Sabharwal, Alka; Grover, Gurprit

Article

COMPARISON OF DIABETIC NEPHROPATHY ONSET TIME OF TWO GROUPS WITH LEFT TRUNCATED AND RIGHT CENSORED DATA

Statistics in Transition New Series

Provided in Cooperation with:
Polish Statistical Association

Suggested Citation: Sabharwal, Alka; Grover, Gurprit (2018) : COMPARISON OF DIABETIC NEPHROPATHY ONSET TIME OF TWO GROUPS WITH LEFT TRUNCATED AND RIGHT CENSORED DATA, Statistics in Transition New Series, ISSN 2450-0291, Exeley, New York, NY, Vol. 19, Iss. 3, pp. 527-549,
https://doi.org/10.21307/stattrans-2018-029

This Version is available at:
https://hdl.handle.net/10419/207912

STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 527–549, DOI 10.21307/stattrans-2018-029

COMPARISON OF DIABETIC NEPHROPATHY ONSET TIME OF TWO GROUPS WITH LEFT TRUNCATED AND RIGHT CENSORED DATA

Alka Sabharwal[1], **Gurprit Grover**[2]

ABSTRACT

The present paper is concerned with the comparison of the nephropathy onset time of type-2 diabetic patients, grouped on the basis of gender and age at the time of diabetes diagnosis. Diabetic Nephropathy (DN) onset time is assumed to follow Weibull distribution with fixed left truncation. The likelihood ratio test is applied on uncensored cases and Thoman and Bain two sample tests is applied with generated left truncated Weibull distributions. To avoid the model validity issues for left truncated and right censored data (LTRC), the nonparametric approach, suggested by Kaplan and Meier, is used to compare the survival function of two groups over different time periods. Another method based on median survival time of the pooled group is applied to compare the survival function of two groups with LTRC data. The major advantage of developing methods for comparing the nephropathy onset times of DM patients is that the expected DN onset time of new DM patients can be predicted depending on the patient group.

Key words: Kaplan-Meier survival function; survival time; Weibull distribution.

1. Introduction

Diabetes is considered to be the primary cause of nephropathy if subjects develop diabetes after the age of 35 years and if diabetes was present in the subjects for more than 5 years before the initiation of renal replacement therapy (Meredith et al. 2009). Type-2 diabetes is known as adult-onset diabetes as it is primarily seen in middle-aged adults over the age of 40. (Brenner et al. 2003). It has been predicted by Viswanathan (2004) that worldwide the prevalence of diabetes in adults would increase to 5.4% by the year 2025 from the prevalence rate 4.0% in 1995. Rodby (1997) study suggested that type-2 diabetic males are at greater risk of developing nephropathy. Also, it has been shown by Wagle (2010) that serum creatinine levels in males are significantly higher than in females.

[1] Kirori Mal College, Department of Statistics, University of Delhi, India. E-mail: alkasabh@gmail.com.
[2] Department of Statistics, University of Delhi, India. E-mail: gurpritgrover@yahoo.com.

Survival time comparison is one of the main goals of survival studies. Biomedical studies often compare the distributions of failure/survival time variables among two or more groups. Rossing et al. (1996) study compared three levels of albuminuria in insulin dependent diabetic patients. Joss et al. (2002) work concludes that survival time of type-2 diabetic patients, once diabetic nephropathy has developed, becomes even worse after starting dialysis. Bruce, Sheppard and others (2004) compared survival times of three categories: no diabetes, diabetes without peripheral vascular disease and renal failure, and diabetes with peripheral vascular disease and/or renal failure. Ashfaq et al. (2006) compared survival time of diabetic and non-diabetic groups to observe the effect of vein graft intervention. In all the above mentioned studies the authors estimated the survival function by Kaplan-Meier method and applied log rank test to compare the survivability of groups. Villar et al. (2007) suggested Cox proportional hazard model to study the effect of renal replacement therapy on the survival times of type-1 diabetic, type -2 diabetic and non-diabetic patients. Jianguo Sun and others (2008) compared survival times of two groups by applying generalized log-rank test. They discuss a class of generalized log-rank tests for incomplete survival data and establish their asymptotic properties and illustrated their study with diabetic patient data.

There are two broad approaches to compare the survival distribution among two groups: non-parametric and parametric. The Weibull family is commonly used in the statistical analysis of lifetime or response time data from reliability experiments and survival studies. To compare two Weibull distributions likelihood ratio test can be used for small samples. This test is based on the identification of the likelihood function. In parametric problems, the likelihood is usually a well-defined quantity. Thoman and Bain (1969) proposed a test to compare shape parameters in two Weibull distributions with the scale parameters unknown, along with a procedure which tests the equality of scale parameters.

The main objectives of this paper are to compare the DN onset time of (i) male and female groups, and (ii) the groups whose age at diabetes diagnosis is less than or equal to 45 years and more than 45 years. In this paper we consider data sets representing the survival time until the occurrence of an event of interest which is diabetic nephropathy. Two methods, i.e. parametric and nonparametric methods are used to compare the survival times of two groups, in the above two cases. Under parametric method, we have compared two left truncated Weibull distributions by applying likelihood ratio test and the two sample test proposed by Thoman and Bain (1969), on the generated left truncated Weibull distributions. Nonparametric methods are applied with left truncated and right censored data. The survival functions are first estimated by applying Kaplan-Meier (KM) method for the groups in both the cases. Then, weighted KM method with appropriate weights is used to test the equality of survival functions, by comparing the difference in the survival functions over time. We have also modified Brookmeyer and Crowley (1982b) method, which is based on the median survival time, to test the equality of median survival of two groups in both the cases. The remainder of the paper is organized as follows. In section 2 development of the models is discussed. Section 3 applies the models to a data set of type-2 diabetic patients (diagnosed of diabetes as per ADA standards) from the data base of Dr. Lal's Path Lab, Delhi, India. Although some work has been

done on the estimation of survival times of diabetic patients, to the best of our knowledge there is no study that has systematically compared the onset DN times of two groups with type-2 diabetes under both the methods: parametric and nonparametric . Some concluding remarks are made in section 4.

2. Development of the model

Observations are made on m mutually independent type-2 diabetic patients. These m patients, on certain criterion, are divided into two groups of sizes m_1 and m_2, so that $m = m_1 + m_2$. Survival time, i.e. DN onset time is not known for all m individuals. It is known for n_1 from m_1 and n_2 from m_2 patients. Thus, survival time is known for $n = n_1 + n_2$ patients and $t_1, t_2, ...t_n$ denote the DN onset time for the combined data. The data at time t (end of study) from two groups can be summarized in a 2 X 2 table as given in Table 1.

Table 1. Distribution of diabetic patients among different groups

Group	Uncensored	censored	Total
Group 1	n_1	$m_1 - n_1$	m_1
Group 2	n_2	$m_2 - n_2$	m_2
Total	n	$m - n$	m

2.1. Likelihood ratio test to compare two left truncated Weibull distributions

Let $t_1, t_2, ...t_{m_1}$ and $t_1, t_2, ...t_{m_2}$ be the times of patients from two groups of sizes m_1 and m_2. The disease time from two groups is assumed to follow Weibull distribution characterized by two parameters, shape (γ) and scale (λ), which are unknown. All the patients included in this study are patients with diabetic history of more than 5 years. The probability density function and survival function for the i^{th} patient belonging to the j^{th} group ($j = 1, 2$) are given as follows:

$$f_j(t_i) = \frac{\lambda_j \gamma_j \ (\lambda_j t_i)^{\gamma_j - 1} \ \exp(-(\lambda_j t_i)^{\gamma_j})}{\exp(-(5\lambda_j)^{\gamma_j})} \qquad t_i \geq 5, \ \ i = 1, 2....., m_j \tag{1}$$

$$S_j(T_i) = \frac{\exp(-(\lambda_j T_i)^{\gamma_j})}{\exp(-(5\lambda_j)^{\gamma_j})} \qquad \lambda_j > 0, \gamma_j > 0 \tag{2}$$

$$L_j = \prod_1^{m_j} (f(t_{ij}))^{\delta_{ij}} (S(T_{ij}))^{1-\delta_{ij}}$$

where δ_{ij} is zero if the i^{th} patient of the j^{th} group is censored and δ_{ij} is unity when the i^{th} patient of the j^{th} group is uncensored, i.e. DN onset time is known. $\mathrm{L_j}(\lambda_j, \gamma_j)$, $j = 1, 2$, denoting the log likelihood functions of the two groups, are given by

$$L_j(\lambda_j, \gamma_j) = \sum_{1}^{m_j} \delta_{ij}((\log \gamma_j + \gamma_j \log \lambda_j - (5\lambda_j)^{\gamma_j}) + (\gamma_j - 1)\log t_{ij} - (\lambda_j t_{ij})^{\gamma_j}) - \sum_{1}^{m_j} (1-\delta_j)((5\lambda_j)^{\gamma_j} + (\lambda_j T_{ij})^{\gamma_j}) \quad (3)$$

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

$$\lambda_j^{\gamma_j}\left(\sum_{1}^{m_j} \delta_{ij} t_{ij}^{\gamma_j} + \sum_{1}^{m_j} (1-\delta_{ij}) T_{ij}^{\gamma_j} - m_j 5^{\gamma_j}\right) - \sum_{1}^{m_j} \delta_{ij} = 0 \quad (4)$$

$$\lambda_j^{\gamma_j}\left(\sum_{1}^{m_j} \delta_{ij} t_{ij}^{\gamma_j} \log(\lambda_j t_{ij}) + \sum_{1}^{m_j} (1-\delta_{ij}) \log(\lambda_j T_{ij}) T_{ij}^{\gamma_j} - n 5^{\gamma_j} \log(5\lambda_j)\right) - \sum_{i=1}^{m_j} \delta_{ij}\left(\frac{1}{\gamma_j} + \log(\lambda_j t_{ij})\right) = 0 \quad (5)$$

The maximum likelihood estimates of λ_j and γ_j are obtained as $\tilde{\lambda}_j$ and $\tilde{\gamma}_j$ using error and trial method.

To compare the two left truncated Weibull distributions using likelihood ratio test (Lee, 2003), the null hypothesis is given by

$$H_0 : \lambda_1 = \lambda_2 = \lambda \quad \text{and} \quad H_0 : \gamma_1 = \gamma_2 = \gamma \quad (6)$$

where λ and γ are unknown. To test the above null hypothesis, we compute the statistic

$$X_L = 2(\mathrm{L_1}(\hat{\lambda}_1, \hat{\gamma}_1) + \mathrm{L_2}(\hat{\lambda}_2, \hat{\gamma}_2) - L(\tilde{\lambda}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\gamma})) \quad (7)$$

$L(\tilde{\lambda}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\gamma})$, the log likelihood value of the combined group with $\tilde{\lambda}$ and $\tilde{\gamma}$ as maximum likelihood estimators, is defined as:

$$L(\tilde{\lambda}, \tilde{\gamma}, \tilde{\lambda}, \tilde{\gamma}) = \sum_{1}^{uncen} \delta_{ij}((\log \tilde{\gamma} + \tilde{\gamma} \log \tilde{\lambda} - (5\tilde{\lambda})^{\tilde{\gamma}}) + (\tilde{\gamma} - 1)\log t_{ij} - (\tilde{\lambda} t_{ij})^{\tilde{\gamma}}) - \sum_{1}^{censored} (1-\delta_j)((5\tilde{\lambda}_j)^{\tilde{\gamma}_j} + (\tilde{\lambda}_j T_{ij})^{\tilde{\gamma}_j}) \quad (8)$$

This likelihood ratio test statistic defined in equation (7), first propounded by Fisher (1922), gives (twice) the log likelihood of the ratio of one hypothesis vs. the

other. It is used to compare two left truncated distributions with fixed truncation time. We reject H_0 if $X_L > \chi^2_{2,\alpha}$, or equivalently, $P(\chi^2_2 > X_L) < \alpha$.

2.2. Likelihood ratio test to compare two left truncated Weibull distributions for uncensored cases

Let $t_1, t_2, ...t_{n_1}$ and $t_1, t_2, ...t_{n_2}$ be the observed exact DN onset times of patients from two groups of sizes n_1 and n_2. The DN onset time from two groups is assumed to follow Weibull distribution characterized by two parameters, shape (γ) and scale (λ), which are unknown. The probability density function for the i^{th} patient belonging to the j^{th} group ($j = 1, 2$) is given above in equation 1 under section 2.1.

$L_j(\lambda_j, \gamma_j)$, $j = 1, 2$, denoting the log likelihood functions of the observed survival time from two groups, are given by

$$L_j(\lambda_j, \gamma_j) = n_j(\log\gamma_j + \gamma_j \log\lambda_j - (5\lambda_j)^{\gamma_j}) + (\gamma_j - 1)\sum_{i=1}^{n_j} \log t_i - \sum_{i=1}^{n_j} (\lambda_j t_i)^{\gamma_j};$$
$$j = 1, 2 \quad (9)$$

The maximum likelihood estimates of λ_j and γ_j are obtained as $\hat{\lambda}_j$ $\hat{\gamma}_j$, respectively.

To compare the two left truncated Weibull distributions using likelihood ratio test

$$H_0 : \lambda_1 = \lambda_2 = \lambda \quad \text{and} \quad H_0 : \gamma_1 = \gamma_2 = \gamma$$

where λ and γ are unknown. To test the above null hypothesis, we compute the statistic

$$X_L = 2(L_1(\hat{\lambda}_1, \hat{\gamma}_1) + L_2(\hat{\lambda}_2, \hat{\gamma}_2) - L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma}))$$

$L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma})$, the log likelihood value of the combined group of uncensored cases, is defined as:

$$L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma}) = (n_1 + n_2)(\log\tilde{\gamma} + \tilde{\gamma}\log\hat{\lambda} - (5\hat{\lambda})^{\tilde{\gamma}}) + (\tilde{\gamma} - 1)(\sum_{i=1}^{n_1} \log t_i + \sum_{i=1}^{n_2} \log t_i) - (\sum_{i=1}^{n_1} (\hat{\lambda} t_i)^{\tilde{\gamma}} + \sum_{i=1}^{n_2} (\hat{\lambda} t_i)^{\tilde{\gamma}}) \quad (10)$$

This test is sufficient if the data concludes that the two distributions are significantly different but if the data concludes otherwise, an additional two-sample test proposed by Thoman and Bain (1969) for uncensored samples for comparing two Weibull distributions has been applied for left truncated distributions.

2.2. Thoman and Bain two-sample test for comparing two left truncated Weibull distributions

This test assumes that independent samples of equal sizes are obtained from left truncated Weibull distributions as given in 2.1. To compare the two distributions the equality of shape parameter is tested. The null and alternative hypothesis is given as:

$$H_0 : \gamma_1 = \gamma_2 \quad \text{against} \qquad H_1 : \gamma_1 > \gamma_2 \tag{11}$$

To test the above null hypothesis, the following statistic is computed

$$R = \frac{\hat{\gamma}_1 / \gamma_1}{\hat{\gamma}_2 / \gamma_2} \quad \text{and under } H_0, \quad R = \frac{\hat{\gamma}_1}{\hat{\gamma}_2} \quad \text{(Thoman and Bain, 1969)}$$

We reject H_0 if $R > \ell_\alpha$ (using Thoman and Bain, 1969) and conclude that the two Weibull distributions are significantly different. However, if the hypothesis $H_0 : \gamma_1 = \gamma_2$ is not rejected we test the equality of scale parameters.

$$H_0 : \lambda_1 = \lambda_2 \text{ against } H_1 : \lambda_1 < \lambda_2 \tag{12}$$

To test the above null hypothesis we compute the statistic given as follows:

$$G = 0.5(\hat{\gamma}_1 + \hat{\gamma}_2)(\log \hat{\lambda}_1 - \log \hat{\lambda}_2) \quad \text{(Thoman and Bain, 1969)} \tag{13}$$

H_0 is rejected if $G > z_\alpha$, (using Thoman and Bain, 1969) where z_α is $P(G < z_\alpha \mid H_0) = 1 - \alpha$ and $\hat{\gamma}_1$, $\hat{\gamma}_2$, $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the maximum likelihood estimators of γ_1, γ_2, λ_1 and λ_2 respectively.

2.3 Weighted Kaplan-Meier method to compare survival distributions of two groups

We consider the classical two-sample censored data with fixed left truncation survival analysis problem, with survival continuous and censoring independent of survival in each group. To compare the survival distributions of two groups of sizes m_1 and m_2, as defined in section 2, let $t_1 < t_2 < \ldots\ldots\ldots < t_n$ denote the ordered survival time of two groups taken together. Let O_{ij}, C_{ij}, Y_{ij} be, respectively, the number of events, the number of censored observations and the number at risk at time t, in the j^{th} group, $j = 1, 2$. Let $\hat{S}_j(t)$ be the Kaplan-Meier (KM) estimator (Klien and Moeschberger, 2003; Pepe and Fleming, 1989) of the event distribution using data in the j^{th} group and $\hat{H}_j(t)$ be the KM estimator of the time to censoring in the j^{th} group, that is $\hat{H}_j(t) = \prod_{t_i \le t} [1 - C_{ij} / Y_{ij}]$, and $\hat{S}_p(t)$ be the KM estimator based on the combined group.

To test the equality of the two survival distributions the hypothesis is defined as

$$H_0 : S_1(t) = S_2(t) \quad \text{against} \quad H_1 : S_1(t) > S_2(t) \tag{14}$$

Or $H_0 : S_1(t) = S_2(t)$ against $H_1 : S_1(t) \neq S_2(t)$

The test statistic is given by

$$Z = W_{KM} / \hat{\sigma}_p$$

where

$$W_{KM} = \sqrt{\frac{m_1 m_2}{m}} \sum_{i=1}^{D-1} [t_{i+1} - t_i] w(t_i) [\hat{S}_1(t_i) - \hat{S}_2(t_i)] \tag{15}$$

KM suggested a weight function, w(t), in the study period t_D defined as

$$w(t) = \frac{m \hat{H}_1(t) \hat{H}_2(t)}{m_1 \hat{H}_1(t) + m_2 \hat{H}_2(t)} \ , \quad 5 \leq t \leq t_D$$

The variance ($\hat{\sigma}_p$) is given by

$$\hat{\sigma}_p^2 = -\sum_{i=1}^{D-1} [\hat{S}_p(t_i) - \hat{S}_p(t_{i-1})] \frac{T^2{}_i}{\hat{S}_p(t_i) \hat{S}_p(t_{i-1})} \frac{m_1 \hat{H}_1(t_{i-1}) + m_2 \hat{H}_2(t_{i-1})}{m \hat{H}_1(t_{i-1}) \hat{H}_2(t_{i-1})} \tag{16}$$

where

$$T_i = \sum_{k=i}^{D-1} [t_{k+1} - t_k] w(t_k) \hat{S}_p(t_k)$$

The sum in (12) has only nonzero contributions as t_i is the onset time and for censored observations $\hat{S}_p(t_{i-1}) - \hat{S}_p(t_i)$ =0.

The test statistic Z has an approximate standard normal distribution. Suppose the alternative hypothesis, $S_1(t) > S_2(t)$, reject H_0 if Z comes out to be greater than Z_α ($Z > Z_\alpha$) where Z_α is given by $P(Z > Z_\alpha \mid H_0) = \alpha$. If the alternative hypothesis, $S_1(t) \neq S_2(t)$, reject H_0 if Z comes out to be greater than $Z_{\alpha/2}$.

We suggest a new weight function defined as:

$$w_{AG}(t) = \frac{(n/m) \hat{H}_1(t) \hat{H}_2(t)}{(n_1/m_1) \hat{H}_1(t) + (n_2/m_2) \hat{H}_2(t)} \ ; \ 5 \leq t \leq t_D \tag{17}$$

to test the equality of the two survival distributions. In this case the test statistic is computed as follows:

$$Z = W_{AG} / \hat{\sigma}_{AG}$$

where W_{AG} and $\hat{\sigma}_{AG}$ can be computed like W_{KM} and $\hat{\sigma}_p$ by replacing m, m_1 and m_2 by $n/m, n_1/m_1$ and n_2/m_2, respectively.

2.4. Modified Brookmeyer and Crowley method to compare survival distributions of two groups

Brookmeyer and Crowley (1982b) have suggested an alternate method (Klien and Moeschberger, 2003) to test the equality of survival times of two groups. This method is based on median survival time rather than comparing the difference in the survival functions over time. Let $t_1, t_2 ....., t_m$ denote the diabetic duration of the two groups taken together, where O_{ij}, Y_{ij}, $\hat{S}_j(t)$ ($j = 1, 2$) and $\hat{S}_p(t)$ are the same as defined in the procedure given in section 2.3.

To test the equality of the two median survival times the null hypothesis is given as follows:

$$H_0 : M_d(S_1(t)) = M_d(S_2(t)) \text{ against } H_1 : M_d(S_1(t)) \neq M_d(S_2(t)) \tag{18}$$

and the test statistic is defined as follows:

$$U = \frac{m(S_1(\hat{M}) - 0.5)}{\sigma^2}$$

Brookmeyer and Crowley derived a method where median survival time is based on the common survival function, $S_w(t) = \frac{m_1 \hat{S}_1(t) + m_2 \hat{S}_2(t)}{m}$.

We modified this survival function, which meets the situation when the number of events and their onset and censoring times in $\hat{S}_1(t)$ and $\hat{S}_2(t)$ are different. In our case median survival time is based on common pooled survival function. To compute $S_1(\hat{M})$, σ^2 and $\hat{M}$, the algorithm used (Klien and Moeschberger, 2003) is given as follows:

- Arrange the onset time as $t_1 < t_2 < ....., t_n$ in a pooled sample. If for some t_i, $\hat{S}_p(t_i) =$ 0.5, then $\hat{M} = t_i$.
- If no event time gives a value of $\hat{S}_p$ equal to 0.5 then let M_L be the largest event time with $\hat{S}_p(M_L)$ > 0. 5 and let M_U be the smallest event with $\hat{S}_p(M_U)$ < 0.5. Then, the median lies in the interval (M_L, M_U) and is obtained by linear interpolation, i.e.

$$\hat{M} = M_L + \frac{(0.5 - \hat{S}_p(M_L))(M_U - M_L)}{(\hat{S}_p(M_U) - \hat{S}_p(M_L))} \tag{19}$$

- Obtain the death time in the j^{th} sample such that $T_{Lj} \leq \hat{M} < T_{Uj}$. The estimated probability of survival beyond $\hat{M}$ in the j^{th} sample is obtained by linear interpolation as

$$\hat{S}_j(\hat{M}) = \hat{S}_j(T_{Lj}) + \frac{(\hat{S}_j(T_{Uj}) - \hat{S}_j(T_{Lj}))(\hat{M} - T_{Lj})}{(T_{Uj} - T_{Lj})}, \quad j = 1, 2 \tag{20}$$

- For obtaining σ^2, let t_{ij} denote the distinct death time in the j^{th} sample, d_{ij} the number of deaths at time t_{ij} and Y_{ij} the number at risk at time t_{ij}. Then, σ^2 is given as follows:

$$\sigma^2 = \frac{m_2^{\,2}}{m}(V_1 + V_2) \tag{21}$$

$$V_j = [\hat{S}_j(T_{Uj})(\frac{\hat{M} - T_{Lj}}{T_{Uj} - T_{Lj}})]^2 \sum_{t_{ij} \leq T_{Uj}} \frac{O_{ij}}{Y_{ij}(Y_{ij} - O_{ij})} \; +$$

$$+\{[\hat{S}_j(T_{Lj})(\frac{T_{Uj} - \hat{M}}{T_{Uj} - T_{Lj}})]^2 + 2(\frac{(\hat{M} - T_{Lj})(T_{Uj} - \hat{M})}{(T_{Uj} - T_{Lj})^2})\hat{S}_j(T_{Uj})\hat{S}_j(T_{Lj})]\} \sum_{t_{ij} \leq T_{Lj}} \frac{O_{ij}}{Y_{ij}(Y_{ij} - O_{ij})}$$

The test statistic U follows chi-square distribution with one degree of freedom. Reject H_0 if U comes out to be greater than $Z_{\alpha/2}$, where Z_α is given by $P(U > Z_\alpha \mid H_0) = \alpha$.

3. Application

The methods discussed in section 2 are applied to the data obtained through a house to house survey of diabetic patients who were referred for pathological tests to Dr. Lal path lab, Delhi, India. A retrospective study is conducted on the collected data. Since our study is focused on diabetic nephropathy only, the patient's data indicating effect on eyes, heart, etc. are excluded. Thus, a total of 132 patients were selected who were diagnosed as diabetics as per ADA standards with minimum 5 years duration. Out of these 132 patients, 60 were uncensored with diabetic nephropathy and 72 were censored/non-diabetic nephropathy, all aged between 44.45 ± 4.79 years (mean ± SD). The patients in the diabetic nephropathy group were distributed according to gender and age at the time of diagnosis as displayed in Table 3. The demographic and risk variables

recorded were: age at the time of diagnosis, duration of disease, fasting blood glucose (FBG), diastolic blood pressure (DBP), systolic blood pressure (SBP), low-density lipoprotein (LDL) and serum creatinine (SrCr) as given in Table 2.

Table 2. Mean± SD of demographic and risk variables of 132 patients of two groups; non-diabetic nephropathy(NDN) and diabetic nephropathy(DN)

Variable	NDN=Censored	DN=Uncensored
Group size	72	60
Age at diabetes diagnosis (years)	44.01±4.36	45.00± 5.28
Duration of diabetes (years)	10.28±5.70	14.10±5.05
FBG (mg/dl)	133.80±17.48	142.04±14.39
DBP (mmHg)	82.39±6.08	91.97±9.42
SBP (mmHg)	125.12±12.40	142.82±13.88
LDL (mg/dl)	91.80±18.75	107.44±14.27
SrCr (mg/dl)	1.00±0.151	1.67±0.2833

* NDN denotes non-diabetic nephropathy and DN denotes diabetic nephropathy

Table 3. Distribution of patients among different groups based on gender and age at diabetes diagnosis.

Group	Uncensored	Censored	Total group size
Male	28	31	59
Female	32	41	73
Diagnosis≤45years	33	48	81
Diagnosis>45years	27	24	51
Total	60	72	132

3.1. Comparison of patients on the basis of gender and age at the time of diagnosis of diabetes by applying likelihood test

The motivation of our model development is to compare the time of type-2 diabetic patients on the basis of gender and age at the time of diagnosis of diabetes. Out of 132 cases, there were 59 males and 73 females. Available diabetic data has been used to test the equality of male and female disease time, which is assumed to follow Weibull distribution with fixed left truncation time as 5 years. To find the likelihood ratio test statistic, X_L , and to test the hypothesis (5), we have found the maximum likelihood estimators of unknown parameters for male, female and combined groups and their respective log-likelihood values as given in Table 4. The observed value of the likelihood ratio test statistic, X_L ,

came out to be 34.9082. Since test statistic X_L is found to be greater than $\chi^2_{2,\alpha}$, p < 0.001, we reject H_0 and conclude that the two Weibull distributions are significantly different.

Secondly, we have classified the time on the basis of age at diabetes diagnosis. Out of 132 cases with diabetic nephropathy, the age at diabetes diagnosis of 81 patients was less than or equal to 45 years and of 51 patients was greater than 45 years. Available data has been used to test the equality of the two groups, which are based on age at diabetes diagnosis and are assumed to follow Weibull distribution with fixed left truncation time as 5 years. To test the hypothesis (5), the above procedure is repeated to compute the log-likelihood ratio test statistic and its value came out to be X_L = 16.2872. Again, since test statistic X_L is greater than $\chi^2_{2,\alpha}$, p <0.001, we reject H_0 and conclude that the DN onset times of two groups are significantly different or the two Weibull distributions are significantly different. The log-likelihood values and maximum likelihood estimators of the parameters are given in Table 4.

Table 4. Maximum Likelihood estimators (MLE) of λ and γ along with log likelihood value of two groups when the grouping variable is (i) gender (ii) age at diabetes diagnosis

(i) Group Size	**Male** m_1 **=59**	**Female** m_2 **=73**	**Total** $m_1 + m_2$ **=132**
MLE of shape parameter	$\hat{\gamma}_1$ =3.4115	$\hat{\gamma}_2$ =3.61382	$\tilde{\gamma}$ =4.04138
MLE of scale parameter	$\hat{\lambda}_1$ =0.0661	$\hat{\lambda}_2$ =0.0718	$\hat{\lambda}$ =0.080238
log likelihood value	$L_1(\hat{\lambda}_1, \hat{\gamma}_1)$ =-46.1095	$L_2(\hat{\lambda}_2, \hat{\gamma}_2)$ =-60.1714	$L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma})$ =-123.735
(ii) Group Size	**Diagnosis ≤45yrs** m_1 **=81**	**Diagnosis >45yrs** m_2 **=51**	**Total** $m_1 + m_2$ **=132**
MLE of shape parameter	$\hat{\gamma}_1$ = 3.4967	$\hat{\gamma}_2$ =3.8199	$\tilde{\gamma}$ =4.04138
MLE of scale parameter	$\hat{\lambda}_1$ =0.0616	$\hat{\lambda}_2$ =0.0.0527	$\hat{\lambda}$ =0.080238
log likelihood value	$L_1(\hat{\lambda}_1, \hat{\gamma}_1)$ =-51.2137	$L_2(\hat{\lambda}_2, \hat{\gamma}_2)$ =-64.3777	$L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma})$ =-123.735

3.2. Comparison of DN onset times of patients on the basis of gender and age at the time of diagnosis of diabetes by applying likelihood test

Further, we have classified the survival time, i.e. DN onset time, on the basis of gender. Out of 60 uncensored cases with diabetic nephropathy, there were 28 males and 32 females. Available diabetic data has been used to test the equality of male and female DN onset time, which is assumed to follow Weibull distribution with fixed left truncation time as 5 years. To find the likelihood ratio test statistic, X_L, and to test the hypothesis (5), we have found the maximum likelihood estimators of unknown parameters for male, female and combined groups and their respective log-likelihood values as given in Table 5. The observed value of the likelihood ratio test statistic, X_L, came out to be 85.056. Since test statistic X_L, is found to be greater than $\chi^2_{2,\alpha}$, $p < 0.001$, we reject H_0 and conclude that the two Weibull distributions are significantly different. It has been shown graphically that the DN onset times of two groups are significantly different, as displayed in Figure 1.

DN onset time is classified on the basis of age at diabetes diagnosis. Out of 60 uncensored cases with diabetic nephropathy, the age at diabetes diagnosis of 33 patients was less than or equal to 45 years and the age at diabetes diagnosis of 27 patients was greater than 45 years. Available data has been used to test the equality of the two groups, which are based on age at diabetes diagnosis and are assumed to follow Weibull distribution with fixed left truncation time as 5 years. To test the hypothesis (5), the above procedure is repeated to compute the log-likelihood ratio test statistic and its value came out to be $X_L = 14.538$. Again, since test statistic X_L is greater than $\chi^2_{2,\alpha}$, $p < 0.001$, we reject H_0 and conclude that the DN onset times of two groups are significantly different or the two Weibull distributions are significantly different. The log-likelihood values and maximum likelihood estimators of the parameters are given in Table 5. It has been also shown graphically that the DN onset times of the two groups are significantly different, as displayed in Figure 2.

Thus, log-likelihood ratio test is found to be sufficient, as in all the above cases the data conclude that the DN onset times of two groups are significantly different.

Table 5. Maximum Likelihood estimators (MLE) of λ and γ along with log likelihood value for two groups when the grouping variable is (i) gender (ii) age at diabetes diagnosis

(i) Group Size	**Male** n_1 **=28**	**Female** n_2 **=32**	**Total uncensored** $n_1 + n_2$ **=60**
MLE of shape parameter	$\hat{\gamma}_1$ = 3.113	$\hat{\gamma}_2$ = 1.561	$\tilde{\gamma}$ =2.811
MLE of scale parameter	$\hat{\lambda}_1$ =0.075	$\hat{\lambda}_2$ =0.062	$\hat{\lambda}$ =0.080
log likelihood value	$L_1(\hat{\lambda}_1, \hat{\gamma}_1)$ =-81.761	$L_2(\hat{\lambda}_2, \hat{\gamma}_2)$ =-60.993	$L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma})$ =-185.283
(ii) Group Size	**Diagnosis ≤45yrs** n_1 **=33**	**Diagnosis >45yrs** n_2 **=27**	**Total uncensored** $n_1 + n_2$ **=60**
MLE of shape parameter	$\hat{\gamma}_1$ = 3.221	$\hat{\gamma}_2$ = 3.014	$\tilde{\gamma}$ =2.811
MLE of scale parameter	$\hat{\lambda}_1$ =0.077	$\hat{\lambda}_2$ =0.072	$\hat{\lambda}$ =0.080
log likelihood value	$L_1(\hat{\lambda}_1, \hat{\gamma}_1)$ =-94.142	$L_2(\hat{\lambda}_2, \hat{\gamma}_2)$ =-83.871	$L(\hat{\lambda}, \tilde{\gamma}, \hat{\lambda}, \tilde{\gamma})$ =-185.283

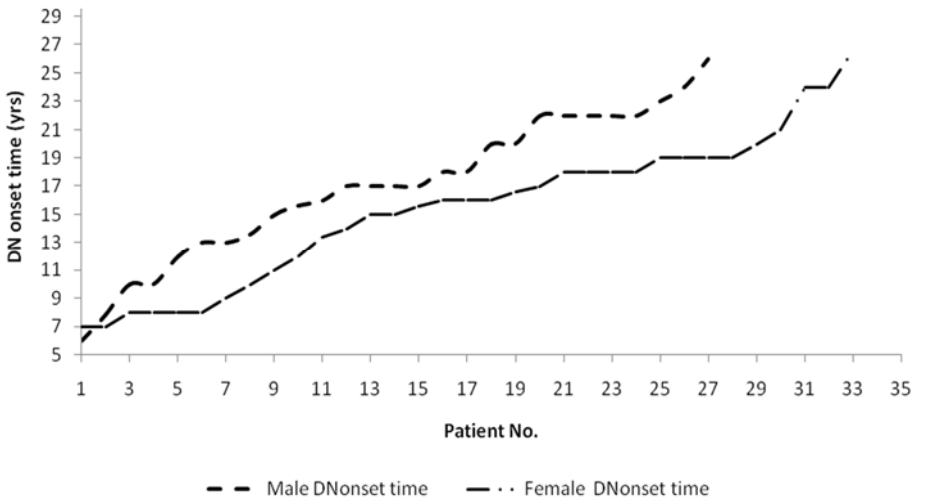

Figure 1. Onset time of diabetic nephropathy, uncensored cases; group1 =28 male patients; group2 = 32 female patients

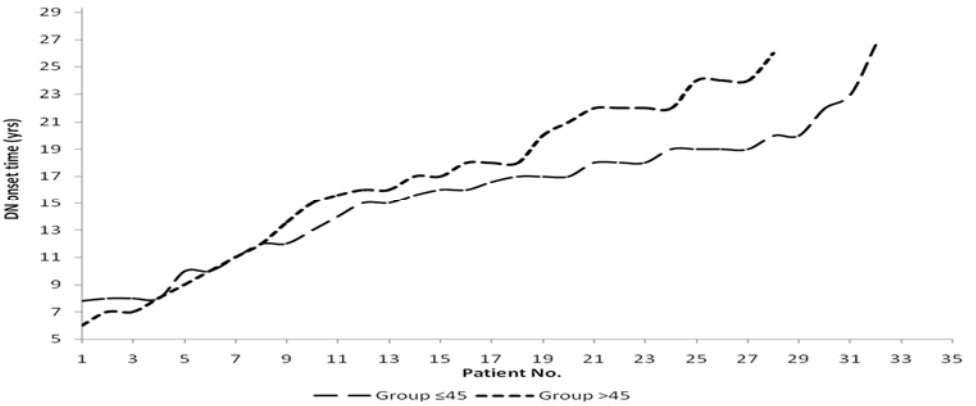

Figure 2. Onset time of diabetic nephropathy, uncensored cases; group1 =33, age at diagnosis less than or equal to 45years; group2 = 27, age at diagnosis greater than 45 years

3.2. Left truncated Weibull distributions are generated for the application of Thoman and Bain two sample test

In this section we have shown the application of two-sample test proposed by Thoman and Bain (1969), for cases when the likelihood ratio test is not sufficient. To compare the survival distributions of two groups of patients of same sizes with fixed left truncation, we generated three left truncated Weibull distributions with maximum likelihood estimators (MLE) of λ, γ parameters. The group1 is generated with MLE of λ, γ as $(\tilde{\lambda}, \tilde{\gamma})$ = (0.077, 3.221) for sizes 10, 15, 20, 30 and 40. The group2 is generated with MLE of λ, γ as $(\tilde{\lambda}, \tilde{\gamma})$ = (0.073, 3.0144) for same sizes. The combined group is generated with MLE of λ, γ as $(\tilde{\lambda}, \tilde{\gamma})$ = (0.080, 2.811) for the combined sample sizes as 20, 30, 40, 60 and 80. The log likelihood is computed for each size corresponding to each group, which is given in column 2, 4 and 6, respectively, in Table 6. The likelihood ratio test statistic has been obtained and simultaneously the significant value for each case has been computed as given in the last column in table 6. The cases where p < 0.05, for group sizes 30 and 40 likelihood ratio test has been found to be sufficient, but if p > 0.05, for group sizes 10, 15 and 20 we fail to reject the null hypothesis and the likelihood ratio test is not sufficient as likelihood ratio test is more appropriate for asymptotic cases.

Thoman and Bain two-sample test has been applied for cases with sample sizes 10, 15 and 20, respectively, where likelihood ratio test is not sufficient. According to this test, first the equality of the shape parameter is tested by finding $R = \frac{\tilde{\gamma}_1}{\tilde{\gamma}_2}$, under H_0. It has been found that in all the three sample sizes 10, 15 and

20, the test statistic R is less than ℓ_α at α =.05. Thus, we fail to reject the equality of shape parameters (Tables in Thoman and Bain, 1969). The respective values of R for the three sample sizes are given in Table 7. After accepting the equality of shape parameter, the equality of scale parameters is tested by finding the statistic G, for all the three sample sizes. Since the value of the test statistic G is found to be less than z_α at α =.05, we fail to reject H_0 (Tables in Thoman and Bain, 1969). Respective comparisons are given in Table 7. Thus, from the generated study we conclude that for sample sizes 10, 15 and 20, there is no significant difference between the two left truncated Weibull distributions, whereas for samples sizes 30 and 40, there is a significant difference between the two generated left truncated Weibull distributions.

Table 6. Generated left truncated Weibull distribution with log-likelihood values and test statistic values of samples with different sizes

Sample Size (group-1)	Generated Weibull $\hat{\lambda}_1$ =0.077 $\hat{\gamma}_1$ =3.221	Sample Size (group-2)	Generated Weibull $\hat{\lambda}_2$ =0.072 $\hat{\gamma}_2$ = 3.014	Sample Size (combined group)	Generated Weibull $\tilde{\lambda}$ =0.080 $\tilde{\gamma}$ =2.811	X_L =2($L_1(\hat{\lambda}_1,\hat{\gamma}_1)$ + $L_2(\hat{\lambda}_2,\hat{\gamma}_2)$ + $L(\tilde{\lambda},\tilde{\gamma},\hat{\lambda},\tilde{\gamma})$)	Significant Value
	Log Likelihood		Log Likelihood		Log Likelihood		
10	-18.566	10	-27.945	20	-46.913	0.806	0.857*
15	-28.905	15	-25.655	30	-55.771	2.423	0.369*
20	-36.120	20	-41.717	40	-78.002	0.332	0.932*
30	-56.411	30	-57.292	60	-120.468	13.532	0.002
40	-78.775	40	-125.537	80	-214.058	19.492	<0.001

* In these cases Thoman and Bain test procedure has been used for comparing the two survival distributions.

Table 7. Thoman and Bain test procedure for comparing two survival distributions

Test	Statistic	Sample size	Inference
$H_0 : \gamma_1 = \gamma_2$ $H_1 : \gamma_1 > \gamma_2$	R^*=1.069	$n_1 = n_2$ =10	R <1.893, Fail to reject H_0
		$n_1 = n_2$ =15	R <1.688, Fail to reject H_0
		$n_1 = n_2$ =20	R <1.553, Fail to reject H_0
$H_0 : \lambda_1 = \lambda_2$ $H_1 : \lambda_1 > \lambda_2$	G^*=0.088	$n_1 = n_2$ =10	G <0.992, Fail to reject H_0
		$n_1 = n_2$ =15	G <0.704, Fail to reject H_0
		$n_1 = n_2$ =20	G <0.593, Fail to reject H_0

$^* R = \frac{\hat{\gamma}_1}{\hat{\gamma}_2}$, $\quad G = 0.5(\hat{\gamma}_2 + \hat{\gamma}_1)(Log(\hat{\lambda}_2) - Log(\hat{\lambda}_1))$

3.3. Weighted Kaplan-Meier method to compare DN onset time of two groups of patients, including censored cases

The KM method is useful to compare the survival function (with different weights) of two groups over different time. In this case, firstly, the 132 type-2 diabetic patients have been divided into two groups (on the basis of gender) of 59 males and 73 females. The duration of diabetes of 132 individuals is mutually independent. The DN onset times are known for 28 males and 32 females. The DN onset times of these 60 patients have been arranged as $t_1 < t_2 < ....., t_n$ in a pooled sample. The survival function has been estimated for both the groups and also for the pooled sample, using KM method. This survival function is redistributed according to the time interval as given in Table 8. To test the equality of the two survival distributions, we have found the weight function, as suggested by KM, at each t_i. The values of the weight function W_{KM}, estimate of the standard deviation $\hat{\sigma}_p$ and the test statistic Z came out to be 1.347, 0.204 and 6.599, respectively, as given in Table 8. The null hypothesis has been rejected since the p-value came out to be less than 0.001 ($p < 0.001$). Further, with the same objective but using a different weight function we have obtained the values of W_{AG} and $\hat{\sigma}_{AG}$, by replacing m, m_1 and m_2 by $n/m, n_1/m_1$ and n_2/m_2, respectively. The values of W_{AG}, $\hat{\sigma}_{AG}$ and test statistic Z, came out to be 0.998, 0.091 and 10.932, respectively. Again, the null hypothesis has been rejected since p-value came out to be less than 0.001 ($p < 0.001$). Thus, we conclude that the two survival functions are significantly different. We have also compared graphically the survival function (onset DN times) of male and female groups, as shown in Figure 3.

Now, secondly, the 132 diabetic patients have been divided into two groups (on the basis of age at diabetes diagnosis) of 81 patients whose age at diabetes diagnosis was less than or equal to 45 years and of 51 patients whose age at diabetes diagnosis was greater than 45 years. The DN onset times are known for 33 patients whose age at diabetes diagnosis was less than or equal to 45 years and for 27 patients whose age at with diabetes diagnosis was greater than 45 years. The DN onset times of these 60 patients have been arranged as $t_1 < t_2 < ....., t_n$ in a pooled sample. The survival function has been estimated for both the groups. This survival function is redistributed according to the time interval as given in Table 8. To test the equality of the two survival distributions, we have used the same procedure as given above, with two weight functions. The two test statistic, with different weight functions, came out to be 5.252 and 6.057, respectively, as given in Table 9. Thus, we reject the null hypothesis in both the cases, since p-value came out to be less than 0.001 ($p < 0.001$) and conclude that the two survival distributions are significantly different. We have also compared graphically the DN onset times of two groups, classified on the basis of age at diabetes diagnosis, as shown in Figure 4.

Table 8. Estimates of survival functions with group variable as gender and age at diabetes diagnosis, using KM method

Duration of diabetes	Survival function		Survival function		Pooled Survival Function
	Group1 =Male	Group2 =female	Group1=diabetes diagnosis age ≤45	Group1=diabetes diagnosis age >45	
6≤ t_i <7	0.983	1	1	0.979	0.992
7≤ t_i <8	0.958	0.966	0.982	0.937	0.963
8≤ t_i <9	0.958	0.885	0.915	0.914	0.918
9≤ t_i <10	0.958	0.863	0.894	0.888	0.906
10≤ t_i <11	0.906	0.841	0.851	0.888	0.870
11≤ t_i <12	0.906	0.794	0.808	0.859	0.845
12≤ t_i <13	0.850	0.794	0.786	0.859	0.819
13≤ t_i <14	0.793	0.768	0.764	0.802	0.778
14≤ t_i <15	0.765	0.743	0.743	0.773	0.751
15≤ t_i <16	0.708	0.666	0.653	0.716	0.684
16≤ t_i <17	0.680	0.559	0.572	0.656	0.616
17≤ t_i <18	0.566	0.536	0.497	0.567	0.533
18≤ t_i <19	0.507	0.512	0.429	0.504	0.459
19≤ t_i <20	0.507	0.442	0.304	0.504	0.398
20≤ t_i <21	0.439	0.421	0.243	0.465	0.352
21≤ t_i <22	0.439	0.301	0.243	0.465	0.352
22≤ t_i <23	0.256	0.271	0.195	0.310	0.248
23≤ t_i <24	0.192	0.271	0.130	0.310	0.124
24≤ t_i <25	0.128	0.090	0.130	0.078	0.083
t_i≥25	0.064	0	0.065	0	0.041

Table 9. Comparison of Diabetic Nephropathy onset time of two groups with different weight functions

Grouping variable	Weight function	Statistic	Sqrt(variance)	Z
Gender	$w(t)=\frac{m\hat{H}_1(t)\hat{H}_2(t)}{m_1\hat{H}_1(t)+m_2\hat{H}_2(t)}$	W_{KM} =1.347	$\hat{\sigma}_p$ =0.204	6.599
	$w_{AG}(t)=\frac{(n/m)\hat{H}_1(t)\hat{H}_2(t)}{(n_1/m_1)\hat{H}_1(t)+(n_2/m_2)\hat{H}_2(t)}$	W_{AG} =0.998	$\hat{\sigma}_{AG}$ =0.091	10.932
Diabetes diagnosis age	$w(t)=\frac{m\hat{H}_1(t)\hat{H}_2(t)}{m_1\hat{H}_1(t)+m_2\hat{H}_2(t)}$	W_{KM} =1.529	$\hat{\sigma}_p$ =0.291	5.252
	$w_{AG}(t)=\frac{(n/m)\hat{H}_1(t)\hat{H}_2(t)}{(n_1/m_1)\hat{H}_1(t)+(n_2/m_2)\hat{H}_2(t)}$	W_{AG} =1.085	$\hat{\sigma}_{AG}$ =0.179	6.057

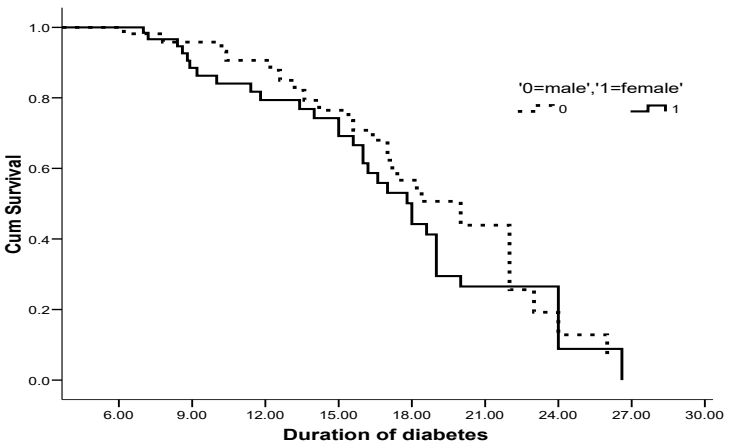

Figure 3 Comparison of DN onset time of male and female type-2 DM patients, using the KM estimator

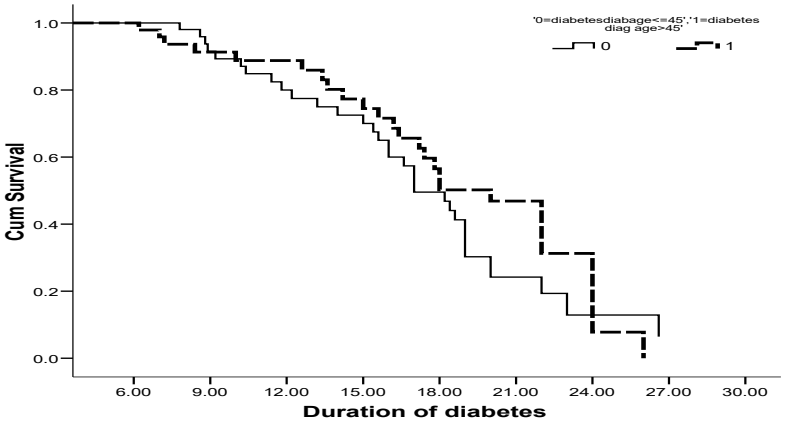

Figure 4 Comparison of DN onset time of diabetic patients with age at diabetes diagnosis ≤45years and diabetes diagnosis age>45years, using the KM estimator

5.3.4. Modified Brookmeyer and Crowley method to compare DN onset time of two groups of patients based on the median survival function of DN onset time

In this part, an alternative method has been used based on median survival time rather than comparing the difference in the survival functions over time. Again, firstly, the 132 type-2 diabetic patients have been divided into two groups of 59 males and 73 females. Using KM method survival function has been estimated at each event point for male, female and pooled sample. We have arranged the onset times in ascending order in a pooled sample and in this case it has been found that there exist no t_i, for which $\hat{S}_p(t_i) = 0.5$. Then, to find the estimated median survival time $\hat{M}$, we have computed M_L, which is the largest event time with $\hat{S}_p(M_L)$ > 0. 5 and M_U be the smallest event with $\hat{S}_p(M_U)$ < 0.5. Then, the median lies in the interval (M_L, M_U) and is obtained by linear interpolation, i.e. M_L and M_U as 18.0 and 18.2, respectively, and the corresponding values of $\hat{S}_p(M_L)$ and $\hat{S}_p(M_U)$ came out to be 0.502 and 0.488, respectively. Thus the median lies in the interval (18, 18.2) and by linear interpolation it came out to be 18.034. Using the survival time for both the samples, we have computed T_{Lj} and T_{Uj}, the death time in the j^{th} sample such that the estimated median survival time lies between these two death times. We

have found that the onset time in the male sample satisfied the condition $T_{Lj} \leq \hat{M} < T_{Uj}$ as the median lies in between 18 and 20 (18<18.034<20). The estimated probability of survival, $\hat{S}_1(\hat{M})$, by linear interpolation, beyond $\hat{M}$ in the male sample came out to be 0.508. To find the variance we have obtained the values of V_1 and V_2 as 0.016 and 0.0002, respectively. Finally, the variance and the test statistic came out to be 0.515 and 0.764, respectively, as given in Table 9. Thus, we fail to reject the null hypothesis as p= 0.447 and conclude that there is no significant difference in the median survival time of DN onset time of male and female groups.

The above test, which is based on median survival time, can also be applied when data is divided into two groups on the basis of age at diabetes diagnosis. Out of 132 individuals there were 81 patients whose age at diabetes diagnosis was less than or equal to 45 years and there were 51 patients whose age at diabetes diagnosis was greater than 45 years. The median survival time came out to be, i.e. 18.034, as it is based on the pooled sample. Using the survival time for both the samples, we have found that the onset time of patients whose age at diabetes diagnosis was greater than 45 years satisfy the condition $T_{Lj} \leq \hat{M} < T_{Uj}$ as 18<18.034<20. The estimated probability of survival, $\hat{S}_1(\hat{M})$, by linear interpolation, beyond $\hat{M}$ for patients whose age at diabetes diagnosis was greater than 45 years, came out to be 0.503. To find the variance, we first obtained the values of V_1 and V_2 as 0.014 and 0.012, respectively. Finally, the variance and the test statistic came out to be 0.827 and 0.475, respectively, as given in Table 10. Thus, we fail to reject the null hypothesis as p=0.496 and conclude that there is no significant difference in the median survival time of DN onset times of age at diabetes diagnosis less than or equal to 45 years and age at diabetes diagnosis greater than 45 years.

Table 10. Comparison of DN onset time between two groups based on median survival time

Grouping variable	Median survival time based on weighted sample	Variance	Test Statistic
Gender	18.034	0.515	0.764
Age at diabetes diagnosis	18.034	0.827	0.475

4. Discussion

The aim of this study is to compare the survival time (onset time of nephropathy from the diagnosis of type-2 diabetes) of two groups of patients with type-2 diabetes. The major advantage of developing methods for comparing the DN onset time of type-2 DM patients is that the DN onset time pattern of new DM patients can be predicted depending on the patient group. Also, it can be used as a baseline for further studies.

Firstly, we have used parametric methods to compare the survival times of two groups, on the basis of gender and age at the time of diagnosis of diabetes, assuming that DN onset times follow Weibull distribution. The two parameters of the Weibull parametric distribution provide additional flexibility that potentially increases the accuracy of the description of collected survival data, since the shape parameter allows the hazard function to increase or decrease with increasing time (Collet, 2003). The likelihood ratio test is applied here on collected data of type-2 DM patients with minimum 5 years of duration of diabetes, to compare groups: male and female and of patients whose age at diabetes diagnosis is less than or equal to 45 years and greater than 45 years. This test is widely used in comparing two survival distributions for the cases where sample sizes are not small and the equality of two Weibull distributions is rejected (Lee, 2003). We have applied two sample test of Thoman and Bain (1969) to compare small samples of equal sizes, and cases where we fail to reject the equality of shape and scale parameter. Thus, in some cases limitation of likelihood ratio test can be handled with Thoman and Bain test. If the data comes from a Weibull distribution, the most accurate way in this case is to use a parametric test such as likelihood ratio test or the method of comparing maximum likelihood estimates proposed by Thoman and Bain (Lam and Spelt, 2007). In all the above tests we have considered only uncensored cases. Clearly, considering only uncensored cases will increase the mean of the survival time (Li and Lagakos, 1997).

To avoid the model validity issues for left truncated and right censored data (LTRC), we have used the nonparametric approach, supported by the well-developed Kaplan-Meier product limit estimator and related techniques, in the second part of the methods, to compare the survival functions of two groups, on the basis of gender and age at the time of diagnosis of diabetes. We have used the KM estimator to estimate the survival function for both the groups because this estimate is the most important and widely used method to estimate the survivor function and it is a generalization of the empirical survivor function, which accommodates censored observations also (Collet, 2003). The KM method with weights W_{KM}, a censored data analog of the two sample t-test, suggest a test which down weights differences late in time when there is heavy censoring (Klien and Moeschberger, 2003). The weight W_{AG}, suggested by us, using the KM method also includes the ratio in which censored observations appear in the data. This method helps us to compare the survivor function of two groups over different time periods, as the data includes patients with minimum 5.6 years and maximum 27-year duration of disease.

Another method to compare the survival function of two groups with LTRC data, based on median survival function, derived by Brookmeyer and Crowley, is

where median survival time is based on the common survival function, $S_w(t) = \frac{m_1\hat{S}_1(t) + m_2\hat{S}_2(t)}{m}$. We have modified $S_w(t)$, by pooled survival function, which meets the situation when the number of events in $\hat{S}_1(t)$ and $\hat{S}_2(t)$ is different as well as censoring and event pattern of group is different. But this pooled survival function depends on the censoring patterns in the two samples. It is found that two groups are not significantly different when we have compared the median survival function of DN onset times. Since the distribution of survival time tends to be positively skewed, the median is the preferred summary measure of the location of the distribution (Collet, 2003).

From the parametric and nonparametric methods used in this study we conclude that the DN onset time of male group differs significantly from that of the female group. Also, the DN onset time of patients whose age at diabetes diagnosis is less than or equal to 45 years differs significantly from that of the group whose age at diabetes diagnosis is more than 45 years. For future studies, the procedures used in this paper can be used for the comparison of survival time of two independent groups in any biomedical study. The application of these models to other complex event history data can also be explored.

REFERENCES:

ASHFAQ, S., GHAZZAL, Z. DOUGLAS, J. S., MORRIS, D. C., VELEDAR, E., WEINTRAUB, W. S., (2006). Impact of diabetes on five-year outcomes after vein graft interventions performed prior to the drug-eluting stent era. The journal of invasive Cardiology, 16(3), pp. 100–105.

COLLETT, D., (2003). Modeling survival data in medical research. 2nd ed. London: Chapman and Hall/CRC.

CZEKALSKI, S., (2005). Diabetic nephropathy and cardiovascular diseases. Annales Academiae Medicae Bialostocensis, 50, pp. 122–125.

FELLINGER, E. K., LEAVITT, B. J., HEBERT, J. C., (2002). Serum levels of prophylactic cefazolin during cardiopulmonary bypass surgery. Ann. Thorac. Surg., 74, pp. 1187–1190.

GROSS, J. L., DE AZEVEDO, M. J., SILVEIRO, S. P., CANANI, L. H., CARAMORI, M. L., ZELMANOVITZ,T., (2005). Diabetic nephropathy: diagnosis, prevention, and treatment. Diabetes Care, 28, pp. 164–176.

JOSS, N., PATERSON, K. R., DEIGHAN, C. J., SIMPSON, K. AND BOULTON-JONES, J. M., (2002). Survival Analysis Techniques for Censored and Truncated data. 2nd ed. Springer Science Business Media Inc.

LAM, R. AND SPELT, J. K., (2007). Comparison of Weibull Small Samples using Monte Carlo Simulations. Quality and Reliability Engineering International, 23, pp. 503–513.

LEE, E. T., WANG, W. Y., (2003). Statistical Methods for Survival Data Analysis, 3rd ed. New Jersey: John Wiley & Sons.

LI, Q. H., LAGAKOS, S. W., (1995). Use of the Wei-Lin-Weissfield Method for the Analysis of a Recurring and a Terminating Event. Statistics in Medicine, 16, pp. 925–940.

PEPE, M. S., FLEMING,T. R., (1989). Weighted Kaplan-Meier Statistics: A Class of Distance Tests for Censored Survival Data. Biometrics; 45(2), pp. 497–507.

ROSSING, P., TARNOW, L., NIELSEN, F. S., BOELSKIFTE, S., BRENNER, B. M., PARVING, H-H., (1995). Short stature and diabetic nephropathy. BMJ, 310, pp. 525–528.

THOMAN, D. R. BAIN, L. J., (1969). Two sample tests in the Weibull Distribution. Technometrics, 11 (4), pp. 805–815.

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), pp. 3070–3076.

VISWANATHAN, V., (2004). Prevention of diabetic nephropathy: A diabetologist's perspective. Indian J. Nephrol, 14, pp. 157–162.

WILD, S., GREEN, G. A., SICREE, R., KING, H., (2004). Global prevalence of diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care, 27, pp. 1047–1053.