

A Parametric Approach to Estimate Survival Time of Diabetic Nephropathy with Left Truncated and Right Censored Data

Gurprit Grover

Department of Statistics, University of Delhi, Delhi, India

E-mail: gurpritgrover@yahoo.com

Alka Sabharwal (Corresponding author)

Kirori Mal College, University of Delhi, Delhi, India

E-mail: sabharwal_alka@hotmail.com

Received: January 29, 2012 Accepted: March 1, 2012 Published: May 1, 2012

doi:10.5539/ijjsp.v1n1p128 URL: <http://dx.doi.org/10.5539/ijjsp.v1n1p128>

Abstract

Diabetic Nephropathy (DN) is the deleterious effect of diabetes mellitus on renal structure, or a function caused by it. The rate of rise in, serum creatinine is observed for the progression of DN. Retrospective data is collected from 132 patients. As 60 patients form uncensored cases, remaining 72 patients are censored cases. In this paper we have developed three models. The first model is based only on uncensored cases with left truncation. The second model also includes censored data. Fisher's information matrix has been applied to find variances and asymptotic confidence interval for the estimated parameters. The method of maximum likelihood is used to estimate unknown parameters in both the models. In the last model, onset times are arranged as to estimate mean onset time of DN for all the patients by using application of order statistic. We aim to predict onset time of DN for new diabetic patients.

Keywords: Diabetic Nephropathy, Fisher's information matrix, Left truncated and right censored data, Order statistic, Serum creatinine, Survival time, Weibull distribution

1. Introduction

Diabetic nephropathy (DN) develops in 20 - 40% of the patients within 10 to 15 years after the onset of diabetes. About one third of those affected eventually have progressive deterioration of renal function (Remmuzz, Schipath & Rugenetiti, 2002). Data from the UKPDS demonstrated that 25% of patients approximately with type-2 diabetes develop worse diabetic nephropathy by 10 years (UKPDS, 1998a, 1998b). Also, it is estimated that almost 50% of patients develop DN within 19 years from diagnosis of diabetes. According to the American Diabetic Association in year 2004, about 20-30% of patients with type-1 or type-2 diabetes develop nephropathy (ADA, 2004). Thus, DN is a serious problem in terms of financial load, morbidity and mortality in the developed world (Banerjee, Ghosh & Saha, 2005).

In many prospective and retrospective studies, survival data are subject to left truncation in addition to the usual right censoring. There is a widespread application and use of left truncation and right censored data (LTRC) in survival analysis. Alioum and Commenges have given proportional hazard models for arbitrarily censored and truncated data to estimate the distribution of time from diabetes onset to development of DN (Alioum & Commenges, 1996). Enzo, Vittorio, and others have applied conditional probability to find that a diabetic patient will develop a second complication, given that they had already developed the first complication (Enzo, Vittorio et al., 2003). They also propose the Bayes' formula for the same problem. Chappell, Fine and Jiang have considered semi-parametric analysis of survival data with LTRC to quantify the association between time to DN and diabetes related death and to estimate the probability of developing DN after being diagnosed with insulin dependent diabetes (Chappell, Fine & Jiang, 2005). Sparling, Younes, and Lachin have applied a parametric family of regression models for interval-censored event-time data accommodating both fixed (e.g. baseline) and time-dependent covariates. Their model employs a three-parameter family of survival distributions including special cases of Weibull, negative binomial, and log-logistic distributions, and can be applied to data with left, right, interval, or non-censored event times. The model is applied by them on diabetic patients' data to describe the effects of longitudinal measures of glycemia over time on the risk of progression of diabetic retinopathy (Sparling, Younes & Lachin, 2006).

Medical and epidemiological studies are mostly conducted with an interest in measuring the occurrence of an outcome event. In this paper we are estimating the time from onset of diabetes to development of DN subject to censoring and truncation. Weibull distribution is fitted on both uncensored and LTRC data. Weibull distribution fitting is tested through

(i) chi-square test for goodness of fit and (ii) log-cumulative hazard plot for uncensored cases. Fitting of the Weibull distribution for censored cases is tested by applying (i) Hollander and Proschan test and (ii) modified Cox-Snell residuals. The unknown parameters shape γ , and scale, of the distribution are estimated by the method of maximum likelihood. Fisher information matrix has been used to construct variances and asymptotic confidence intervals of the unknown parameters for LTRC data. The DN onset times of uncensored out of total cases are arranged in ascending order for the application of order statistic as: $t_{(1)}, t_{(2)}, \dots, t_{(i)}, \dots, t_{(n_1)}$. The probability density functions and means of minimum and maximum DN onset time have been estimated using the application of order statistic. Also, we have estimated mean DN onset times for all the patients who are under advanced nephropathy. Observed DN onset times have been compared with DN onset times obtained by using the application of order statistic. Besides introduction, the course of this paper is as follows: In section 2 developments of the models is discussed. Section 3 applies the models to type-2 diabetic patient data and section 4 contains discussion.

2. Methodology

2.1 Model for Uncensored but Left Truncated Data

Let $t_{(1)}, t_{(2)}, \dots, t_{(i)}, \dots, t_{(n_1)}$ be the exact DN onset times of n_1 individuals out of n patients which are known under this study. Left truncation arises when a subject is not included in the study because its diabetic nephropathy onset times (event of interest) originated prior to the starting time of the study. The DN onset times are assumed to follow Weibull distribution. The probability density and survival function of a $W(\lambda, \gamma)$ for the i^{th} ($i = 1, 2, \dots, n_1$) patient are given by:

$$f(t_i) = \frac{\gamma(\lambda)^\gamma t_i^{\gamma-1} \exp(-(\lambda t_i)^\gamma)}{\exp(-(5\lambda)^\gamma)} \quad (1)$$

$$S(t_i) = \frac{\exp(-(\lambda t_i)^\gamma)}{\exp(-(5\lambda)^\gamma)}; t \geq 5, \lambda > 0; \gamma > 0 \quad (2)$$

2.1.1 Estimation of Unknown Parameters Using Maximum Likelihood Estimation

The log likelihood function is given as follows:

$$\log L = n_1 \log \gamma + n_1 \gamma \log \lambda - n_1 (5\lambda)^\gamma + (\gamma - 1) \sum_{i=1}^{n_1} \log t_i - \sum_{i=1}^{n_1} (\lambda t_i)^\gamma \quad (3)$$

The maximum likelihood estimates of λ and γ is found by differentiating above function with respect to λ and γ , and equating the derivatives to zero. The resulting equations are:

$$n_1 - (\lambda)^\gamma [n_1 (5)^\gamma + \sum_{i=1}^{n_1} (t_i)^\gamma] = 0 \quad (4)$$

$$\frac{n_1}{\gamma} + n_1 \log \lambda + \sum_{i=1}^{n_1} \log t_i - n_1 (5\lambda)^\gamma \log 5\lambda - \sum_{i=1}^{n_1} (\lambda t_i)^\gamma (\log \lambda t_i) = 0 \quad (5)$$

The maximum likelihood estimates (MLE) of λ and γ obtained as $\tilde{\lambda}$, $\tilde{\gamma}$ by solving above two equations (4) and (5) simultaneously.

2.1.2 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time from the Diagnosis of Diabetes

Using MLE of λ and γ as $\tilde{\lambda}$, $\tilde{\gamma}$ we obtained estimated mean, variance and 95% confidence interval of DN onset time/survival time, which are given by:

$$E(\text{onset time}) = \int_5^\infty t \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t)^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(5\tilde{\lambda})^{\tilde{\gamma}})} dt \quad (6)$$

$$\text{Var}(\text{onset time}) = \int_5^\infty t^2 \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t)^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(5\tilde{\lambda})^{\tilde{\gamma}})} dt - E^2(\text{onset time}) \quad (7)$$

$$95\% \text{ C. I of onset time} = E(\text{onset time}) \pm 1.96 \sqrt{\text{Var}(\text{onset time})/n_1} \quad (8)$$

2.2 Model for Left Truncated and Right Censored Data

All subjects under study experienced an initial event E1 (diagnosed as diabetes as per ADA standards), but not all of them experienced a second event E2 i.e. the event of interest (onset of renal disease/diabetic nephropathy) till the study was terminated on November 2007. Right censoring happens when DN onset time of a subject is not completely observed as

a second event, the end time of the study is fixed and patients enter the study at different times i.e. duration of diabetes is different for different patients. The form of censoring is generalized Type I censoring (Lee & Wang, 2003). This can be explained as: suppose that there are four patients with different duration of time and study is terminated at fixed time, where each individual’s minimum diabetes duration is 5 years (left truncated). Then the starting time is backed up to 5 years for four individuals. DN onset time for first patient is observed before the end of the study, it is an uncensored case with ($\delta_1 = 1$); DN onset time for second patient cannot be observed before the end of the study, it is a censored case with ($\delta_2 = 0$); DN onset time for third patient is known and an uncensored case with ($\delta_3 = 1$) and DN onset time for fourth patient is unknown and is a censored case with ($\delta_4 = 0$). The DN onset time of first and third patient may not be same, also, censoring time of second and fourth may not be same, because patients under study are with different duration of diabetes.

There are n patients under study, out of which diabetic nephropathy onset times for n_1 patients are uncensored and DN onset time for remaining $n - n_1$ patients are censored. The DN onset time from the diagnosis of diabetes is a non-negative random variable which is assumed to follow Weibull distribution .The probability density function of a $W(\lambda, \gamma)$ for the i^{th} ($i = 1, 2, \dots, n$) patient is given by (Lee & Wang, 2003)

$$f(t_i) = \begin{cases} \frac{\gamma(\lambda)^\gamma t_i^{\gamma-1} \exp(-(\lambda t_i)^\gamma)}{\exp(-(\lambda T_i)^\gamma)}, & 5 \leq t_i \leq T_i \\ \frac{\exp(-(\lambda T_i)^\gamma)}{\exp(-(\lambda T_i)^\gamma)}, & t_i > T_i \end{cases} \tag{9}$$

2.2.1 MLE for Unknown Parameters for Right Censored and Fixed Left Truncated Weibull Distribution

When the censoring times T_i are different for different patients, the likelihood function for the survival times is defined as:

$$L = \prod_1^n (f(t_i))^{\delta_i} (S(T_i))^{1-\delta_i}$$

where δ_i is zero, if for the i^{th} patient DN onset time is censored and δ_i is unity when DN onset time is known. The corresponding log likelihood function is given by:

$$\begin{aligned} \log L &= \sum_1^n \delta_i \log \left(\frac{\gamma(\lambda)^\gamma t_i^{\gamma-1} \exp(-(\lambda t_i)^\gamma)}{\exp(-(\lambda T_i)^\gamma)} \right) + \sum_1^n (1 - \delta_i) \log \frac{\exp(-(\lambda T_i)^\gamma)}{\exp(-(\lambda T_i)^\gamma)} \\ &= \sum_1^n \delta_i (\log \gamma + \gamma \log \lambda + (\gamma - 1) \log t_i - (\lambda t_i)^\gamma + (\lambda T_i)^\gamma) + \sum_1^n (1 - \delta_i) ((\lambda T_i)^\gamma - (\lambda T_i)^\gamma) \end{aligned} \tag{10}$$

The maximum likelihood estimates of λ and γ is found by partially differentiating the above function with respect to λ and γ , and equating the derivatives to zero. The resulting equations are:

$$\lambda^\gamma \left(\sum_1^n \delta_i t_i^\gamma + \sum_1^n (1 - \delta_i) T_i^\gamma - n 5^\gamma \right) - \sum_1^n \delta_i = 0 \tag{11}$$

$$\lambda^\gamma \left(\sum_1^n \delta_i t_i^\gamma \log(\lambda t_i) + \sum_1^n (1 - \delta_i) \log(\lambda T_i) T_i^\gamma - n 5^\gamma \log(5\lambda) \right) - \sum_{i=1}^n \delta_i \left(\frac{1}{\gamma} + \log(\lambda t_i) \right) = 0 \tag{12}$$

The MLE of λ and γ are obtained as $\tilde{\lambda}, \tilde{\gamma}$ by solving above (11) and (12) equations by trial and error method.

2.2.2 Approximate Fisher Information Matrix to Construct Asymptotic Confidence Intervals of the Unknown Parameters

In this sub-section, the approximate Fisher information matrix of maximum likelihood estimators of parameters from Weibull distribution when the data are left truncated and right censored has been obtained, which can be used to construct asymptotic variance and confidence intervals of the parameters (Gross & Clark, 1975). The Fisher information matrix can be written as:

$$\begin{aligned} \begin{pmatrix} [-\frac{\partial^2}{\partial \lambda^2} \log L]_0 & [-\frac{\partial^2}{\partial \lambda \partial \gamma} \log L]_0 \\ [-\frac{\partial^2}{\partial \gamma \partial \lambda} \log L]_0 & [-\frac{\partial^2}{\partial \gamma^2} \log L]_0 \end{pmatrix}^{-1} &= \begin{pmatrix} Var(\tilde{\lambda}) & Cov(\tilde{\lambda}, \tilde{\gamma}) \\ Cov(\tilde{\lambda}, \tilde{\gamma}) & Var(\tilde{\gamma}) \end{pmatrix} \\ \frac{\partial^2}{\partial \lambda^2} \log L &= \gamma \lambda^{\gamma-1} \left(\sum_1^n \delta_i t_i + \sum_1^n (1 - \delta_i) T_i^\gamma - n 5^\gamma \right) \end{aligned} \tag{13}$$

$$\frac{\partial^2}{\partial \lambda \partial \gamma} \log L = \frac{\partial^2}{\partial \gamma \partial \lambda} \log L = \left[\left(\sum_1^n \delta_i + n5^\gamma \lambda^\gamma (1 + \gamma \log(5\lambda)) \right) / \lambda \right] - \lambda^{\gamma-1} \left[\sum_1^n \delta_i t_i^\lambda (\gamma \log(\lambda t_i) + 1) + \sum_1^n (1 - \delta_i) T_i (\gamma \log(\lambda T_i) + 1) \right] \quad (14)$$

$$\frac{\partial^2}{\partial \gamma^2} \log L = \frac{\sum_1^n \delta_i}{\gamma^2} + \lambda^\gamma \left[\sum_1^n \delta_i \log^2(\lambda t_i) t_i^\gamma + \sum_1^n (1 - \delta_i) \log^2(\lambda T_i) T_i^\gamma - n \log^2(5\lambda) 5^\gamma \right] \quad (15)$$

The approximate 95% confidence intervals for λ and γ are $\tilde{\lambda} \pm 1.96 \sqrt{\text{Var}(\tilde{\lambda})}$ and $\tilde{\gamma} \pm 1.96 \sqrt{\text{Var}(\tilde{\gamma})}$, respectively; where $\text{Var}(\tilde{\lambda})$ and $\text{Var}(\tilde{\gamma})$ are the diagonal elements of the variance-covariance matrix.

2.2.3 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time

Using maximum likelihood estimates of λ and γ for LTRC data as $\tilde{\lambda}$, $\tilde{\gamma}$ we obtained the estimated mean, variance and 95% confidence interval of onset time /survival time which are given by:

$$E(\text{DN onset time} | 5 \leq \text{time} \leq 26.6) = \frac{\int_5^T t \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t)^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} dt}{\int_5^T \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t)^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} dt} \quad (16)$$

$$E((\text{DN onset time})^2 | 5 \leq \text{time} \leq 26.6) = \frac{\int_5^T t^2 \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t)^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} dt}{\int_5^T \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t)^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} dt} \quad (17)$$

$$\text{Var}(\text{DN onset time}) = E((\text{DN onset time})^2 | 5 \leq \text{time} \leq 26.6) - E^2(\text{DN onset time} | 5 \leq \text{time} \leq 26.6) \quad (18)$$

$$95\% \text{ C. I} = E(\text{DN onset time}) \pm 1.96 \sqrt{\text{Var}(\text{DN onset time})/n} \quad (19)$$

Where, T denotes the maximum duration of diabetes among n_1 (uncensored) patients under study. The above results are compared with those obtained for uncensored cases and observed cases.

2.3 The Probability Density Function and Mean Time of First, Last and Order Statistic for the Patients Who Proceed Towards Nephropathy

Let $\tau_1 \leq \tau_2 \leq \dots \leq \tau_i \leq \dots \leq \tau_n$ be the ordered duration of n diabetic patients and n_1 be the number of uncensored diabetic nephropathy of patients with ordered onset times as, $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(i)} \leq \dots \leq t_{(n_1)}$ where $t_{(1)}$ and $t_{(n_1)}$ denote the minimum and maximum time taken by type-2 diabetic patient proceeding towards nephropathy, respectively, among n_1 patients. The DN onset times $t_{(1)}, t_{(2)}, \dots, t_{(i)}, \dots, t_{(n_1)}$ are following left truncated Weibull distribution with parameters λ and γ as given in equation (1).

The probability density function and mean of first order statistic or the patient who has taken minimum time to proceed towards nephropathy is given as:

$$f_{t_{(1)}}(t) = n_1 \left[\frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \right]^{n_1-1} \gamma (\lambda)^\gamma (t)^{\gamma-1} \frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \quad (20)$$

$$E(t_{(1)}) = \int_5^T t_{(1)} n_1 \left[\frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \right]^{n_1-1} \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} (t_{(1)})^{\tilde{\gamma}-1} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} dt_{(1)} \quad (21)$$

Here we are dealing with right censored data and T denotes the maximum duration of diabetes among n_1 patients under study.

The probability density function and mean of r^{th} order statistic is given as:

$$f_{t_{(r)}}(t) = rB(n_1, r) \left[1 - \frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \right]^{r-1} \gamma (\lambda)^\gamma (t)^{\gamma-1} \frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \left[\frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \right]^{n_1-r} \quad (22)$$

$$E(t_{(r)}) = rB(n_1, r) \int_5^T t_{(r)}^{\tilde{\gamma}} \left[1 - \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} \right]^{r-1} \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} \left[\frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{5}\tilde{\lambda})^{\tilde{\gamma}})} \right]^{n_1-r} dt_{(r)}$$

The probability density function and mean of n_1^{th} order statistic or the patient who has taken maximum time to proceed towards nephropathy is given as:

$$f_{t_{(n_1)}}(t) = n_1 t_{(n_1)}^{\gamma-1} \left[1 - \frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \right]^{n_1-1} \gamma (\lambda)^\gamma \frac{\exp(-(\lambda t)^\gamma)}{\exp(-(\lambda 5)^\gamma)} \quad (23)$$

$$E_{t_{(n_1)}}(t) = n_1 \tilde{\gamma}(\tilde{\lambda})^{\tilde{\gamma}} \left[\int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{\exp(-(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-(\tilde{\lambda}5)^{\tilde{\gamma}})} dt_{(n_1)} + (-1)(n_1 - 1) \int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{\exp(-2(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-2(\tilde{\lambda}5)^{\tilde{\gamma}})} dt_{(n_1)} + \right. \\ \left. (-1)^2((n_1 - 1)(n_1 - 2)/2) \int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{\exp(-3(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-3(\tilde{\lambda}5)^{\tilde{\gamma}})} dt_{(n_1)} + \dots + (-1)^{n_1-1} \int_5^T t_{(n_1)}^{\tilde{\gamma}} \frac{\exp(-n_1(\tilde{\lambda}t)^{\tilde{\gamma}})}{\exp(-n_1(\tilde{\lambda}5)^{\tilde{\gamma}})} dt_{(n_1)} \right]$$

Thus, the mean time of first or minimum, last or maximum and r^{th} order statistic for the patients who proceed towards nephropathy are computed.

2.4 Model Checking

There are a number of techniques for evaluating the fit of parametric survival models, including analogues of residuals and influence diagnostics (Collett, 2003). To assess the adequacy of the present model, two methods have been applied for the testing of the distribution assumption for both the uncensored and censored data. Firstly for uncensored data, log-cumulative hazard plot is used, if the single sample of the survival data is available. For obtaining log-cumulative hazard plot, first compute the Kaplan Meier estimate of the survivor function, $\hat{S}(t)$ for each uncensored observation and then find the survivor function, $\hat{S}(t)$ of the fitted distribution. If the plot of $\log(-\log S(t))$ against $\log(t)$ is an approximate straight line, and close to the line obtained from plotting $\log(-\log \hat{S}(t))$ against $\log(t)$ then the fitted distribution assumption is tenable. Secondly, the fitting has been tested by applying Karl-Pearson’s chi-square goodness of fit test. In case of censored data, firstly, the modified Cox-Snell residual plot is used for fitting distribution, proposed by Crowley and Hu. To include censored observation, Cox-Snell residual is modified by the addition of a positive constant Δ , the excess residual. Suppose that the i^{th} survival time is a censored observation, t_i^* , and let t_i be the actual, but unknown, survival time, so that $t_i > t_i^*$. Then modified Cox-Snell residual can be used for i^{th} censored observation as: $r'_{C_i} = r_{C_i}$ for uncensored and $r'_{C_i} = r_{C_i} + \Delta$ for censored observation, where $r_{C_i} = -\log \hat{S}_i(t_i^*) = \hat{H}_i(t_i^*)$. The, $\hat{S}_i(t_i^*)$ and $\hat{H}_i(t_i^*)$ are the estimated KM survival and cumulative hazard functions, respectively, for the i^{th} individual at the censored survival time. After computing r'_{C_i} , it is plotted against the $-\log(S(t))$ where $S(t)$ is the survival function of the fitted distribution. If this plot is close to a straight line passing through the origin, this suggests that model fitted to the data is satisfactory. Secondly, Hollander and Proschan test has been applied to test the fitted distribution survival function with the survival function obtained by applying Kaplan Meier method (Hollander & Proschan, 1979).

3. Applications

The methods discussed above are applied to the following data concerning nephropathy of type-2 diabetic patients. Up-to-date pathological reports/records of 132 diabetic patients, using a common path lab, are collected through a house to house Survey. Retrospective study is conducted on the collected data. Patients with less than 5 years diabetic history are not included or data considered is left truncated which arises when a patient is not included in the study because its DN onset time (event of interest) originated prior to the starting time of the study. Under truncation these individuals are never considered for inclusion into the study.

On the basis of serum creatinine the event of interest nephropathy, using reference as the rate of rise in SrCr, a well-accepted marker for the progression of DN, (creatinine value 1.4 to 3.0 mg/dl) is the indicator for impaired renal function (Adler, Stevens, Manley et al., 2003). Also, the renal health/dysfunction can be estimated from serum creatinine level for 132 patients under study. Thus, at the end of the study, out of 132 patients, there are only 60 diabetic nephropathy cases and remaining 72 are treated as censored. Data is simultaneously left truncated and right censored. Table-1 represents the duration-distribution of type-2 diabetic patients who developed diabetic nephropathy. It shows that out of 60 DN cases, 16.67% are with less than 10 years of diabetes, 25% with 10 to 15 years, 38.33% with 15 to 20 years and 20% are of more than or equal to 20 years of duration of diabetes. In all there are 45.45% cases with diabetic nephropathy.

The minimum age at which diabetes is diagnosed is found to be 29 yrs and maximum age as 58 years, as available from data. Table 2 represents descriptive statistics of 132 patients giving minimum, maximum and mean \pm S.D of the variables; age at diagnosis, duration of disease and SrCr for two groups; censored or non diabetic nephropathy (NDN) and uncensored or diabetic nephropathy group.

<Table 1-2>

3.1 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time from the Diagnosis of Diabetes for Uncensored Cases

For modelling the survival time (from the diabetes diagnosis till the onset of nephropathy) for the uncensored cases only from the data, the minimum onset time is observed as 6 years and the maximum onset time is 26.6 years. Left truncated Weibull distribution is assumed for this case. Fitting is tested, firstly, by applying Karl-Pearson’s chi-square goodness of fit test and calculated chi-square value came out to be 2.732 with $p=0.435$. Therefore, we conclude that there is

insufficient evidence to say that data are not from left truncated Weibull distribution with $\tilde{\lambda}=0.079515$ and $\tilde{\gamma}=2.811274$. Secondly, graphical method, log-cumulative hazard plot is also used to test the fitting of distribution. For obtaining this plot, the values $\log(-\log(\hat{S}(t)))$, where, $\hat{S}(t)$ is Kaplan-Meier estimate of survival, and $\log(-\log(S(t)))$, where $S(t)$ is fitted distribution estimates of survival are computed. Then plotting, $\log(-\log(\hat{S}(t)))$ against $\log(\text{survival time})$ and $\log(-\log(S(t)))$ against $\log(\text{survival time})$, the resulting log-cumulative hazard plot is shown in Figure 1. In this figure the line $\log(-\log(S(t)))$ against $\log(\text{survival time})$ is reasonably straight and close to the other line. This means that the assumption of left truncated Weibull distribution for the uncensored data of DN onset times is quite plausible (Collett, 2003).

Using maximum likelihood estimates as $\tilde{\lambda}=0.079515$ and $\tilde{\gamma}=2.811274$, estimated mean DN onset time is obtained as 15.19228, variance of DN onset time as 17.09349 and 95% confidence interval as (14.14615, 16.23842). These results are also compared with mean, variance and 95% confidence interval of DN onset time obtained from the observed data, which are displayed in Table 4.

<Figure 1>

3.2 Estimated Mean, Variance and 95% Confidence Interval of DN Onset Time from the Diagnosis of Diabetes by Including Censored Cases

For modelling the DN onset times including censored cases, the minimum duration of diabetes is found to be 5.6 years and maximum as 27 years as available from the data. From the data we have only partial information about these 72 cases, as they are censored cases. For this model, data is left truncated and right censored simultaneously, and Weibull distribution is assumed for this case also. Fitting is tested, firstly, on applying Hollander and Proschan goodness of fit test for censored data. The observed value of the test statistic is 1.256186 with $p=0.182$, which does not fall in the rejection area. Therefore, we conclude that there is insufficient evidence to say that data are not from Weibull distribution with $\tilde{\lambda}=0.066834$ and $\tilde{\gamma}=1.78211$ estimated by applying maximum likelihood method. Secondly, graphical method, i.e., modified Cox-Snell residual plot (Gross & Clark, 1975) is also used to test the fitting of distribution. For obtaining this plot, $r = -(\log(S(t)))$ is plotted against $-\log(\hat{S}(t))$, where $S(t)$ and $\hat{S}(t)$ are same as defined in section 3.1. The resultant line is modified Cox-Snell residual plot and line is close to straight line as shown in Figure 2.

Therefore, from both the methods Weibull distribution is found to be an appropriate parametric model for DN onset times.

3.2.1 Construct Asymptotic Confidence Intervals of the Estimated Parameter by Using Fisher Information Matrix

Again using maximum likelihood estimators as $\tilde{\lambda}=0.066834$ and $\tilde{\gamma}=1.78211$ of unknown parameters for censored case, the Fisher information matrix is used to obtain the asymptotic confidence intervals, $Var(\tilde{\lambda})$ and $Var(\tilde{\gamma})$, the results are depicted in Table 3. The estimated mean DN onset time is found to be 13.9058552 with corresponding variance and 95% confidence interval are obtained as 13.953207 and (13.4213197, 14.695785), respectively. The results are displayed in Table 4.

<Figure 2>

<Table 3-4>

3.3 Estimated Mean Time of First, Last and r^{th} Order Statistic for the Patients Who are under Advanced Nephropathy Group

For modelling the nephropathy onset time from diagnosis of diabetes for estimating mean DN onset time for every patient, the application of order Statistic has been used. We have used this model to estimate mean DN onset time for 27 patients with mean SrCr = 1.91mg/dl, who are under advanced nephropathy group out of 60 uncensored cases. Estimated minimum and maximum mean DN onset time are obtained and the results came out to be 5.646346 years and 24.58083 years respectively. Also, estimated mean DN onset times for 27 patients are compared with observed onset times and results are displayed in Table 5. Comparison is also illustrated graphically in Figure 3. It is observed from graph that the mean values of onset of DN for each patient up to 10^{th} order are close to observed onset time, but as the order increases the difference between observed and estimated mean value increases as higher ordered values depend on all the prior ordered values. Also, it can be observed from Table 5. The mean and variance of means of mean DN onset time obtained through application of order statistic are 15.71540 and 37.949, respectively. SPSS for Windows, Version 15 and MATLAB, Version 6.5 statistical packages were used for the calculation and analysis.

<Table 5>

<Figure 3>

4. Discussion

The aim of this paper is to estimate nephropathy onset time arising out of type-2 diabetic patients. Previous studies reveal that type-2 DM is the most common form of diabetes constituting 90% of the diabetic population and nephropathy is a life-threatening complication of diabetes mellitus. The DN onset times calculations throughout this paper are based on the values of serum creatinine, since, Bio-Stratum at their 64th Scientific Sessions ADA meet accepted rate of rise in serum creatinine as a marker for the progression of diabetic nephropathy.

There is a widespread application and use of left truncation and right censored data (LTRC) in survival analysis. Data used in this paper are simultaneously left truncated and right censored. Thus, no information was available for the patients whose DN onset time appear prior to five years and partial information was available for the patient whose DN onset times could not be observed till the end of study.

The main objectives of this paper are to estimate mean, variance and 95% confidence interval of DN onset time by fitting (i) model for uncensored cases, and (ii) model including censored cases. Weibull distribution is assumed for both the cases as it is very flexible and because of flexible shape and ability to model a wide range of failure rates, the Weibull has been successfully used in many applications (Collett, 2003; Lee & Wang, 2003). The distribution fitting is tested through graphical method, to access whether a particular theoretical distribution provides an adequate fit to the data (Lee & Wang, 2003). Also, most common Chi-square goodness of fit test is used for testing the fitting for uncensored data and Hollander and Proschan goodness of fit test is used for testing the fitting of data that includes censored observations. Weibull distribution is found to be appropriate in both the cases. Previous studies suggest that when survival time has a specific statistical distribution; the statistical power of parametric distribution is higher than nonparametric and semi-parametric models.

Firstly, for uncensored cases using Weibull model the mean of DN onset time is found to be 15.19228 years whereas the the observed mean of DN onset time is 15.21333 years. Thus, the estimated and the observed mean DN onset times are found to be remarkably similar. Secondly, mean of DN onset time, 13.905855 years, is obtained by including censored cases. The mean of DN onset time based on uncensored observations is higher than the case which includes censored observations and this matches with Li and Lagakos finding suggesting that considering only uncensored cases will increase the mean of the survival time (Li & Lagakos, 1997). The mean DN onset times obtained from both the cases are almost consistent with the previous findings which suggest that diabetic nephropathy develop within 10 to 15 years after the onset of diabetes (ADA, 2004).

In part three, the model is applied with the application of order statistic to estimate minimum and maximum mean DN onset time and also, mean DN onset time for all the 27 patients out of 60, who are under the advanced nephropathy group. It is found from Table 5, which is visible through graph, that estimating mean DN onset time using the application of order statistic is a useful exercise if the sample size is small. This is further, verified by applying paired t-test. We have divided 27 patients in accordance with the ascending order of DN onset time into two groups of sizes 14 and 13. By applying paired t-test for comparing the observed DN onset times with those obtained from the application of order statistic on first group(comparing first 14 ordered values) the test statistic value is found to be 1.659 with $p=0.121$, ($p > 0.05$) on 13 degrees of freedom. Thus, we conclude that there is not enough evidence to reject the hypothesis as there is no significant difference between the estimated and the observed DN onset times. Then by applying paired t-test on second group (ordered values from 15 to 27) the test statistic value is found to be 14.645 with $p < 0.001$ on 12 degrees of freedom. From this we conclude that difference between observed and estimated mean DN onset time increases with an increase in the order of the observation. The minimum and maximum mean of DN onset times are found to be 5.646346 years and 24.58083 years respectively.

Weibull model can empirically fit a wide range of data. Thus, Weibull distribution may be used to model the survival distribution of a population with increasing, decreasing, or constant risk and can be taken as a baseline model for future studies. The model can be applied for estimating survival times of other complication arising out of type-2 diabetes such as retinopathy, CVD and others. The application of order statistic can be further explored with other survival models and also for other biomedical complications. The major use of estimating the nephropathy onset time of diabetic patient is that current as well as future DN onset time of new diabetic patients can be predicted.

References

- Adler, A. I., Stevens, R. J., Manley, S. E. et al. (2003). Development and progression of nephropathy in type 2 diabetes. United Kingdom Prospective Diabetes Study, 64. *Kidney International*, 63, 225-232. <http://dx.doi.org/10.1046/j.1523-1755.2003.00712.x>
- Alioum, A., & Commenges, D. (1996). A Proportional Hazard models for Arbitrarily Censored and Truncated Data. *Biometrics*, 52, 512-524. <http://dx.doi.org/10.2307/2532891>

- American Diabetes Association. (2004). Nephropathy in Diabetes. *Diabetes Care*, 27, S79-S83. <http://dx.doi.org/10.2337/diacare.27.2007.S79>
- Banerjee, S., Ghosh, U. S., & Saha, S. J. (2005). Role of GFR Estimation in Assessment of the Status of Nephropathy in Type 2 Diabetes Mellitus. *JAPI*, 53, 181-184.
- Chappell, R., Fine, J., & Jiang, H. (2005). Semiparametric Analysis of Survival Data with Left Truncation and Right Censoring. *Biometrics*, 61, 567-575. <http://dx.doi.org/10.1111/j.1541-0420.2005.00335.x>
- Collett, D. (2003). Modeling Survival data in Medical Research. Second edition. London: Chapman and Hall/CRC.
- Enzo, B., Vittorio, C., Nicola, M. Di, et al. (2003). Probabilistic approach to developing nephropathy in diabetic patient with retinopathy. *Statistics in Medicine*, 20(24), 3889-3897.
- Gross, J. A., & Clark, A.V. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences. New York: John Wiley and Sons.
- Hollander, M., & Proschan, F. (1979). Testing to determine the Underlying Distribution Using Randomly Censored Data. *Biometrics*, 35, 393-401. <http://dx.doi.org/10.2307/2530342>
- Lee, Elisa. T., & Wang, J. W. (2003). Statistical Methods for Survival Data Analysis, third edition. New York: John Wiley and Sons.
- Li, Q. H., & Lagakos, S. W. (1997). Use of the Wei-Lin-Weissfeld Method for the Analysis of a Recurring and a Terminating Event. *Statistics in Medicine*, 16, 925-940. [http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19970430\)16:8<925::AID-SIM545>3.0.CO;2-2](http://dx.doi.org/10.1002/(SICI)1097-0258(19970430)16:8<925::AID-SIM545>3.0.CO;2-2)
- Remmuzz, G., Schipath, A., & Rugenititi, P. (2002). Nephropathy in patients with type 2 diabetes mellitus. *New Eng. J. Med.*, 346, 1145-51. <http://dx.doi.org/10.1056/NEJMc011773>
- Sparling, Y. H., Younes, N., & Lachin, J. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*, 7(4), 599-614. <http://dx.doi.org/10.1093/biostatistics/kxj028>
- U. K. Prospective Diabetes Study Group. (1998). Intensive blood glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type-2 diabetes. UKPDS 33. *Lancet*, 352, 837-853. [http://dx.doi.org/10.1016/S0140-6736\(98\)07019-6](http://dx.doi.org/10.1016/S0140-6736(98)07019-6)
- U. K. Prospective Diabetes Study Group. (1998). Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes, UKPDS 38, *BMJ*, 317, 703-713. <http://dx.doi.org/10.1136/bmj.317.7160.703>
- Villar, E., Chang, S. H., & McDonald, S. P. (2007). Incidences, Treatments, Outcomes, and Sex Effect on Survival in Patients with End-Stage Renal Disease by Diabetes Status in Australia and New Zealand (1991-2005). *Diabetes Care*, 30(12), 3070-76. <http://dx.doi.org/10.2337/dc07-0895>

Table 1. Duration-distribution of Diabetic patients till Nov'2007, who developed Diabetic Nephropathy

Duration of Diabetes (years)	Number of diabetic patients	Diabetic Nephropathy (%)
<10	56	16.67
10-15	20	25
15-20	31	38.33
≥20	25	20
Total	132	45.45

Table 2. Descriptive statistics of 132 patients giving minimum, maximum and mean \pm standard deviation of age at diagnosis, duration of diabetes, and serum creatinine for two groups i.e. NDN (censored) and DN (uncensored) Group

Variable	Statistic	Censored (N=72)	Uncensored (N=60)
Age at Diagnosis (years)	Minimum	35	29
	Maximum	58	56
	Mean \pm S.D	44.0108 \pm 4.36	45.0034 \pm 5.2824
Duration of Disease (yrs)	Minimum	5.6	6
	Maximum	27	26.6
	Mean \pm S.D	10.2784 \pm 5.5	14.0931 \pm 5.0528
Serum Creatinine (mg/dl)	Minimum	0.71	1.29
	Maximum	1.39	2.21
	Mean \pm S.D	0.9780 \pm 0.12616	1.6686 \pm 0.28233

Table 3. Variance and asymptotic confidence intervals for the estimates of the parameters obtained by using Fisher information matrix

Parameter	Maximum likelihood estimator	Variance	95% Confidence interval
$\tilde{\lambda}$	0.066834	0.001078	(0.002487, 0.131181)
$\tilde{\gamma}$	1.78211	0.000391	(1.820866, 1.743354)

Table 4. Estimated mean, variance and 95% confidence interval of DN onset time obtained from observed and fitting (i) left truncated Weibull distribution on uncensored cases (ii) left truncated Weibull distribution by including censored cases

Cases	Mean	Variance	95% C.I
Observed uncensored cases	15.21333	22.87686	(14.003195,16.42357)
LT Weibull uncensored cases	15.19228	17.09349	(14.14615,16.23842)
Weibull LTRC data	13.9058552	13.953207	(13.4213197,14.6957854)

Table 5. Estimated and observed DN onset time for 27 patients who are under advance diabetic nephropathy group

Ordered Observation	Diabetes Duration(yrs)	Serum Creatinine (mg/dl)	Diabetic Nephropathy Onset time Observed (yrs)	Estimated Diabetic Nephropathy Onset time (yrs)
1	7.8	1.98	6	5.646346
2	12	1.7	6.6	6.215418
3	13	1.54	7.6	6.297053
4	13	1.66	7.9	7.321488
5	13.6	1.97	8.6	7.522939
6	15	2.14	9.2	10.32963
7	15	2.11	9.6	10.80505
8	15.6	2.21	11	12.42911
9	16	2.18	11.6	12.6954
10	16	1.99	12	14.13889
11	16	1.62	13	14.72397
12	16	1.66	13.6	15.22681
13	16.6	1.89	13.6	15.86091
14	17	1.57	14	16.52023
15	17	1.97	14.6	17.30451
16	17.2	1.6	15	17.76943
17	17.6	2.11	15.6	18.49336
18	18	1.79	15.6	19.55992
19	18	2.18	16	20.68734
20	19	2.19	16.6	20.79337
21	20	1.79	18	21.0774
22	20	1.74	18	21.9298
23	23	1.79	19.6	23.04411
24	24.4	2.05	19.6	23.28542
25	26	2.01	21	23.66707
26	26	1.94	22	24.08705
27	26.6	2.14	22.6	24.58083

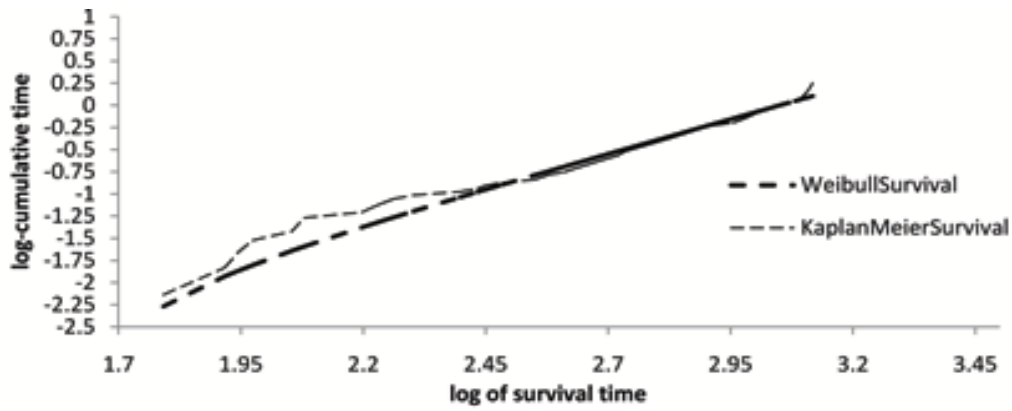


Figure 1. Fitting of Weibull distribution for the uncensored data by plotting (i) $\log(-\log(\hat{S}(t)))$ against $\log(\text{survival time})$ (ii) $\log(-\log(S(t)))$ against $\log(\text{survival time})$

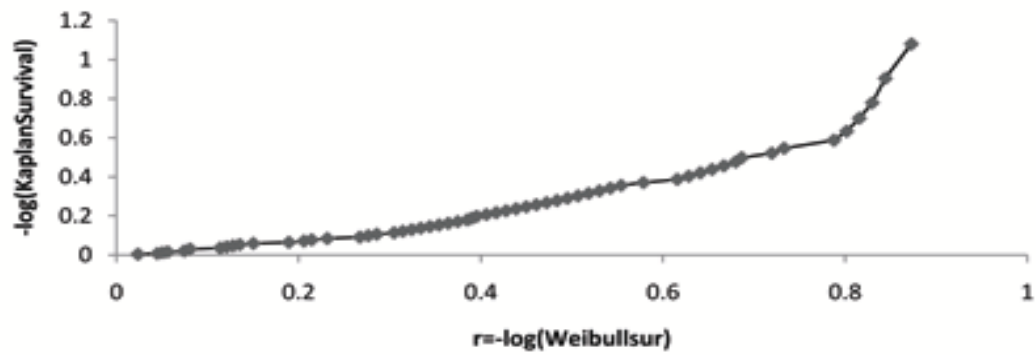


Figure 2. Fitting of Weibull distribution for censored data by modified Cox-Snell residual method by plotting $r=-\log(\text{Weibull Survival})$ against $(-\log(\text{Kaplan-Meier Survival}))$

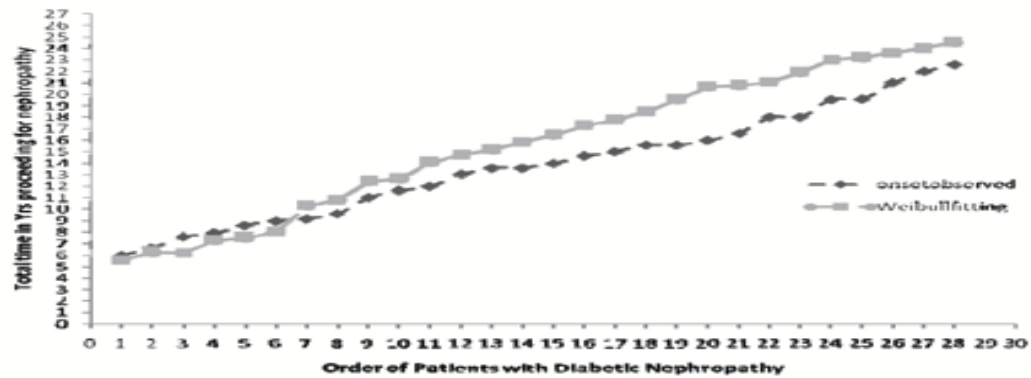


Figure 3. Estimated and observed survival time for 27 patients who are under advance diabetic nephropathy group