Fuzzy Systems and Soft Computing
ISSN : 1819-4362

A HYBRID APPROACH TO SURVIVAL ANALYSIS: MERGING FUZZY LOGIC, WEIBULL SMOOTHING, AND BOOTSTRAPPING

Ramakrishnan. M Associate Professor, Department of Mathematics, RKM Vivekananda College, Chennai, Tamil Nadu - 600004, India. *Email: mramkey@rkmvc.ac.in*
Manikandan. M Research Scholar, Department of Mathematics, RKM Vivekananda College, Chennai, Tamil Nadu – 600004, India. *Email: manijee211993@gmail.com*
Ravanan. R Joint Director of Collegiate Education, Chennai region, Chennai, Tamil Nadu – 15, India. *Email: ravananstat@gmail.com*

Abstract:
Survival analysis is a crucial statistical field used to analyse the time until an event of interest occurs, often applied in medical research, engineering, and social sciences. This study explores and compares multiple advanced statistical methodologies for analysing survival data, including non-parametric, semi-parametric, Weibull smoothing, fuzzy, and bootstrapping techniques, highlighting their theoretical underpinnings, practical applications, and comparative effectiveness. Non-parametric methods, such as the Kaplan-Meier estimator, provide a flexible approach without assuming a specific distribution. Semi-parametric models, notably the Cox proportional hazards model, offer a balance by incorporating covariates without specifying the baseline hazard. Weibull smoothing adds parametric flexibility, accommodating different hazard shapes. Fuzzy methods introduce a novel perspective by handling imprecision and uncertainty in survival data. Bootstrapping enhances the robustness of our estimates through resampling techniques. By integrating these diverse approaches, for the Diabetic retinopathy patient and observed the importance of these techniques and identified the applications of the Bootstrap to Survival Analysis.

Keywords-
Non-Parametric, Semi-parametric, Weibull Smoothing, Fuzzy Model, Bootstrapping

INTRODUCTION

Non-parametric methods in survival analysis are commonly used for their flexibility and minimal assumptions. The Kaplan-Meier estimator remains a cornerstone for estimating survival functions. Recent studies have focused on improving its computational efficiency and extending its applications. For instance, recent advancements involve integrating Kaplan-Meier with machine learning algorithms to enhance predictive performance in large datasets (Wang et al., 2021). Another significant development is the adaptation of the log-rank test for comparing survival curves under complex sampling designs (Smith et al., 2020).

The Cox proportional hazards model is the most widely used semi-parametric method in survival analysis. Recent literature has explored its extensions and adaptations. A notable area of research includes the development of time-varying covariate models to handle non-proportional hazards, as well as the incorporation of high-dimensional data (Liang et al., 2022). Moreover, Bayesian semi-parametric approaches have been proposed to provide more robust parameter estimates and account for model uncertainty (Yu & Zeng, 2021).

Weibull models are favoured for their parametric flexibility, accommodating various hazard shapes. Recent studies have applied Weibull smoothing to improve survival predictions, particularly in biomedical research. Innovations include the use of spline functions to smooth Weibull estimates, offering better adaptability to data with irregular hazard functions (Chen et al., 2021). Another trend is the combination of Weibull models with machine learning techniques to enhance predictive accuracy and interpretability (Kim & Lee, 2020).

Fuzzy logic offers a unique approach to handling uncertainty and imprecision in survival data. Recent applications of fuzzy methods in survival analysis include the development of fuzzy survival models that incorporate linguistic variables and expert knowledge (Hajialiasghari et al., 2021). These

models have shown promise in medical decision-making, where precise data may not always be available. Additionally, the integration of fuzzy clustering with survival analysis has been explored to identify subgroups with distinct survival patterns (Zadeh & Ahmadi, 2022).

Bootstrapping techniques provide a powerful tool for assessing the reliability of survival estimates. Recent research has focused on enhancing bootstrap methods to handle censored data more effectively. For example, advancements in weighted bootstrap techniques have been made to improve the estimation of confidence intervals in complex survival models (Efron et al., 2021). Furthermore, the application of bootstrapping in validating predictive models has been widely explored, ensuring robust performance in various contexts (Davison & Hinkley, 2020). The field of survival analysis continues to evolve with the integration of advanced statistical methods and computational techniques. Non-parametric, semi-parametric, Weibull smoothing, fuzzy methods, and bootstrapping each contribute uniquely to the analysis of survival data, addressing different challenges and improving the robustness and accuracy of survival estimates. Ongoing research and development in these areas promise to further enhance the capabilities of survival analysis in various scientific and practical applications.

MATERIALS AND METHODS

Diabetic retinopathy patient (Vision level improvement is event). The data set were collected from EYDOX hospital private limited, Chennai. Data set contains 91 observations with 12 covariates in follow up period during six years between 2016 to 2022.The covariates are ID, Age, Gender, Marital Status, start time, end time, status, comorbidities, treatment, eyes, Vision levels and glass. Start time is when subject started to the treatment and end time is subject stop the treatment or withdraw from the treatment. Status consider to vision levels (low, medium, high). Comorbidities: Diabetics, Hypertension, Renal status, Cardio, TB, cholesterol. Treatment levels: 1 tablet, 2 tablets and injection and 3 tablet, injection and surgery. Glass: suggested or not. The author consider only 84 patients and remaining observations are eliminated because it contains missing values. 48 cases attained event of interest and the remaining are censored. In this study, attaining Improvement of vision level considered as event and not attained improvement and withdrawn from the study were considered as censored.

1. Survival Statistical Methods

Statistical methods like Nonparametric, Semiparametric and Parametric models are applied to many survival data.

1.1 Kaplan-Meier survival estimate

The Kaplan-Meier (KM) method is a non-parametric method used to estimate the survival probability from observed survival times (Kaplan and Meier, 1958).

The Kaplan-Meier estimator is given by:

$$\hat{S}(t) = \prod_{t_i \le t} \left(1 - \frac{d_i}{n_i}\right) \qquad (1.1)$$

Variance of the Survivorship is obtained using Greenwood's formula [1926]

$$\widehat{Var}\left(\hat{S}(t)\right) = \left(\hat{S}(t)\right)^2 \sum_{t_i \le t} \frac{d_i}{n_i(n_i - d_i)} \qquad (1.2)$$

1.2 The Nelson-Aalen Estimator

Nelson Aalen estimator is also Non-Parametric. It doesn't work with distributions. This method like KM is used to estimate data with censored one. The Nelson-Aalen estimator is presented below

$$\hat{S}(t_i) = \prod_{j=1}^{i} e^{\frac{-d_j}{n_j}} \qquad (1.3)$$

The KM estimator is an approximation of Nelson-Aalen estimator. When d_j is small relative to n_j, which it will except at the longest survival time. The Nelson-Aalen estimator of survival function will always be greater the KM estimator at any given time.

2. Weibull Smoothing:

Rossa and Zielinski proposed a local smoothing of the Kaplan Meier Estimator based on an approximation by means of the piecewise Weibull survival function

$$S_{Wei}(x;\beta,\gamma) = \exp(-\beta x^{\gamma}),\ \beta,\gamma > 0, x > 0 \qquad (2.1)$$

Let us denote $x_1, x_2, \dots, x_N$ the jump points of KM, and $P_1, P_2, \dots, P_N$ the values of the KM at these jump points,

$$P_i = \hat{S}(x_i), \quad i = 1,2,\dots,N$$

Let us define $\overline{P}_i = \frac{P_{i-1}+P_i}{2}$, $i = 1,2,\dots,N$. For i = N we define $\overline{P}_N = \frac{P_N}{2}$ if the last observation is censored and $\overline{P}_N = P_N$ otherwise. For i = 0 we put $\overline{P}_0 = 1$.

Determine the values of β_i and γ_i from the following equations. For i = 1,2, ... N, we have

$$S_{wei}(x_i;\hat{\beta}_i,\hat{\gamma}_i) = \overline{P}_i$$
$$S_{wei}(x_{i+1};\hat{\beta}_i,\hat{\gamma}_i) = \overline{P}_{i+1}$$

Now the smoothed estimator becomes

$$\overline{S}_X(x) = S_{wei}(x_i;\hat{\beta}_i,\hat{\gamma}_i) \quad for\ x \in (x_i, x_{i+1}) \qquad (2.2)$$

The above equations (Non Linear) are solved to get the required beta and gamma parameters, That represents the scale and shape parameters of Weibull distributions respectively. Therefore for a fixed i = 1,2, ... , N we have the equations;

$$\begin{cases} e^{-\hat{\beta}_i {x_i}^{\hat{\gamma}_i}} = \overline{P}_i \\ e^{-\hat{\beta}_i {x_{i+1}}^{\hat{\gamma}_i}} = \overline{P}_{i+1} \end{cases} \qquad (2.3)$$

Taking log on both sides in the equations we get

$$\begin{cases} \hat{\beta}_i {x_i}^{\hat{\gamma}_i} = -\ln \overline{P}_i \\ \hat{\beta}_i {x_{i+1}}^{\hat{\gamma}_i} = -\ln \overline{P}_{i+1} \end{cases} \qquad (2.4)$$

Again taking logarithms on both sides we obtain

$$\begin{cases} \ln \hat{\beta}_i + \hat{\gamma}_i \ln x_i = \ln(-\ln \overline{P}_i) \\ \ln \hat{\beta}_i + \hat{\gamma}_i \ln x_{i+1} = \ln(-\ln \overline{P}_{i+1}) \end{cases} \qquad (2.5)$$

Subtracting the above equations

$$\therefore\ \hat{\gamma}_i = \frac{\ln(-\ln \overline{P}_{i+1}) - \ln(-\ln \overline{P}_i)}{(\ln x_{i+1} - \ln x_i)} \qquad (2.6)$$

From equations we also have

$$\ln \hat{\beta}_i = \ln(-\ln \overline{P}_i) - \overline{\gamma}_i \ln x_i \qquad (2.7)$$

The last expression can be written as

$$\ln(-\ln \overline{P}_i) = \overline{\gamma}_i \ln x_i + \ln \hat{\beta}_i \qquad (2.8)$$

3.1 The Weibull plot

Let us assume the coordinate system (u, v) such that

$u_i = \ln x_i$ and $v_i = \ln(-\ln \overline{P}_i)$ (2.9)

Now we can write the equation in the form

$v_i = \hat{\gamma}_i u_i + \hat{b}_i$ Where $\hat{b}_i = \ln \hat{\beta}_i$.

Thus the estimates $\hat{\gamma}_i$ and $\hat{b}_i$ can also be expressed as follows

$$\hat{\gamma}_i = \frac{v_{i+1}-v_i}{u_{i+1}-u_i} \quad \text{and} \quad \hat{b}_i = v_i - \hat{\gamma}_i u_i \qquad (2.10)$$

Censoring times can change coordinates on the Weibull plot. Estimates of parameters depends on the length of time T. So, source of Uncertainty occurs. The Problem of estimation reduces to linear fuzzy regression analysis.

3. Bootstrapping

First application of bootstrap was made in the context of survival analysis (Efran,1981). Let (X_1, D_1), (X_n, D_n) be identically independently distributed vectors, X and D are independent.

Bootstrap estimate of the variance of the Kaplan-Meier estimator at a point suggested by Efran as follows: Given the sample $(U_1, \delta_1), \ldots, (U_n, \delta_n)$ where U_i= min(X_i, D_i) and

$$\delta_i=\begin{cases}1, & if \quad U_i = X_i \ (uncensored\ observation)\\ 0, & if \quad U_i = D_i (censored\ observation)\end{cases} \qquad (3.1)$$

Take a bootstrap sample (giving probability 1/n to each element of the previous sample):
$(U_1^*, \delta_1^*), \ldots, (U_n^*, \delta_n^*)$
Efron shows that this plane is equivalent to resampling $X_1^*,\ldots, X_n^*$ from $X_1, \ldots, X_n$ and $D_1^*,\ldots, D_n^*$ from $D_1, \ldots, D_n$

U_i^*=min(X_i^*, D_i^*) and $\delta_i^*=\begin{cases}1, & if \quad U_i^* = X_i^*\\ 0, & if \quad U_i^* = D_i^*\end{cases}$ (3.2)

We compute the Kaplan-Meier estimator for the bootstrap sample, $\hat{S}_X^*(t)$ and define the bootstrap estimator of the variance of $\hat{S}_X(t)$, which we may indicate $\hat{\sigma}_B$ as the standard deviation of $\hat{S}_X^*(t)$. After B bootstrap samples, standard deviation of the corresponding Kaplan-Meier estimators at the point t is computed. This is useful to find Standard Error and Confidence limits for the bootstrap survival function estimator.

4. Semi-parametric method

The Cox proportional-hazards model (Cox, 1972) is essentially a regression model commonly used statistical method in medical research for investigating the association between the survival time of patients and one or more predictor variables.

The Cox model is expressed by the hazard function denoted by h(t). Briefly, the hazard function can be interpreted as the risk of event at time t. It can be estimated as follow:

$$h(t, X) = h_0(t) e^{\sum_1^p \beta_i X_i} \qquad (4.1)$$

where,

- *t* represents the survival time
- h(t) is the hazard function determined by a set of p covariates $(X_1, X_2, \ldots., X_P)$
- The coefficients $(\beta_1, \beta_2, \ldots., \beta_p)$ measure the impact (i.e., the effect size) of covariates.
- The term $h_0(t)$ is called the baseline hazard. It corresponds to the value of the hazard if all the X_i are equal to zero (the quantity exp(0) equals 1).

4.1 Estimation of parameter and its Standard Error

Cox (1972) proposed "partial likelihood function" that depends only on the parameter of interest and the resulting parameter estimators from the partial likelihood function would have the same distributional properties as full maximum likelihood estimators

The partial log-likelihood function is then

$$LL(\beta) = L_p(\beta)$$
$$= \sum_{i=1}^{k}(\beta_1 z_{1i} + \beta_2 z_{2i} + \ldots + \beta_p z_{pi}) - \sum_{i=1}^{k} \log\left[\sum_{l \epsilon R(t_{(i)})} \exp\left(\beta_1 z_{1l} + \beta_2 z_{2l} + \ldots + \beta_p z_{pl}\right)\right] \qquad (4.2)$$

Standard errors of the estimates of β_i's can be estimated as $\left(I(\beta)\right)^{-1}$. The estimator of the variance of the estimated coefficient is the inverse of

$$I(\beta) = -\frac{\partial^2 L_p(\beta)}{\partial \beta^2} \qquad (4.3)$$

4.2 Confidence Interval and significance of the covariate

The 100(1-α) percent confidence interval for β_i is $\widehat{\beta_i} \pm Z_{\alpha/2}$(estimated SE of $\widehat{\beta_i}$). A 100(1-α) percent confidence interval for relative risk can be obtained by using the confidence interval for β. Let $\left(\beta_{1L}, \beta_{1U}\right)$ be the 100(1-α) percent confidence interval for β_1; a 100(1-α) percent confidence interval for relative risk is (exp(β_{1L}), exp(β_{1U})). Hazard assumption may be met in many situations, it is not reasonable in others.

- Significance of the covariate is tested by Wald statistic

$$z = \frac{\hat{\beta}}{\widehat{SE(\hat{\beta})}} \qquad (4.4)$$

5. Parametric Models

The estimator of the coefficients in parametric model is obtained by log-likelihood function

$$L(\beta) = \sum_{i=1}^{n} c_i z_i - e^{z_i} \qquad (5.1)$$

where , $z_i = y_i - x_i'\beta$,

$y_i = \ln(t_i)$

$x_i' = (x_{i0}, x_{i1}, \ldots, x_{ip})$ and $x_{i0}=1$

The Likelihood equations are obtained by differentiating the log-likelihood function with respect to the unknown parameters and setting the expressions equal to zero. Here also, the inverse of the observed information matrix provides the estimators of the variances.

Bootstrapping Confidence Intervals

The distribution of $\hat{\theta}^{(b)}$or $\hat{\theta}$ are generated using all the B bootstrap samples. Also, $c_{\frac{\alpha}{2}}$ and $c_{1-\frac{\alpha}{2}}$ are the $\frac{\alpha}{2}$ and 1- $\frac{\alpha}{2}$ quantiles of the $\hat{\theta}^{(b)}$or $\hat{\theta}$ distribution respectively.

Percentile Method

The Percentile confidence interval is derived based on the distribution of $\hat{\theta}^{(b)}$ as $\left[c_{\frac{\alpha}{2}}, c_{1-\frac{\alpha}{2}}\right]$

RESULTS AND DISCUSSION:

Table 1: Comparison of Kaplan-Meier and Bootstrapping survival probability for DEP data

Time	BSSP	2.50%	97.50%	BLCI	KM	2.50%	97.50%	KLCI
2	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	0.0000
3	0.9875	0.9524	1.0000	0.0476	0.9880	0.9648	1.0000	0.0352
6	0.9886	0.9634	1.0000	0.0366	0.9880	0.9648	1.0000	0.0352
12	0.9748	0.9397	1.0000	0.0603	0.9758	0.9431	1.0000	0.0569
14	0.9637	0.9252	1.0000	0.0748	0.9636	0.9239	1.0000	0.0761
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
135	0.7777	0.6744	0.8622	0.1878	0.7764	0.6876	0.8767	0.1891
136	0.7754	0.6785	0.8701	0.1916	0.7764	0.6876	0.8767	0.1891
141	0.7617	0.6548	0.8514	0.1966	0.7609	0.6696	0.8646	0.1950
141	0.7659	0.6667	0.8538	0.1871	0.7609	0.6696	0.8646	0.1950
144	0.7462	0.6456	0.8390	0.1934	0.7450	0.6514	0.8521	0.2007
147	0.7287	0.6322	0.8317	0.1995	0.7292	0.6334	0.8394	0.2059
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
511	0.5507	0.4225	0.6717	0.2492	0.5459	0.4338	0.6871	0.2533
564	0.5286	0.3971	0.6485	0.2514	0.5257	0.4129	0.6693	0.2564
607	0.5020	0.3767	0.6267	0.2500	0.5055	0.3923	0.6513	0.2590
612	0.5036	0.3819	0.6344	0.2525	0.5055	0.3923	0.6513	0.2590
615	0.4805	0.3468	0.6087	0.2619	0.4844	0.371	0.6326	0.2616
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1526	0.1783	0.0638	0.3121	0.2483	0.1824	0.0904	0.3680	0.2776
1578	0.1467	0.0401	0.2662	0.2261	0.1459	0.0638	0.3338	0.2700
1579	0.1139	0.0000	0.2291	0.2291	0.1094	0.0402	0.2982	0.2580
1586	0.0737	0.0000	0.1673	0.1673	0.0730	0.0202	0.2631	0.2428
1754	0.0410	0.0000	0.1185	0.1185	0.0365	0.0055	0.2411	0.2355
2136	0.0599	0.0256	0.1349	0.1093	0.0365	0.0055	0.2411	0.2355

In the above table, we presented survival Probabilities of Kaplan-Meier and Bootstrap survival estimate. Kaplan-Meier estimates gives survival probabilities for Event observations only. On the other hand, Bootstrap survival estimates gives value for censored and uncensored cases.

The median survival time of KM and Bootstrap is 615 days. There are small changes in both methods. From the above table, it is observed that survival probabilities were observed for the observations lies between two event observations in Bootstrap method, but KM gives same probabilities for the observations lies between two event observations.

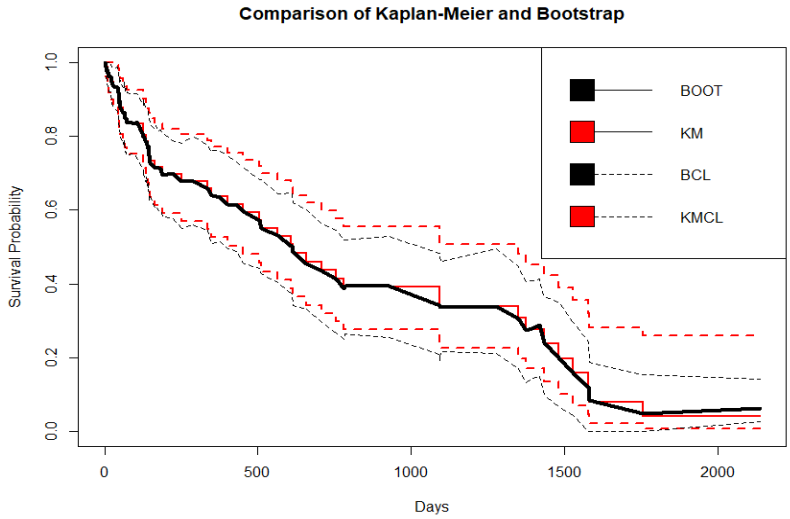

Figure 1. Comparison of KM and Bootstrap with confidence intervals for DEP data

Above Figure 1 indicates, when compared to KM survival curve, Bootstrap Survival Curve is smoother and no jumps.

Table 2: Comparison of Nelson-Aalen and KM Bootstrapping survival probability for DEP data

Time	BSSP	2.50%	97.50%	BLCI	Nelson-Aalen	2.50%	97.50%	NALCI
2	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000	0.0000
3	0.9875	0.9524	1.0000	0.0476	0.9880	0.9650	1.0000	0.0350
6	0.9886	0.9634	1.0000	0.0366	0.9880	0.9650	1.0000	0.0350
12	0.9748	0.9397	1.0000	0.0603	0.9759	0.9435	1.0000	0.0565
14	0.9637	0.9252	1.0000	0.0748	0.9638	0.9244	1.0000	0.0756
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
135	0.7777	0.6744	0.8622	0.1878	0.7779	0.6895	0.8776	0.1880
136	0.7754	0.6785	0.8701	0.1916	0.7779	0.6895	0.8776	0.1880
141	0.7617	0.6548	0.8514	0.1966	0.7625	0.6717	0.8656	0.1938
141	0.7659	0.6667	0.8538	0.1871	0.7625	0.6717	0.8656	0.1938
144	0.7462	0.6456	0.8390	0.1934	0.7468	0.6537	0.8532	0.1995
147	0.7287	0.6322	0.8317	0.1995	0.7311	0.6358	0.8405	0.2047
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
511	0.5507	0.4225	0.6717	0.2492	0.5497	0.4380	0.6898	0.2517
564	0.5286	0.3971	0.6485	0.2514	0.5297	0.4174	0.6723	0.2549
607	0.5020	0.3767	0.6267	0.2500	0.5097	0.3970	0.6545	0.2575
612	0.5036	0.3819	0.6344	0.2525	0.5097	0.3970	0.6545	0.2575
615	0.4805	0.3468	0.6087	0.2619	0.4889	0.3758	0.6360	0.2601
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1526	0.1783	0.0638	0.3121	0.2483	0.1940	0.1000	0.3760	0.2760
1578	0.1467	0.0401	0.2662	0.2261	0.1588	0.0736	0.3428	0.2692

Time	BSSP	2.50%	97.50%	BLCI	Nelson-Aalen	2.50%	97.50%	NALCI
1579	0.1139	0.0000	0.2291	0.2291	0.1237	0.0497	0.3079	0.2582
1586	0.0737	0.0000	0.1673	0.1673	0.0886	0.0289	0.2721	0.2433
1754	0.0410	0.0000	0.1185	0.1185	0.0537	0.0121	0.2384	0.2263
2136	0.0599	0.0256	0.1349	0.1093	0.0537	0.0121	0.2384	0.2263

In the above table, we presented survival Probabilities of Nelson-Aalen and with KM Bootstrap. Nelson-Aalen estimates gives Survival Probabilities for Event observations only. On the other hand, Bootstrap survival estimates give value for censored and uncensored cases.

The median survival time of NA and Bootstrap is 615 days. There are small changes in both methods. From the above table, it is observed that survival probabilities were observed for the observations lies between two event observations in Bootstrap method, but NA gives same probabilities for the observations lies between two event observations.

The above Table 2 and 3 shows, when time increases; length of the confidence intervals for the bootstrap is smaller than compared to other traditional methods.

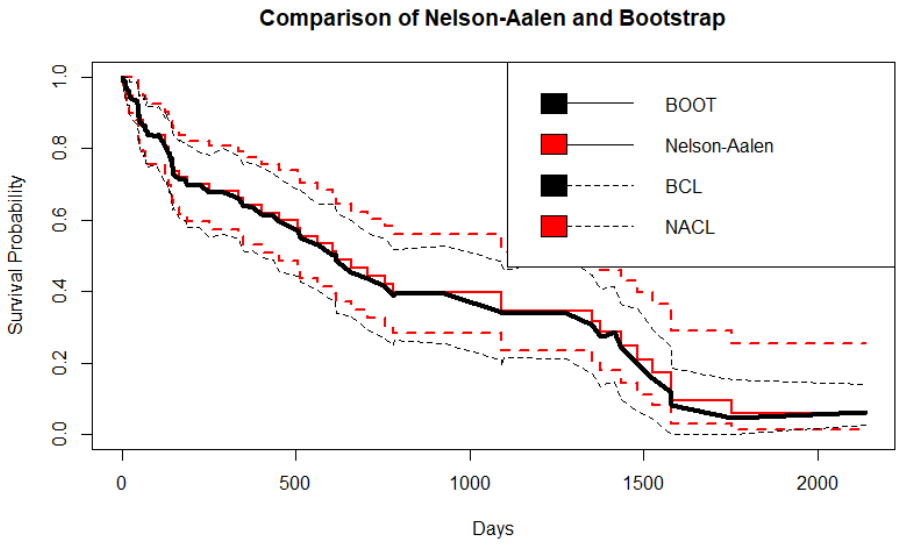

Figure 2. Comparison of Nelson-Aalen with KM Bootstrap with confidence intervals for DEP data

Above Figure 2 indicates, when compared to NA survival curve, Bootstrap Survival Curve is smoother and no jumps.

Table 3: Survival estimates of Kaplan-Meier and Nelson-Aalen with Bootstrapping survival probability for DEP data

Time	KM Bootstrap	Kaplan-Meier	Nelson-Aalen
2	1.0000	1.0000	1.0000
3	0.9875	0.9880	0.9880
6	0.9886	0.9880	0.9880
12	0.9748	0.9758	0.9759
14	0.9637	0.9636	0.9638
⋮	⋮	⋮	⋮
135	0.7777	0.7764	0.7779
136	0.7754	0.7764	0.7779
141	0.7617	0.7609	0.7625
141	0.7659	0.7609	0.7625
144	0.7462	0.7450	0.7468
147	0.7287	0.7292	0.7311
⋮	⋮	⋮	⋮

Time	KM Bootstrap	Kaplan-Meier	Nelson-Aalen
511	0.5507	0.5459	0.5497
564	0.5286	0.5257	0.5297
607	0.5020	0.5055	0.5097
612	0.5036	0.5055	0.5097
615	0.4805	0.4844	0.4889
⋮	⋮	⋮	⋮
1578	0.1467	0.1459	0.1588
1579	0.1139	0.1094	0.1237
1586	0.0737	0.0730	0.0886
1754	0.0410	0.0365	0.0537
2136	0.0599	0.0365	0.0537

It is observed that, the median survival time exists at 615 days in all the methods. KM Bootstrap survival estimates give Survival Probabilities for censored and uncensored cases for each time.

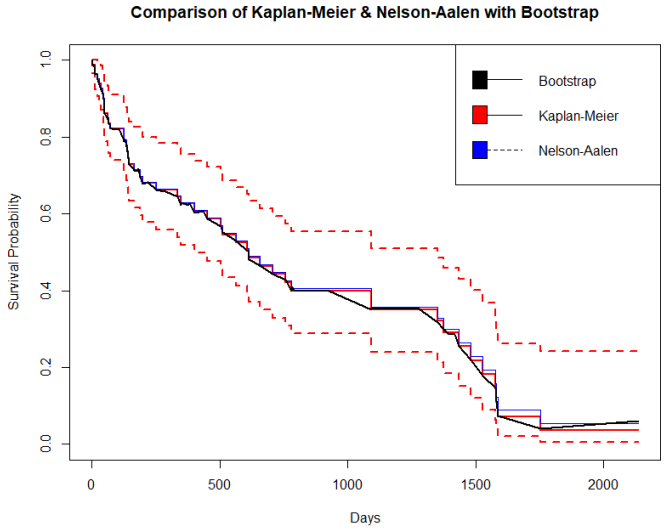

Figure 3. Survival estimates of KM and Nelson-Aalen with KM Bootstrap with confidence intervals for DEP data

Above Figure 3 indicates, when compared to KM and NA survival curve, Bootstrap Survival Curve is smoother and no jumps. Mean width of the confidence interval obtained from KM Bootstrap (0.1912) is less than mean width of the confidence interval obtained from KM (0.2020518) and NA (0.2007675).

Model Assumption checking

The hazard ratio for the two groups should remain proportional under the proportional hazard assumption, which implies that the hazard ratio will remain constant over time. Scaled Schoenfeld residuals are statistical tests and graphical representations that verify the proportional hazard assumption. It verifies the assumption of proportional hazard.

Table 4: Score Tests and p-Values for Proportional Hazards on each of the Covariates as Well as the Global Test for the Model Fit to the DEP Data

	Chisq	p-value
Gender	1.497	0.221
ıent	0.591	0.442
Comorbidities	6.821	0.033
Age	5.686	0.017
Eyes	2.073	0.355
GLOBAL	15.223	0.033

The results in Table 4 indicate that the covariates Gender, Treatment and Eyes satisfies the proportional hazard assumption at 5% level of significance but the covariates Comorbidities and Age satisfies the PH assumption at 1% level of significance. Scatter Plot (Figure 4) of the scaled Schoenfeld residuals also reiterates the results of the score tests.

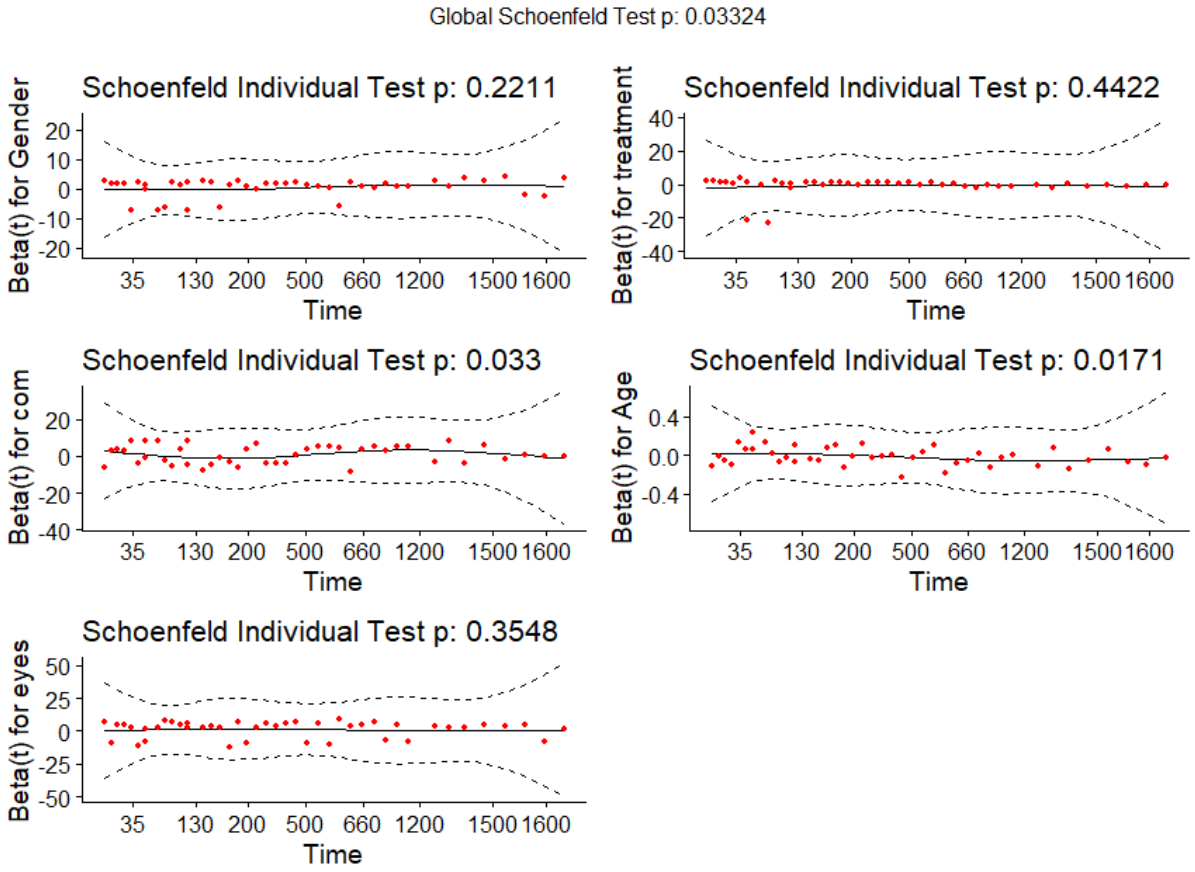

Figure 4. Plotting scaled Schoenfeld residuals against survival time to examine the proportional hazards assumption for DEP patients

A solid line and two dotted lines accompany each other in Figure 4, to represent the results graphically. The smoothing spline fit to the plot is shown by the solid line, while the dotted lines show ± 2 standard error bands around the fit. The Proportional Hazard assumption was not violated in the DEP data. So, the DEP data can be analysed using Cox regression.

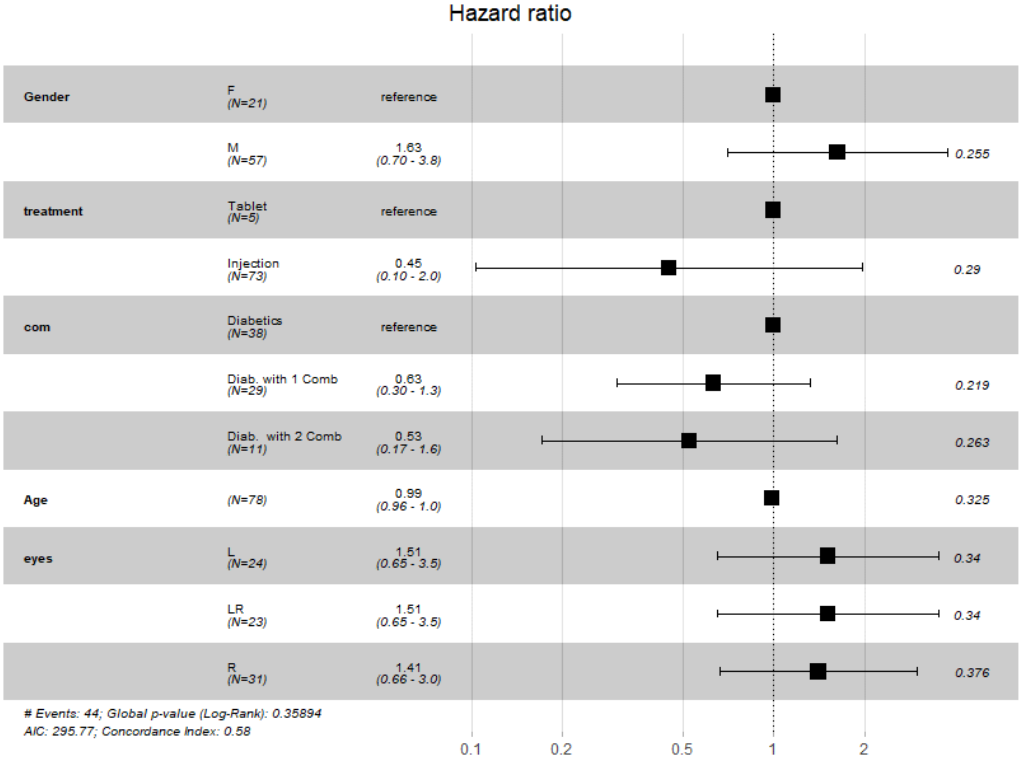

Figure 5. Forest plot of estimates of hazard ratios for the final model fit to the DEP data

Forest plots, as illustrated in Figure 5 for DEP data, are a convenient approach to display and graphically compare the results of multivariate Cox models.

Table 5: Estimated Summary for Main Effects Proportional Hazards Model and Bootstrap technique containing covariates for the DEP Patients

	Cox PH		Bootstrap	
Variables	**coefficient**	**p-value**	**coefficient**	**p-value**
Gender Male	0.4875	0.2550	0.4875	0.3360
Treatment Injection	-0.7969	0.2900	-0.7969	0.5080
comDiab. With 1	-0.4611	0.2190	-0.4611	0.3620
comDiab. With 2	-0.6428	0.2630	-0.6428	0.1990
Age	-0.0128	0.3250	-0.0128	0.4480
Eyes LR	0.4128	0.3400	0.4128	0.3560
Eyes R	0.3412	0.3760	0.3412	0.3190

From the above table, it is understood that the parameter estimates are obtained same values using Cox PH and Bootstrap technique. With out using Standard error, the bootstrap estimates are estimated. All the covariates are not statistically significant because p-values are >0.05. In this study, the number of female patients is less.

Table 6: The AIC/BIC estimated value for Exponential, Weibull, Log-logistic and Lognormal in DEP data.

Models	**AIC**	**BIC**
Exponential	**689.6824**	**708.5361**
Weibull	690.6685	711.8789
Loglogistic	695.6228	717.5549
Log-Normal	696.3445	717.5549

Fitted various Parametric survival models for DEP data and compared the performance of these Parametric models. To select the best Parametric model for DEP data, AIC/BIC values are estimated.

The AIC/BIC estimated value for Exponential, Weibull, Log-Logistic and Lognormal are shown in the Table 6. The least AIC and BIC value gives the best fit model. From the above table, it is observed that Exponential model is suitable for DEP data.

Table 7: Estimated Exponential AFT model and Bootstrap fitted for DEP data

	AFT		Bootstrap	
Variables	**Value**	**p-value**	**Estimate**	**p-value**
GenderM	-0.1914	0.6400	-0.1914	0.7060
treatmentInjection	0.8263	0.2700	0.8263	0.5380
comDiab with 1.	0.3636	0.3100	0.3636	0.4060
comDiab with 2.	0.6705	0.2400	0.6705	0.1880
Age	0.0086	0.5000	0.0086	0.7070
eyesLR	-0.4340	0.3100	-0.4340	0.2930
eyesR	-0.3585	0.3500	-0.3585	0.2600

From the above table, The AFT Exponential model and Bootstrap technique shows that all the covariates are not statistically significant because p values are >0.05. Both Semi-Parametric and Parametric models confirm that all the variables in the study are not statistically significant related to survival time. Both estimates give the same values for estimates and p-values.

None of the variables show statistically significant associations with the hazard in this analysis, as all p-values are greater than 0.05. This suggests that the factors studied (Gender, Treatment, Diabetics comorbidities, Eyes condition, and Age) do not significantly influence the survival in this dataset.

Weibull Model fit for the Eye data is confirmed by the following diagram

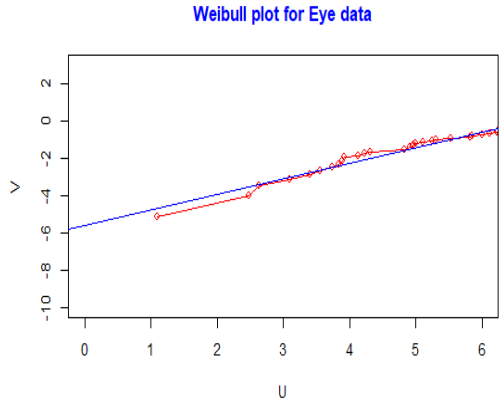

Figure 6. Model fit for Weibull Distribution of Eye Data

From the Figure, it is seen that estimated parameters fit well for the Weibull Structure, as the corresponding plot follows a near linear structure. Table, when the data contains more censoring, the estimated value of parameters differ. But for the eye data which contains 36 censored observations, it is observed that there is change in the values of the parameter. Censoring plays important role in estimating parameters. So, it is observed that this Parameter estimation in Semi-Parametric model leads this survival data into fuzzy model.

Table 8: List of Transformed Values: Coordinates and for the Weibull locally smoothed estimator for Eye Data

Time	KM	kbar	uk	vk	gamma	beta
3	0.9880	0.9940	1.0986	-5.1090	0.7999	0.0025
12	0.9758	0.9819	2.4849	-4.0001	3.3754	0.0000
14	0.9636	0.9697	2.6391	-3.4798	0.7742	0.0040
⋮	⋮	⋮	⋮	⋮	⋮	⋮
74	0.8210	0.8279	4.3041	-1.6668	0.2501	0.0644
125	0.7916	0.8063	4.8283	-1.5357	1.5892	0.0001
135	0.7764	0.7840	4.9053	-1.4134	1.7997	0.0000

⋮	⋮	⋮	⋮	⋮	⋮	⋮
659	0.4634	0.4739	6.4907	-0.2920	0.8406	0.0032
707	0.4423	0.4528	6.5610	-0.2329	0.8712	0.0026
756	0.4212	0.4318	6.6280	-0.1745	1.8508	0.0000
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1579	0.1094	0.1277	7.3645	0.7219	4.2280	0.0000
1586	0.0730	0.0912	7.3690	0.8733	1.9203	0.0000
1754	0.0365	0.0547	7.4697	1.0667	NA	NA

Above Table, shows that the Weibull locally smoothed estimator and the Coordinates for Eye data contains censoring observations. As a result, the censoring variables create complexity, which can be a source of fuzziness. This suggests the researcher to approach survival model in the form of fuzzy linear regression with U_k being crisp and dependent variables V_k being fuzzy. This fuzzy concept used to estimate survival function. From the above table the slope (β) values of all interval Eye data for nearly zero.

Uncertainty observed from Semi-Parametric and Parametric models when removing some censored observations leads the researcher to apply fuzzy linear regression for Eye data. The fuzzy linear coefficients will be considered to have symmetric triangular membership functions for the sake of simplicity.

First converted the crisp numbers into symmetric fuzzy numbers after introduced small random spreads. Spreads were calculated using random method and standard deviation from the data. In Eye data, both methods give same interpretation. For defuzzification, the centroid method was applied to fuzzy number to convert into crisp number.

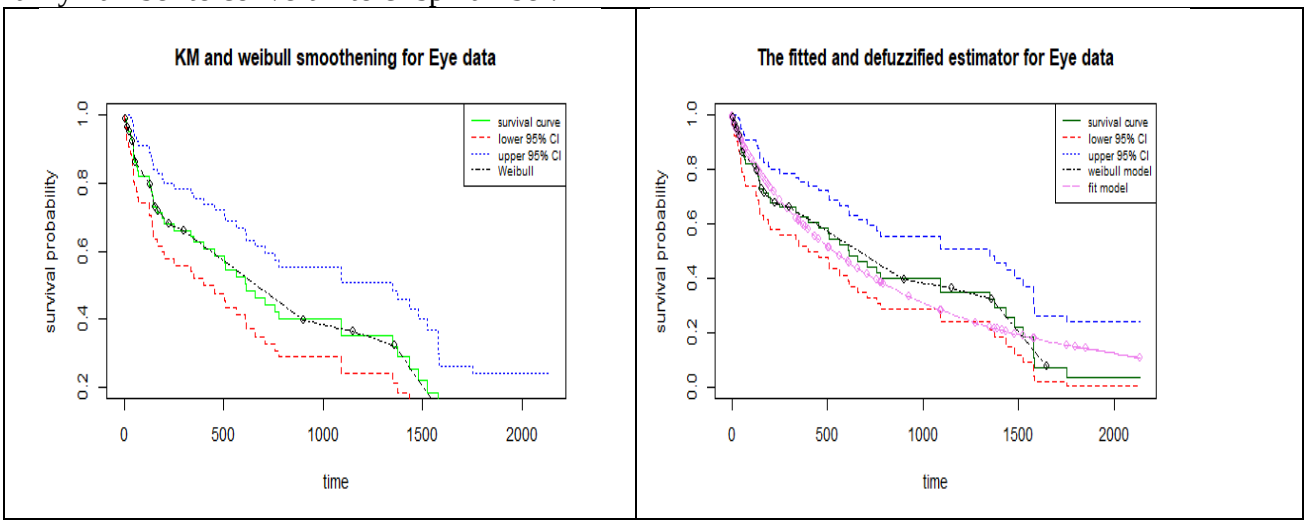

Figure 7. Comparison of KM and Weibull Smoothing and Comparison of KM, Weibull Smoothing and Fuzzy (Fit Model) for Eye data

The plot comparing the KM and Weibull smoothed values is provided for Eye data in Figure 7. Also plot comparing the KM, Weibull smoothed values and the Fuzzy Survival Model is provided for Eye data. Weibull Smoothing and the fuzzy Model present varying survival probabilities, instead of a flat constant one provided by KM Method for Eye data. Also, it is seen that the estimated probabilities using Weibull smoothing and the fuzzy Model fall in between the confidence limits of the KM Estimator.

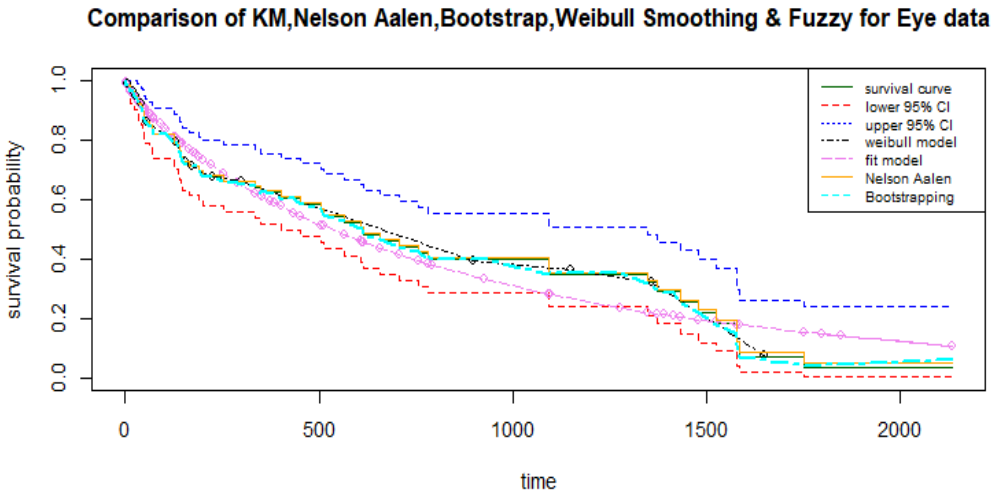

Figure 8. Comparison of KM, Nelson Aalen, Bootstrap, Weibull Smoothing and Fuzzy for Eye data

This plot provides, Kaplan-Meier curve (green) provides an empirical estimate of the survival function without assuming any specific distribution. The Nelson-Aalen estimator (magenta) offers a cumulative hazard function which can be used to derive the survival function. The Weibull smoothing (black dashed line) fits the survival data using a parametric Weibull model. This model assumes a specific form for the hazard function and can provide a smoother estimate. The bootstrap method (Cyan line) is used to provide a sense of the variability in the survival estimates, which helps in understanding the stability of the observed patterns. Confidence Intervals, (red and blue dotted lines) provide a range within which the true survival function is expected to lie with 95% confidence. These intervals indicate the uncertainty around the survival estimates.

The Kaplan-Meier and Nelson-Aalen estimators are both non-parametric methods and provide similar patterns, though the Nelson-Aalen might appear slightly smoother. The Weibull model offers a different perspective by assuming a parametric form, which can be more informative if the underlying assumptions hold true. Bootstrapping highlights the variability in the estimates, and the results show that the observed survival curve is quite stable with narrow confidence intervals.

CONCLUSIONS:

This study examines the effectiveness of bootstrapping on the confidence intervals of various parameter of interest while handling survival data. Applying Nonparametric methods KM and NA to survival data, it is observed that Nelson-Aalen survival estimator is always greater than or equal to KM estimator. Kaplan-Meier and Nelson-Aalen estimates gives survival probabilities for event observations only, but Bootstrap survival estimates gives value for both censored and uncensored cases. Survival probabilities were observed for the observations lies between two event observations in Bootstrap method, but KM gives same probabilities for the observations lies between two event observations. when time increases; length of the confidence intervals for the bootstrap is smaller than compared to other traditional methods. Bootstrap Survival Curve is Smoother and no jumps for probability of survival. Mostly Bootstrapping performs better when sample size was increased. Increasing of Bootstrap samples gives only small variation. The parameter estimates obtained from Cox PH and Bootstrap technique gives same value. Without using Standard error, the bootstrap estimates are estimated. Survival methods like Kaplan-Meier and Weibull Smoothing with the Fuzzy logic approach applied to the data set. Estimate probabilities using Weibull smoothing and the fuzzy Model fall in between the confidence limits of the KM Estimator. Both Semi-Parametric and Parametric models give the same values for estimates and p-values. Combining these methods finally all the curves in between the confidence interval of KM and can often provide a more comprehensive analysis, leveraging the strengths of each approach. Each method's integration with bootstrapping provides a more comprehensive analysis by quantifying the uncertainty and variability of survival

estimates. This combination is especially powerful in survival analysis, where censored data and small sample sizes often pose significant challenges. The choice of method should be guided by the data characteristics and the specific goals of the analysis, with bootstrapping serving as a valuable tool for enhancing the reliability of the results.

REFERENCES:

1. Aalen, O.O. 1978. Non-Parametric Inference for a Family of Counting Processes. The Annals of Statistics. 6, 701-726.
2. Borgan, O., & Liestol, K. (2000). Bootstrap methods in survival analysis. Lifetime Data Analysis.
3. Chaudhuri, P., & Lo, S. H. (2009). Bootstrap methods for interval estimation in survival analysis. Journal of the Royal Statistical Society: Series B.
4. Chen, X., Wang, Y., & Liu, Y. (2021). Applications of spline functions in Weibull smoothing for survival data. Journal of Biomedical Research.
5. Cox, D. R. and Oakes, D. (1984). Analysis of Survival Data. UK: Chapman and Hall.
6. Cox, D.R. (1972). Regression models and life tables. Journal of the Royal Statistical Society, Series B, 34:187-202.
7. Davison, A. C., & Hinkley, D. V. (2020). Bootstrap methods and their application. Cambridge University Press.
8. Efron, B., Hastie, T., & Tibshirani, R. (2021). Improvements in bootstrap methods for censored data. Statistical Science.
9. Hajialiasghari, F., Nasrabadi, A. M., & Maleki, M. (2021). Fuzzy survival models: An approach to handling imprecision in medical decision-making. Artificial Intelligence in Medicine.
10. Kalbfleisch, J. D., & Prentice, R. L. (2002). The Statistical Analysis of Failure Time Data. Wiley Series in Probability and Statistics.
11. Kaplan EL and Meier PL. 1958. Non parametric estimation from incomplete observations. Journal of the American Statistical Association; 53, 457-81.
12. Kim, J., & Lee, S. (2020). Combining Weibull models with machine learning techniques for survival prediction. IEEE Transactions on Biomedical Engineering.
13. Liang, H., Wang, M., & Sun, X. (2022). Bayesian semi-parametric approaches in survival analysis with high-dimensional data. Statistics in Medicine.
14. Mohan, N., Ramakrishnan, M and Ravanan, R. (2022). The Comparison of Survival approach with Fuzzy Logic for National Stock Exchange data during Covid-19 in India. International Journal of Mathematics and Statistics, 23 (2), 41-51.
15. Pokorny, K. and Sule, D. (2004). Empirical Fuzzy Estimate of the Survival Curve. International Journal of Uncertainty, Fuzziness and Knowledge Based Systems. 12(3), 347-356.
16. Shao, J., & Tu, D. (2012). The Jackknife and Bootstrap. Springer Science & Business Media.
17. Smith, R., Jones, L., & Patel, A. (2020). Enhancements in the log-rank test for complex survival data. Journal of Statistical Methods.
18. Therneau, T. M., & Grambsch, P. M. (2000). Modeling Survival Data: Extending the Cox Model. Springer.
19. W. Kosinski and W. Piasecki (2008). On a class of defuzzification functionals, Proccedings of the International Multiconf erence on Computer Science and Information Technology, 2008. IMCSIT 2008, 145-150, Wisla, Poland.
20. Wang, L., Zhang, H., & Li, J. (2021). Integrating Kaplan-Meier estimator with machine learning for large survival datasets. Journal of Computational Statistics.
21. Yu, Y., & Zeng, D. (2021). Bayesian semi-parametric models for survival analysis. Journal of the American Statistical Association.
22. Zadeh, L. A., & Ahmadi, M. (2022). Fuzzy clustering in survival analysis: Applications and innovations. Soft Computing Journal.
23. Zeng, J. and Zhang, A. (2018). Statistical Analysis of Fuzzy Linear Regression Model Based on Different Distances. International Journal of Fuzzy Logic Systems, 8(12), 1-13.