link is used to estimate the duration of diabetes based on serum creatinine and number of success the results are shown in Table 2.

It is clear from the table that for estimating duration

Table 1 : AIC, BIC, AIC_C values of different distributions

Distribution	Link	AIC	BIC	AIC_C
Gamma	Log	-173.745	-165.789	-172.929
Inverse Gaussian	Log	-172.812	-164.856	-171.996
Normal	Identity	105.391	111.358	105.871
Multinomial	Cumulative logit	26.001	51.858	35.101

of diabetes serum creatinine and number of success are significant as the p-values are less the .0100. The fitted gamma generalized linear model is

$\mathbf{Log(Dur)_i = 0.132\ -0.058(SrCr)_i+0.002(no.\ of\ succ)_i}$ (17)

Now, Table 3 gives the AIC, BIC AIC_C values of the distributions fitted for the generalized linear model

$\mathbf{Log(Dur)_i = \beta_0 + \beta_1(GFR)_i + \beta_2(no.\ of\ success)_i}$ (18)

where,

Dur= Duration of diabetes

GFR= Glomelural Filtration Rate

No. of success= Number of times SrCr exceeds the normal range

In Table 3 we found that the values of AIC, BIC, AIC_C are minimum for Inverse Gaussian distribution. And Inverse Gaussian comes out to be the most appropriate distribution for the duration of diabetes of model (17). Inverse Gaussian generalized linear model with log link is used for estimating the duration of diabetes based on GFR and no. of success. The results are shown in Table 4.

Thus the fitted Inverse Gaussian generalized linear model with log link is

$\mathbf{Log(Dur)_i = -.266 + .010(GFR)_i + .002(no.\ of\ succ)_i}$ (19)

The estimates of shape and scale parameters of

Table 2 : Gamma Generalized Linear Model with log link for estimating the duration of diabetes based on serum creatinine and number of success

Variable	Parameter	Standard	95% CI		Wald Chi-square	p-value
	estimate	error	Upper	Lower		
Intercept	0.132	0.0379	0.058	0.206	12.163	0.00
SrCr	-0.058	0.0049	-0.068	-0.048	138.071	0.00
Succ	0.002	0.0005	0.001	0.003	14.725	0.00
Dispersion	0.002	0.0004	0.001	0.003		
AIC	-173.745					
BIC	-165.789					
AIC_C	-172.929					
Link =Log						
Log $(dur)_i$ =0.132 -0.058$(SrCr)_i$ + 0.002$(Succ)_i$						

*degree of freedom for each intercept is 1

Table 3 : AIC, BIC, AIC_C values for different distributions

Distribution	Link function	AIC	BIC	AIC_C
Gamma	Log	-190.843	-182.887	-190.026
Inverse Gaussian	Log	-193.339	-185.383	-192.523
Normal	Identity	-184.512	-176.556	-183.696
Multinomial	Cumulative logit	127.748	153.604	136.848

Table 4 : Inverse Gaussian generalized linear model estimating the duration of diabetes based on GFR and number of success

Variable	Parameter	Standard	95% CI		Wald	p-value
	estimate	Error	Upper	Lower	Chi-square	
Intercept	-0.266	0.0153	-0.296	-0.236	303.937	0.00
GFR	0.010	0.0007	0.009	0.011	209.553	0.00
Succ	0.002	0.0004	0.001	0.003	17.996	0.00
Dispersion	0.001	0.0003	0.001	0.002		
AIC	-193.339					
BIC	-185.383					
AIC_C	-192.523					
Link = Log						
$Log(Dur)_i$ = -0.266 +0.010$(SrCr)_i$ + 0.001$(Succ)_i$						

*Degree of freedom for each intercept is 1

Table 5 : Survival functions by using Gamma and Inverse Gaussian generalized linear model and Kaplan Meier

Duration of diabetes	S_{Gamma}(t)	$S_{Inverse\ Gaussian}$(t)	S_{KM}(t)
7	0.9075	0.9287	0.8333
8	0.7500	0.7245	0.6852
9	0.6019	0.6298	0.5370
10	0.4630	0.4595	0.4074
11	0.3703	0.3989	0.3519
12	0.3056	0.3220	0.2778
13	0.2407	0.2540	0.2222
14	0.1852	0.1755	0.1667
15	0.1019	0.0683	0.0556
19	0.0370	0.0370	0.0370
20	0.0185	0.0185	0.0185
25	0.0000	0.0000	0.0000

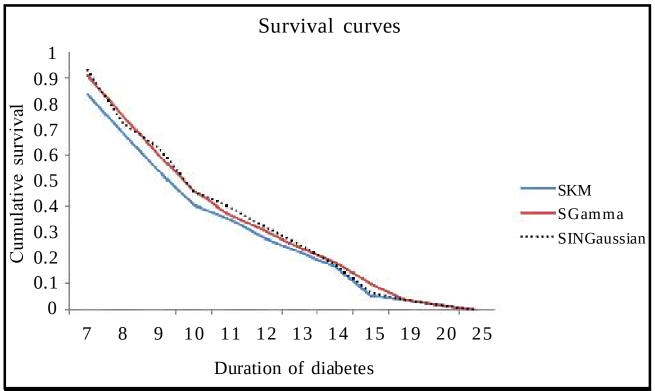

Fig. 1 : Survival curves using Gamma and Inverse Gaussian generalized linear model and Kaplan Meier Method

the gamma and inverse Gaussian distributions are used to estimate the survival functions. The survival estimates obtained by Kaplan Meier method, gamma and inverse Gaussian distributions are shown in Table 5. Fig. 1 shows the survival curve plotted using the gamma and inverse Gaussian generalized linear models and Kaplan Meier. The mean duration of diabetes of patients who develop diabetic nephropathy under gamma and inverse Gaussian distributions is 10.3 while under Kaplan Meier is 9.9.

Diabetic nephropathy is the leading cause of chronic kidney diseases and end stage renal failure. Throughout the world the number of the people developing type 2-DM is increased dramatically (WHO). In the course of diabetes mellitus diabetic nephropathy occurs in 30%-40% in patients with type 1 diabetes (USRDS) and in 25%-40% in patients with type 2-diabetes (Hall, 2006). So researchers are making different efforts to apply technologies to come up with results which are helpful to the medical field. We have also made a contribution in this regard by obtaining the survival functions of the type 2 diabetic patients. Gamma distribution comes out to be the best fit for model I and inverse Gaussian distribution is best fit for model II on comparing their AIC,BIC and AIC_C values. Survival function of type 2 diabetic patients are obtained based on gamma and inverse Gaussian distributions. And these survival functions are compared with the estimates obtained by the Kaplan Meier method. The survival function based on gamma and inverse Gaussian distributions and those obtained by the Kaplan Meier method are approximately same. Model II (proposed model) provides an alternative approach to obtain the survival function of type 2 diabetic patients.

Authors' affiliations:
Yasmeena Ismail, S.A. Mir and M.A. Bhat, S.K. University of Agricultural Sciences and Technology of Kashmir, Shalimar, Srinagar (J&K) India

■ REFERENCES

Akaike, H. (1973). Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 255-265.

Akram, M., Ullah, M.A. and Taj, R. (2007). Survival analysis of cancer patients using parametric and non-parametric approaches. *Pakistan Veterinary J.*, **27** : 194.

Cox, C., Chu, H., Schneider, M.F. and Muñoz, A. (2007). Parametric survival analysis and taxonomy of hazard functions for the Generalized Gamma Distribution. *Statistics Med.*, **26** :

4352-4374.

Cox, J. and Mann, M. (2008). Maxquant enables high peptide identification rates, individualized Ppb-Range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.*, **26** : 1367-1372.

Grover, G., Sabharwal, A.S.A. and Mittal, J. (2013). An application of gamma generalized linear model for estimation of survival function of diabetic nephropathy patients. *Internat. J. Statistics Med. Res.*, **2** : 209-219.

Hakulinen, T. and Tenkanen, L. (1987). Regression analysis of relative survival rates. *Appl. Statistics*, **36** (3) : 309-317.

Hall, Phillip M. (2006). Mechanisms in Diabetic Nephropathy Prevention of Progression in Diabetic Nephropathy. *Diabetes Spectrum*, **19**(1): 18-24.

Hurvich, C.M. and Tsai, C.L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76** (2) : 297-307.

Karen, A. (2006). Application of the generalized linear model to the prediction of lung cancer survival. 2006; 1-18. http:// analytics. ncsu.edu/sesug/2006/ST09_06.PDF

Kass, R.E. and Raftery, A.E. (1995). Bayes factors. *J. American Statistical Association*, **90** : 773-795.

McCullagh, P. and Nelder, J.A. (1989). Generalized linear models, No. 37 in Monograph on Statistics and Applied Probability.”

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized Linear Models. *J. Royal Statistical Society. Series A (General),* **135** (3) : 370-384.

Schwarz, G. (1978). Estimating the dimension of a model. *The Ann. Statistics*, **6** : 461-464.

Stroup, W.W. and Kachman, S.D. (1994). Generalized Linear Mixed Models-an Overview. Annual Conference on Applied Statistics in Agriculture

US Renal Data System and USRDS (2003). Annual Data Report; Atlas of end stage renal diseases, in the united states. Bethesda MD. National Institute of Health. National Insuitute of Diabetes, Digestive and Kidney Disease.

Yuan, X., Hong, D. and Shyr, Y. (2007). Survival model and estimation for lung cancer patients 2007; 201-22. http:// capone.mtsu.edu/ dhong/YuanHongShyr07.pdf.

World Health Organisation (2004). The diabetes program.