Asian European Journal of Probability and Statistics
Vol. 1, No-2, (2024), pp. 119-135

Evaluating the Performance of the World Top Eight T20 Run Scorers Using Survival Analysis

Ramakrishnan. M[a] and Sathishkumar. M[b] and Viswanathan. N[c]

[a] *Associate Professor, Department of Mathematics, RKM Vivekananda College, Chennai, Tamil Nadu - 600 004, India. Email: mramkey@rkmvc.ac.in*
[b]*Research Scholar, Department of Mathematics, RKM Vivekananda College, Chennai, Tamil Nadu – 600 004, India. Email: m.sathish1098@gmail.com*
[c]*Associate Professor and Head, Department of Statistics, Presidency College, Chennai, Tamil Nadu – 600 005, India. Email: visustats10@gmail.com*

ARTICLE HISTORY
Compiled April 16, 2025

Received 28 July 2024; Accepted 27 December 2024

ABSTRACT
Cricketing knowledge tells us that batting is more difficult early in a player's innings but becomes easier as a player familiarizes themselves with the conditions. A comprehensive dataset of T20 matches is utilized to study the impact of different factors on the survival of batsmen in the highly dynamic and fast-paced T20 format. Survival analysis, in the context of T20, models dismissal as an event. Each batsman's innings represents a "lifetime" until dismissal occurs. This research compares the effectiveness of Non-Parametric, Semiparametric, and Parametric survival analysis methods using T20 data. This analysis utilizes several survival models, including the Kaplan-Meier method for estimating survival rates, the Log-rank test for comparing survival differences between groups, and the Cox Proportional Hazards model for calculating hazard ratios. Additionally, AIC and BIC values were employed to identify the most appropriate survival distribution for each player, which was then applied into a parametric regression model to generate time ratios for each group. Furthermore, Conditional Survival Probabilities can be beneficial for team management in determining or adjusting the batting order during a match based on the current game situation and the opposing team. World's top eight T20 run scorer Batsmen up to May 31st, 2024, was taken from www.espncricinfo.com for this study. For this analysis carried out from the R Programming Language and its packages like "survival" and "surviminer" were used.

KEYWORDS
Survival Analysis; Non-Parametric; Semiparametric; Parametric models; Conditional Survival Probability

1. Introduction

One of the most well-liked sports in the world now is cricket. Now-a-days, this game is played throughout countries, states, and even cities. There are three different ways to play cricket: Twenty20 (T20), One Day International (ODI), and Test Cricket.

CONTACT Ramakrishnan. M[a]. Email: mramkey@rkmvc.ac.in

One-day cricket was introduced in the 1960s as an alternative to the Test Cricket characterized by more aggressive batting, colorful uniforms and fewer matches ending in draws. ODI cricket is limited to fifty overs. The biggest event in ODI cricket takes place after every four years when the World Cup of Cricket (WCC) is organized by the International Cricket Council (ICC) which is the global governing body for cricket games. Later in 2003, T20 form of the cricket game was introduced with focus on gaining wider audience and with emphasis on power hitting. Cricket in T20 format is limited to twenty overs. The present research is related to the game of T20's. The topic of this study is the T20 league. Batting is the heart of cricket and bowling is its backbone. According to cricket understanding, batting is more difficult early in a player's innings but improves as players become more comfortable with the pitch conditions.

Survival analysis is a collection of methods for analysing data in which the outcome variable is the time until the occurrence of a specified event. It differs from other fields of Statistics mainly because of the incomplete information that arise due to censoring. A batsman's innings might be thought of as a lifespan (Kachoyan & West, 2016). The authors defined it as "when the batsman goes out to bat, he is 'born' and 'lives' for a certain number of balls before he is dismissed". A dismissal was referred to as a batsman 'death' which is the event of interest. When a batsman was not dismissed during a match that particular observation was referred to as a censored observation. The majority of earlier research that examined individual batsmen's survival ability, i.e. (Kimber & Hansford, 1993; Kachoyan & West, 2016; Brown, 2017; Saikia & Bhattacharjee, 2018), was conducted using the survival function of the number of balls faced till dismissal.

The Kaplan-Meier (1958) method for estimating the survival function, Log-rank statistics, Mantal (1966) for comparing two survival distributions, and the Cox (1972) Proportional Hazards model for quantifying the effects of covariates on survival time are the most significant developments in this field. Cox regression can be employed to determine the variables that significantly influence the outcome of interest, with results expressed as hazard ratios. Ramakrishnan and Ravanan (2013) implemented Non-Parametric approaches and tests across a variety of areas, including sociology and medicine. Mohan et al. (2022) used Survival Methods to study the impact of Covid-19 on NSE sectoral indices. Brief summary of Semiparametric models used in survival analysis by ShaojunGuo and Donglin Zeng (2013). NanamiTaketomi et al., (2022) reviewed important Survival Parametric distribution models. The implementation of Parametric survival models was demonstrated by Mukesh Kumar et al., (2019) through the utilization of R software, which is freely accessible.

Kimber and Hansford (1993) demonstrated utility of Non-Parametric models for estimating hazard of player's. Bracewell and Ruggiero (2009) suggested 'Ducks n runs'distribution for scores of zero to overcome inability of geometric distribution under inflated number of scores of zero. A survival rate criterion is considered by van Staden (2010) for evaluating the performance of batsmen. Saikai and Bhattacharjee (2018) examined survival ability of batsmen in IPL 2012. A detail study regarding research directions in cricket was considered by Swartz (2017). Sachin S.S. Kottearachchi et al. (2022) assessing the survival abilities of the opening batting performance is crucial for addressing Sri Lankan cricket's current decline. Preetham HK and Kumar (2023) proposed a method to predict IPL match outcomes and inning scores. To achieve this, they suggested using various machine learning techniques on specific datasets. Shah et al. (2023) investigated the survival probabilities of the world's top ten batsmen, and their findings can be utilised as a new criterion for evaluating batsmen because they indicate a batsman's capability to survive on the crease.

In this paper, we have taken past record of World Top eight highest run scorers in T20 Matches. This analysis utilizes several survival models, including the Kaplan-Meier method for estimating survival rates, the Log-rank test for comparing survival differences between groups, and the Cox Proportional Hazards model for calculating hazard ratios. Additionally, AIC and BIC values were employed to identify the most appropriate survival distribution for each player, which was then applied into a Parametric regression model to generate time ratios for each group. Furthermore, Conditional Survival Probabilities can be beneficial for team management in determining or adjusting the batting order during a match based on the current game situation and the opposing team.

The study begins with a concise introduction and a comprehensive literature review. Section 2 provides an overview of Non-Parametric, Semiparametric, and Parametric estimation methods for analyzing time-to-event data, with a focus on classical approaches. Section 3 outlines performance measures using the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These survival analysis techniques are subsequently applied to real-world data, specifically the T20 dataset, with the results presented in Section 4. Finally, Section 5 concludes with a summary of the findings.

2. SURVIVAL FUNCTIONS AND METHODS

The survival function is of atmost priority in the field of survival analysis is defined as the probability of survival beyond time t.

$$S(t) = P(T > t) = 1 - F(t)$$

Where T is a random variable denotes the time that the event occurs. The survival function is the complement of the Cumulative Density Function (CDF),

$$F(t) = \int_0^t f(u)du$$

Where f(t) is the probability density function. The hazard function h(t) gives the instantaneous potential per unit timefor the event to occur, given that the individual has survived up totime t.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

The hazard function or the instantaneous rate at which an event occurs at time t given survival until time t is given by,

$$h(t) = \frac{f(t)}{S(t)}$$

The survival function can also be stated in terms of the cumulative hazard function,

$$H(t) = \int_0^t h(u)du$$

$$S(t) = e^{-H(t)}$$

The hazard function and the survival function have a straightforward one-to-one connection. The cumulative hazard function is denoted by H(t).

2.1. Non-Parametric Method

Non-Parametric methods in survival analysis play a crucial role in understanding time-to-event data without making strong assumptions about the underlying distribution of the data. The Kaplan-Meier method is the most popular method used for survival analysis. It provides us an opportunity to estimate survival probabilities and compare survival between two or more groups at a given specific time.

The Kaplan-Meier (KM) method

Survival Probabilities were estimated through product of conditional probabilities without assuming distributional form for survival time. In this model, Survival Functions $S(t) = P(T > t)$ is estimated through.

$$\widehat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where d_i and n_i respectively are the number of events that occur and the number of subjects that enters the study at time t_i, which is the i^{th} ordered survival time. Log-rank test is used to compare the survival patterns across ‘Batting First’ and ‘Chasing’.

2.2. Semiparametric Method

A semiparametric model, such as the Cox Proportional Hazards (PH) model, is commonly used in survival analysis to assess the relationship between covariates and the hazard of an event occurring over time.

Cox Proportional Hazard

Cox (1972) proposed the following regression model for the hazard function

$$h\left(t|X,\beta\right) = h_0(t)e^{\sum_1^p \beta_i X_i}$$

The survival time is denoted by t, and the hazard function, represented as $h\left(t|X,\beta\right)$, is influenced by a set of p covariates, which are denoted as $(X_1, X_2, \ldots, X_p)$. The coefficients $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ quantify the effect of these covariates on the hazard function.

Additionally, the term $h_0(t)$ signifies the baseline hazard, serving as a reference point for understanding how the covariates modify the risk of the event occurring over time. Together, these components form a comprehensive framework for analyzing survival data and assessing the relationships between covariates and the hazard of events.

Proportional Hazard Assumptions

The proportional hazards assumption requires that covariates are multiplicatively related to the hazard. To verify Proportional Hazard assumption, test based on Schoenfeld Residuals is used.

2.3. Parametric Methods

Parametric models can be used if the survival time follows to specific distributions. In the Accelerated Failure Time model, the survival proportion of one group at any time t is equivalent to the survival proportion of the second group at time $\phi * t$, with ϕ being a constant. This investigation explores the application of various well-established distribution models, such as the Exponential, Weibull, Lognormal, and Log-logistic distribution. This paper specifically utilizes the Accelerated Failure Time model to analyze the survival patterns of batsmen.

3. PERFORMANCE MEASURES

Performance between the models is compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), a measure of the goodness of fit for statistical models. The AIC and BIC are given by

$$AIC = -2 * (loglikelihood) + 2(k + c)$$

$$BIC = -2 * (loglikelihood) + (k + c)\log(n)$$

, Where k denotes the number of covariates in the model, not including the constant terms, n is the sample size and c is the number of model specific distributional parameters. Smaller AIC and BIC values of the models indicates better model fit.

4. APPLICATION TO T20 CRICKET DATA

A reliable data source for this research was found in Statsguru. Statsguru is ESPN Cricinfo's cricket statistics maintenance database (www.espncricinfo.com). The details of the performance of T20 2024 World's ICC Top eight high scoring batsmen up to May 31st, 2024 are considered in this study. This data T20 considered the high run scorer Batsmen Virat Kohli, Babar Azam, Rohit Sharma, Paul Stirling, Mohammad Rizwan, David Warner, Jos Buttler and Kane Williamson.

This study investigates the effectiveness of various statistical methods (Non-Parametric, Semiparametric, and Parametric methods) for analyzing batsmen survival times, defined as the number of runs scored. Event of interest of this study is 'getting out' in an innings and considering 'not out' as censored. The author aims to assess the probabilities and factors influencing of dismissal in cricket. Virat Kohli, Babar Azam, Rohit Sharma, Paul Stirling, Mohammad Rizwan, David Warner, Jos Buttler and Kane Williamson played 109, 112, 151, 141, 85, 103, 107 and 87 matches respectively.

Kaplan-Meier's method of estimating Survival function is a Non-Parametric analysis. This method considers only time and event.

Table 1. Descriptive statistics of the selected batsmen

Player Names	Number of Matches	Min.Runs	Q1	Median	Mean	Q3	Max.Runs
Virat Kohli	109	0	13.00	29.00	37.04	59	122
Babar Azam	112	0	9.75	31.50	35.92	55	122
Rohit Sharma	143	0	5.50	15.00	27.79	45	121
Paul Stirling	141	0	7.00	19.00	25.45	35	115
Mohammad Rizwan	85	0	11.00	32.00	37.68	67	104
David Warner	103	0	6.50	23.00	30.09	53	100
Jos Buttler	107	0	8.00	22.00	28.50	40	101
Kane Williamson	87	0	10.50	25.00	29.28	42	95

From the above table.1, it can be seen that Rohit Sharma has played the maximum number of matches with 143, while Virat Kohli and Babar Azam have scored the highest maximum runs of 122 each. Average runs scored indicate overall performance. Kohli leads with a mean of 37.04, closely followed by Rizwan (37.68), demonstrating their consistent scoring ability across matches. Paul Stirling has the lowest average at 25.45.

Table 2. Kaplan-Meier and Survival Probabilities for top 8 batsmen

Runs	Virat Kohli	Babar Azam	Rohit Sharma	Paul Stirling	Mohammad Rizwan	David Warner	Jos Buttler	Kane Williamson
	n=109	n=112	n=143	n=141	n=85	n=103	n=107	n=87
10	0.8065	0.7411	0.6500	0.6950	0.7831	0.6873	0.7330	0.7584
30	0.5319	0.5245	0.3619	0.3373	0.5384	0.3829	0.4153	0.4453
50	0.3761	0.3375	0.2559	0.1810	0.3625	0.2743	0.2791	0.2547
70	0.2467	0.1631	0.1268	0.1080	0.2460	0.1379	0.1289	0.0772
100	0.1583	0.0669	0.0498	0.0144	0.1298	0.0460	0.0552	0.0000

n-Number of Matches Played

This table 2 shows the Kaplan-Meier survival probabilities for the top 8 batsmen based on the number of runs scored in T20 cricket matches. The numbers in the table represent the probability of a player surviving (not getting out) after scoring a certain number of runs in a match. Kohli has 80.65% chance of remaining not out after scoring 10 runs, after scoring 30 runs is 53.19% chance, and so on. Babar, Rizwan, Warner, Buttler and Williamson also show relatively high probabilities of remaining not out after scoring 50 runs (0.3375, 0.3625, 0.2743, 0.2791 and 0.2547, respectively). Kohli demonstrating the highest survival probabilities among all other top batsmen. This also suggests that Kohli Converts a good start into half century and century, which is confirmed by the number of centuries he has scored. Williamson and Stirling have a lower probability of survival compare to others Players. The table also includes the number of matches played (n) for each player, which ranges from 85 to 143 matches. The survival probabilities provide insights into the batting performance and consistency of these players in different run-scoring scenarios.

Figure 1 shows that all players except Kane Williamson hit more than 100 runs. Rohit and Williamson suggest that batsmen have a higher probability of getting out

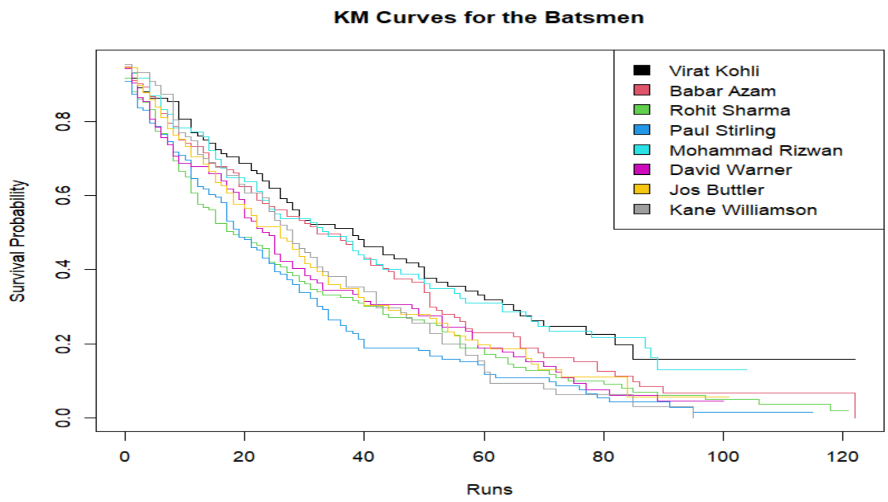

Figure 1. Survival Curves of Top Eight Batsmen

as they score more runs. Overall Kohli, Babar and Rizwan have a high survival probability of getting good runs exhibit the highest survival probabilities, indicating better performance and consistency in scoring runs compared to the other batsmen.

Similarly, survival probabilities of each batsman according to their innings and p-value for corresponding Log-rank statistic are given in the following Table 3.

Table 3. Survival Probabilities of T20 top 8 batsmen Corresponding to Innings

Players	Runs	10	30	50	70	100	p-value
Kohli	Batting First (63, 32.14)	0.6984	0.4105	0.2773	0.1977	0.1648	0.0136*
	Chasing (46,43.74)	0.9556	0.7044	0.5187	0.3175	-	
Babar	Batting First (59,37.47)	0.7797	0.5593	0.3153	0.1356	0.0678	0.8264
	Chasing (53,34.19)	0.6981	0.4861	0.3646	0.1975	0.0705	
Rohit	Batting First (77,32.58)	0.7089	0.4226	0.3099	0.1713	0.0623	0.0499*
	Chasing (66,22.19)	0.5813	0.2906	0.1913	0.0773	0.0386	
Stirling	Batting First (54,26.80)	0.7222	0.3519	0.1852	0.0926	0.0185	0.9784
	Chasing (87, 24.62)	0.6782	0.3280	0.1795	0.1212	-	
Rizwan	Batting First (41,43.24)	0.8537	0.5610	0.4088	0.2811	0.2024	0.1938
	Chasing (44,32.5)	0.7123	0.5158	0.3174	0.2116	-	
Warner	Batting First (40,29.95)	0.6500	0.3250	0.2250	0.1250	0.0500	0.7309
	Chasing (63,30.17)	0.7111	0.4202	0.3062	0.1472	-	
Buttler	Batting First (50,31.68)	0.7400	0.5152	0.3252	0.0965	0.0322	0.5598
	Chasing (57,25.72)	0.7268	0.3203	0.2349	0.1495	-	
Williamson	Batting First (49,28.33)	0.7959	0.4209	0.1994	0.0443	-	0.1043
	Chasing (38,30.50)	0.7096	0.4815	0.3377	0.1313	-	

(a,b): where a denotes number of matches played and b denotes average runs

Table 3 presents the survival probabilities of the top 8 T20 run scorer, considering their performance in different innings scenarios: Batting first and Chasing. The survival probabilities are provided at different run milestones (10, 30, 50, 70, 100) for each player. The Players Kohli, Rohit, Stirling, Buttler and Rizwan have high probabilities of survival when chasing compared to Facing. But Babar and Williamson started taking runs in Facing innings compared to Chasing. We expect centuries from Kohli, Stirling, Rizwan, Warner and Buttler when they are playing Facing innings but Rohit and Babar have high chances for getting centuries in both innings. From

Non-Parametric Log-rank test, the p-value column indicates the statistical significance of the survival probabilities. From Non-Parametric Log-rank test, it is observed that there is a significant difference in getting runs in Facing and Chasing innings for Kohli and Rohit at 5% level of significance. Hence it is concluded that there is a significant difference between scoring pattern of two innings. But there is no significant difference observed from all other batsmen.

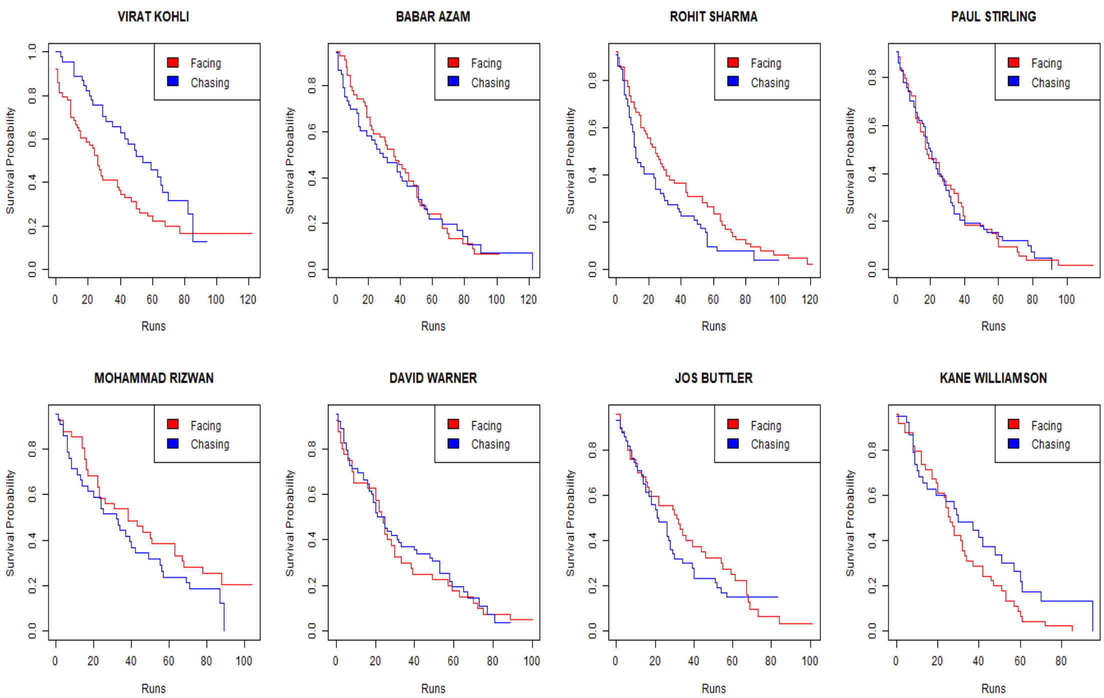

Figure 2. Survival Curves of Batsmen Corresponding to Innings

Figure 2 graphically represents the survival curves for every innings of batsmen. This diagram also confirms that there is a significant difference in getting runs in Facing and Chasing innings for Kohli and Rohit at 5% level of significance. But there is no significant difference observed from all other batsmen.

The following Table 4, shows Chi-Square statistic and p-Values of Log-rank test for the comparison survival distribution among the players.

Table 4. Chi-Square and p-Value for All Players

Batsmen	Kohli	Babar	Rohit	Stirling	Rizwan	Warner	Buttler	Williamson
Virat Kohli		2.32/0.13	11.08/<0.001**	18.6/<0.001**	0.11/0.74	7.48/0.01*	5.36/0.02*	7.73/0.01*
Babar Azam			3.59/0.06*	9.24/<0.001**	1.24/0.27	1.71/0.19	1.08/0.3	2.52/0.11
Rohit Sharma				1/0.32	7.21/0.01*	0.2/0.66	0.84/0.36	0.1/0.75
Paul Stirling					13.4/<0.001**	1.84/0.17	3.18/0.07	1.8/0.18
Mohammad Rizwan						4.89/0.03*	3.65/0.06	5.48/ 0.02*
David Warner							0.13/0.71	0.03/0.86
Jos Buttler								0.14/0.71
Kane Williamson								

a/b: a denotes chi-square statistic, b denotes its corresponding p-value; *denotes 5% level of significance; **denotes 1% level of significance

Table 4. confirms that Kohli has a significant difference in the survival pattern of

scoring runs among all players at the 5% level of significance. Furthermore, excluding Babar and Rizwan, there is a significant difference between Kohli's and the other players' survival distributions at the 5% level. Babar and Stirling also exhibit a significant difference in their survival patterns at 1% level. The difference increases for Rohit and Rizwan, as well as Stirling and Rizwan, with a significant difference in their survival distributions at the 1% level. Rizwan's survival pattern differs significantly from those of Warner, Buttler, and Williamson at the 1% significance level. This information can be valuable for team selection, strategy formulation, and understanding player dynamics in cricket.

Table 5. Chi-Square and p-Value for All batsmen Corresponding to Innings

	Innings	Babar	Rohit	Paul	Rizwan	Warner	Buttler	Williamson
Virat Kohli	I	0.05/0.83	0.31/0.57	3.6/0.06	1.8/0.18	1.32/0.25	0.09/0.76	2.43/0.12
	II	5.8/0.02*	20.97/<0.001**	20.04/<0.001**	4.31/0.04*	10.74/<0.001**	11.39/<0.001**	5.07/0.02*
Babar Azam	I		0.44/0.51	5.35/0.02*	1.97/0.16	1.55/0.21	0.53/0.47	5/0.03*
	II		3.9/0.05*	3.4/0.07	0/0.96	0.44/0.51	0.54/0.46	0/0.99
Rohit Sharma	I			2.53/0.11	3.88/0.05*	0.47/0.49	0/0.96	2.01/0.16
	II			0.18/0.67	4.15/0.04*	2.73/0.1	1.7/0.19	3.94/0.05*
Paul Stirling	I				10.53/<0.001**	0.39/0.53	2.06/0.15	0.06/0.81
	II				3.25/0.07	1.18/0.28	0.69/0.41	3.15/0.08
Mohammad Rizwan	I					5.35/0.02*	3.58/0.06	0.75/<0.001**
	II					0.61/0.44	0.84/0.36	0.02/0.88
David Warner	I						0.36/0.55	0.35/0.55
	II						0.03/0.86	0.43/0.51
Jos Buttler	I							2.21/0.14
	II							0.83/0.36

*denotes 5% level of significance; **denotes 1% level of significance

From Table 5. It is observed that Kohli survival distribution for Chasing is significantly differ from all other Players in this data but Babar Azam only differ from Stirling and Williamson in Facing. Similarly, Riswan differ with Stirling and Williamson while Facing.

These results can be used by team management to understand the performance patterns of these batsmen and make informed decisions about their batting order and strategy, especially in the second innings of the match. Survival probabilities of all batsmen against six selected countries were studied. A batsman country combination was considered for survival probability estimation if the batsman had played at least ten matches against that specific country. These estimates were derived using the Kaplan-Meier model. The survival probabilities of the all-top run scorer batsmen against the six teams. The Log-rank test used for comparing survival differences between groups.

The table 6. Provides survival probabilities of all Batsmen against six different teams (India, Australia, England, New Zealand, Pakistan and Ireland) at different run levels (10, 30, 50, 70, 100).

We can see that Virat has Survival Chances more at initial and after that are highest against Australia and least against New Zealand. Babar Azam had played against more than ten Matches England and New Zealand, it is observed that After he stands initial stage then we expect him to hit half century and century against both countries. Also, Babar have high probabilities of survival against England compared to New Zealand. Rohit has highest survival chance against England but least chance for getting century against Pakistan. He has no chance for getting century against New Zealand, but he has high probability to stand up to 50 runs against New Zealand comparing to others. Rohit has a lower chance of reaching 50 against Pakistan compared to other teams. Rizwan initially exhibits the highest survival chances against New Zealand. He

Table 6. Survival Probabilities of all Players against 6 Teams

Players	Runs	India	Australia	England	New Zealand	Pakistan	Ireland
	NOM		21	20	10	10	2
	10		0.8100	0.7500	0.9000	0.0949	
	30		0.5130	0.4880	0.5000	0.1265	
Virat Kohli	50		0.3590	0.2170	0.2500	0.1610	
	70		0.2870	0.1630		0.1666	
	100		0.1440				
	NOM	4	8	17	24		3
	10			0.8240	0.7917		
	30			0.6370	0.5000		
Babar Azam	50			0.3190	0.3750		
	70			0.1910	0.1406		
	100			0.1270	0.0938		
	NOM		19	14	17	10	3
	10		0.5789	0.7860	0.5647	0.4800	
	30		0.2895	0.3570	0.5020	0.1200	
Rohit Sharma	50		0.2316	0.2140	0.3137		
	70		0.0772	0.1070	0.0627		
	100			0.1070			
	NOM	6	2	2	4	3	
	10						
	30						
Paul Stirling	50						
	70						
	100						
	NOM	4	5	14	20		3
	10			0.7140	0.7500		
	30			0.5000	0.4500		
Mohammad Rizwan	50			0.4290	0.2400		
	70			0.2140	0.1800		
	100						
	NOM	9		14	10	16	2
	10			0.4286	0.7000	0.6730	
	30			0.2857	0.3000	0.4040	
David Warner	50			0.2143	0.2000	0.1790	
	70			0.0714			
	100						
	NOM	18	15		12	11	1
	10	0.6670	0.8000		0.8330	0.6364	
	30	0.3750	0.4670		0.6250	0.3636	
Jos Buttler	50	0.3000	0.3330		0.3470	0.1818	
	70	0.1500	0.1780		0.1740	0.0909	
	100						
	NOM	13	11	8		21	1
	10	0.8462	0.4545			0.8095	
	30	0.3846	0.1818			0.5565	
Kane Williamson	50	0.2885	0.1818			0.3036	
	70	0.0962	0.0909			0.0632	
	100						

NOM – Number of Matches Played

has nearly a 75% chance of remaining not out after scoring 10 runs and a 45.50% chance after reaching 30 and so on. Also, survival probabilities are higher against England compared to New Zealand. David Warner exhibits the highest initial survival chances against New Zealand, but unable to convert into big scores. Compared to other teams like England and Pakistan, Warner has a higher chance of reaching 50 against England, while his probability of achieving scores against Pakistan is lower. Jos Buttler is most comfortable in the initial stages of an innings against New Zealand compared to another opponent's team. He has an 83% chance of remaining not out after scoring 10 runs against New Zealand. Additionally, Buttler has the highest survival rates against Australia and India. When compared to other teams, his chances of reaching 50 runs are higher against New Zealand, but his probability of scoring a half-century against

Pakistan is lower than against other teams. Kane Williamson initially demonstrate the highest survival chances against India and lowest against Australia.

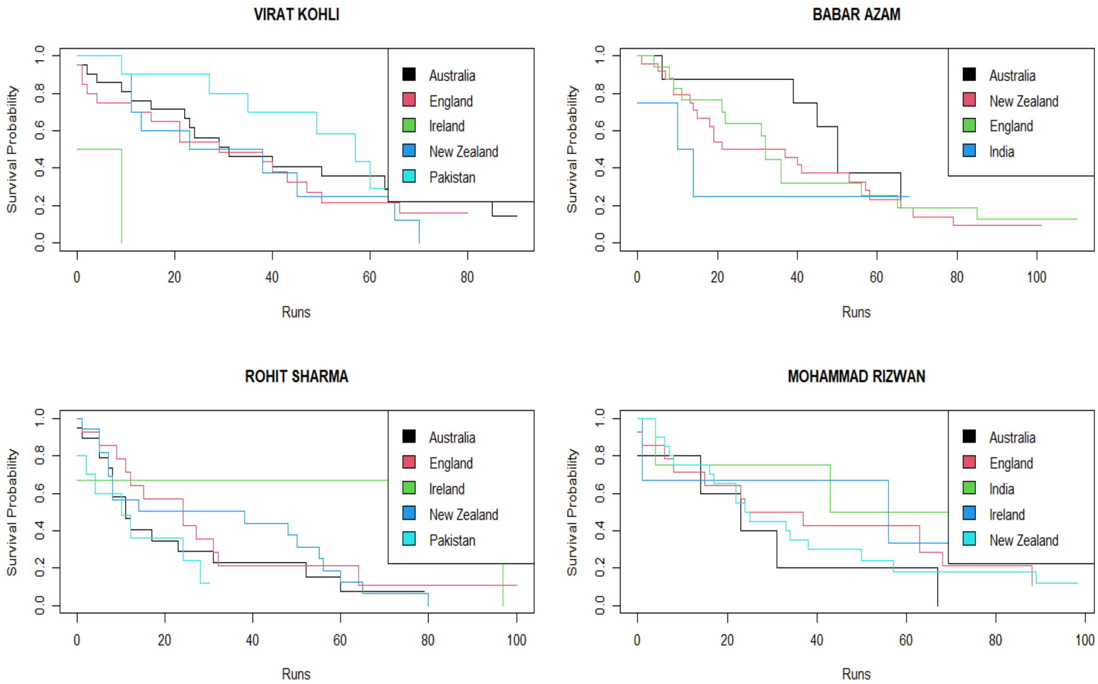

Figure 3. Survival Curves of each batsman against different Countries

From Figure 3, It is observed that Kohli initial survival probability is higher against all teams except Pakistan. His survival pattern is relatively consistent against all countries. Also, his survival chances at initial score and after that are highest against Australia and least against New Zealand. Babar has high probabilities of survival when England compared to New Zealand. Rohit has highest survival chance against England but least chance for getting century against Pakistan. Rizwan's survival probabilities are higher against England compared to New Zealand.

From table 7, This Conditional Survival Probability indicates that a specific batsman's chances of reaching ‘b’ runs while currently at ‘a’ runs reflect their performance ability in the ongoing match. We can see that when Kohli scores 10 runs, his probability of reaching a century is the highest among all players. This indicates that once Kohli finds his pattern, his chances of converting it into a big score are greatest. The chances of a batsman reaching a century after scoring 50 runs are high for Kohli, with Rizwan, Buttler, and Babar closely following him. Additionally, the probability of scoring 50 runs is high for Rohit. For hitting Century, Stirling has lower chances, while Williamson has no chance at all. This is also evident in the number of centuries these batsmen have achieved. Once they settle in at the crease, their scoring tends to be high. Low probabilities for a batsman indicate that they often lose their wicket after getting into a rhythm. This insight can help coaches encourage players to focus on longer innings. These probabilities will be particularly valuable for predicting individual and team scores, which is crucial for both the team and the betting industry.

Batsmen considered in this study are mostly top order batsmen. Semiparametric Cox-PH and Parametric models were used to study the survival pattern present with

Table 7. Conditional Probabilities P(Runs > b | Runs > a) of Batsmen

Players	b	a = 10	a = 30	a = 50	a = 70
	10	1			
	30	0.6596	1.0000		
Kohli	50	0.4663	0.7070	1.0000	
	70	0.3059	0.4638	0.4638	1.0000
	100	0.1963	0.2976	0.4209	0.6417
	10	1.0000			
	30	0.7077	1.0000		
Babar	50	0.4554	0.6435	1.0000	
	70	0.2201	0.3110	0.4832	1.0000
	100	0.0903	0.1276	0.1982	0.4103
	10	1.0000			
	30	0.5568	1.0000		
Rohit	50	0.3938	0.7071	1.0000	
	70	0.1951	0.3504	0.4956	1.0000
	100	0.0767	0.1377	0.1947	0.3929
	10	1.0000			
	30	0.4854	1.0000		
Stirling	50	0.2605	0.5366	1.0000	
	70	0.1554	0.3203	0.5968	1.0000
	100	0.0207	0.0427	0.0796	0.1333
	10	1.0000			
	30	0.6875	1.0000		
Riswan	50	0.4629	0.6733	1.0000	
	70	0.3141	0.4569	0.6786	1.0000
	100	0.1658	0.2412	0.3582	0.5278
	10	1.0000			
	30	0.5571	1.0000		
Warner	50	0.3990	0.7162	1.0000	
	70	0.2006	0.5028	0.5028	1.0000
	100	0.0669	0.1676	0.1676	0.3333
	10	1.0000			
	30	0.5666	1.0000		
Buttler	50	0.3808	0.6720	1.0000	
	70	0.1758	0.3103	0.4618	1.0000
	100	0.0754	0.1330	0.1979	0.4286
	10	1.0000			
	30	0.5871	1.0000		
Williamson	50	0.3359	0.5721	1.0000	
	70	0.1018	0.1734	0.3030	1.0000
	100	0.0000	0.0000	0.0000	0.0000

respect to "innings" (Batting First or Chasing) and their Positions (Top, Middle and Low order). Comparing all the batsmen simultaneously may not be justified due to high degree of heterogeneity present in the batting style and each batsman played in "top order" and "middle order" were compared. Batsmen who typically bat in the first three positions (1, 2, and 3) are designated as "top order" and bat in positions four through seven are classified as "middle order".

The hazard ratios between innings and positions were analyzed using the Cox Proportional Hazards model, with its assumptions evaluated through graphical methods and tests. While using Cox Proportional Hazard model for all batsmen score data with their positions, the following observations were made. The following Table 8 includes Hazard Ratio, Chi-square statistics and p-value for accessing Proportional Hazard (PH) assumptions for the variable Innings and Positions.

From the table 8, Kohli's batting corresponding to innings play a significant role in scoring runs at a 5% level of significance. The hazard ratio for Kohli's innings is 0.57, indicating that the risk of getting out while Chasing is 43% lower compared to

Table 8. Hazard Ratio and Chi-square Statistic for PH Assumption and its corresponding p-values for Covariates Innings and Position of each Batsman

Players	Variable	CoxPH HR	p-Value	PH assumption Chi-square	p-Value
	Innings	0.5675	0.0171*	6.6390	0.0100
Virat Kohli (89,20,0)	Middle	1.2341	0.4809	0.0290	0.8650
	Bottom				
	Innings	1.0515	0.8070	1.8300	0.1760
Babar Azam(110,2,0)	Middle	1.1609	0.8370	4.3500	0.0370
	Bottom				
	Innings	1.4470	0.0425*	0.2140	0.6440
Rohit Sharma (119,24,0)	Middle	1.3456	0.2866	3.6740	0.0550
	Bottom				
	Innings	0.9918	0.9640	0.0360	0.8500
Paul Stirling (141,0,0)	Middle				
	Bottom				
	Innings	1.3648	0.2138	0.0896	0.7650
Mohammad Rizwan (73,10,2)	Middle	2.3519	0.0425*	8.6003	0.0140
	Bottom	7.2795	0.0691		
	Innings	0.9334	0.7510	0.0746	0.7800
David Warner (99,4,0)	Middle	0.9612	0.9400	0.1293	0.7200
	Bottom				
	Innings	1.2921	0.2482	0.6190	0.4300
Jos Buttler (52,55,0)	Middle	2.2007	0.0008*	0.2330	0.6300
	Bottom				
	Innings	0.6582	0.0875	3.3000	0.0690
Kane Williamson (73,13,1)	Middle	1.6518	0.1364	2.7200	0.2570
	Bottom	11.7647	0.0220*		

(a,b,c): a,b and c denotes number of Matches Played as Opener, Middle Order and Low order respectively. *denotes 5%level of significance; **denotes 1% level of significance

Facing. This suggests that Kohli is likely to spend more time at the crease in the second innings than in the first. Similarly, for Rohit, it is found that innings also play a significant role in scoring runs at a 5% level of significance. However, the hazard ratio for Rohit's innings is 1.45; showing that the risk of getting out while Chasing is 1.45 times higher compared to facing other teams. For Mohammad Rizwan and Jos Buttler changing his position of batting order affects getting Runs. Their position of batting order increases then their chance for out also increases at 2.6 times and 2.2 times respectively.

Figure 4 shows, the PH assumption of the Models is verified graphically and using Schoenfeld residual test for Kohli.

Table 9. Hazard ratio between Players Scoring their runs

Players	Coefficient	HR	p-Value
Babar Azam	0.2339	1.2635	0.1235
Rohit Sharma	0.4941	1.6390	0.0006**
Paul Stirling	0.6294	1.8765	<0.0001**
Mohammad Rizwan	0.0549	1.0564	0.7434
David Warner	0.4275	1.5334	0.0055**
Jos Buttler	0.3746	1.4543	0.0168*
Kane Williamson	0.4394	1.5517	0.0066**

The hazard ratio in the table 9 represents the exponentiated coefficient, it denotes the increase or decrease in risk of an event given a unit change in time (in this case time is measures as progress in terms of runs scored) relative to the baseline, which is Kohil's

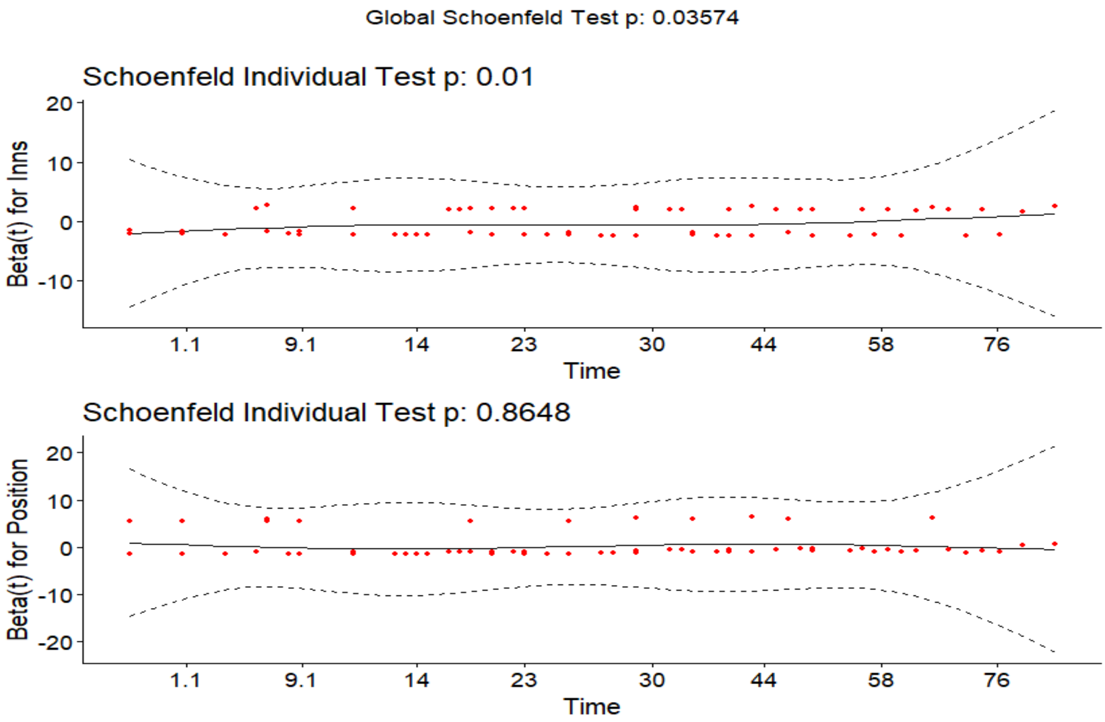

Figure 4. Graph for PH Assumption of Virat Kohli

innings. The hazard ratio for Rohit, Stirling, Williamson and Warner is significantly higher at 1.63, 1.87, 1.55 and 1.53 respectively, indicating a greater chance of getting out compared to Kohli at a 1% level of significance. Buttler also has a significant risk of getting out compared to Kohli at a 5% level of significance. In contrast, Babar and Rizwan are less likely to get out before scoring the next run compared to Kohli, with their results being statistically insignificant.

Four Parametric survival models were applied to the data, and their fit was assessed using graphical techniques and performance metrics such as AIC and BIC. The appropriate models for evaluating the time ratios for innings and positions were identified. Using corresponding distribution selected from performance measures, we obtained Time Ratio for Innings. In order to address the singularities at the score zero, a small positive value 0.001 is replace against Zero and this does not alter the estimates of model parameters considerably.

The table 10, provided shows the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for fitting four different distributions (Exponential, Weibull, Lognormal, Loglogistic) on the runs scored by World's top eight highest run scorer. Weibull is the best fit for runs scored by all Players except Kane Williamson. But Exponential distribution is the best fit for Kane Willamson. These are observed using AIC and BIC measures. By utilizing the appropriate distributions chosen from performance measures, we calculated the Time Ratio for the innings variable. The above table also indicates the significant difference of scoring pattern of Kohli and Rohit between Facing and Chasing innings. The same observations made through Non-Parametric and Semiparametric methods also.

Table 10. Fitting Four distributions on runs scored by Top Eight batsmen of World

Players	Measure	Exponential	Weibull	Lognormal	Loglogistic
	AIC	769.5125	**755.1848**	796.5756	769.6730
	BIC	774.8952	**763.2588**	804.6497	777.7471
Virat Kohli	Time Ratio	1.77	2.20	6.84	3.00
	p-Value	0.02*	0.02*	<0.001**	<0.01**
	AIC	927.9619	**917.2094**	989.7701	948.1618
	BIC	933.3989	**925.3649**	997.9255	956.3173
Babar Azam	Time Ratio	0.93	0.89	0.69	0.74
	p-Value	0.71	0.68	0.48	0.39
	AIC	1128.4919	**1080.6595**	1165.3457	1123.8865
	BIC	1134.4176	**1089.5481**	1174.2342	1132.7750
Rohit Sharma	Time Ratio	0.69	0.68	0.62	0.63
	p-Value	0.04*	0.19	0.38	0.22
	AIC	1133.3329	**1089.5447**	1181.1998	1139.6658
	BIC	1139.2305	**1098.3909**	1190.0461	1148.5121
Paul Stirling	Time Ratio	1.01	1.07	1.12	1.03
	p-Value	0.97	0.81	0.84	0.93
	AIC	646.5853	**636.8753**	679.6942	651.9681
	BIC	651.4706	**644.2032**	687.0221	659.2961
Mohammad Rizwan	Time Ratio	0.71	0.67	0.63	0.65
	p-Value	0.17	0.25	0.47	0.32
	AIC	834.8672	**817.6628**	878.7353	845.1923
	BIC	840.1367	**825.5669**	886.6395	853.0965
David Warner	Time Ratio	1.12	1.20	1.64	1.25
	p-Value	0.60	0.56	0.39	0.57
	AIC	789.4608	**777.1025**	837.5210	800.3496
	BIC	794.8064	**785.1210**	845.5395	808.3681
Jos Buttler	Time Ratio	0.88	0.89	0.72	0.84
	p-Value	0.57	0.71	0.57	0.64
	AIC	**688.1395**	689.2994	761.2072	717.9918
	BIC	**693.0714**	696.6971	768.6049	725.3895
Kane Williamson	Time Ratio	1.35	1.38	1.38	1.22
	p-Value	0.20	0.22	0.56	0.55

5. SUMMARY

Different survival techniques are used to predict the performance of top highest scorer in T20 matches. These survival estimates help to evaluate the batsman while standing on the crease against particular countries and considering his every runs. The performance of each batsman is evaluated based on their scoring runs at the crease, their ability to manage different positions, their effectiveness in either facing or chasing innings, and their run-scoring capabilities against different countries. These metrics assist decision-makers in determining the appropriate batting order for a player against specific opponents. In this study we observed that Kohli has higher survival rate compared to other top scorer, especially he differs from others scoring pattern while playing in Chasing. Application of survival methods outlined in this article can contribute to a better understanding of the risks and advantages of intervention in various circumstances.

FUTURE STUDY

The competing risk analysis can be used for the same study, while considering different terminal events of getting out of any player by Bowled, LBW, Runout, caught etc., and Recurrent event analysis also can apply for this study.

Acknowledgement

Authors would like to express their sincere thanks to the Editor in Chief, the Associate Editor and anonymous referees for their valuable comments that have helped to improved the quality of this paper.

References

[1] Parag Shah, R. D. Chaudhari, and Patel, M. N. (2023). Evaluating Batsman using Survival Analysis. *Statistics and Applications*, Volume 21, No. 1, 2023 (New Series), pp 51–62

[2] Preetham HK, P. K. Prajwal, R. Kumar, N. (2023). Cricket score prediction using machine learning. *International Journal of Innovative Research in Technology*, 9, no. 8, pp. 109–114.

[3] Mohan, N. Ramakrishnan, M. 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, no. 2, pp. 41– 51.

[4] Ramakrishnan, M. Ravanan, R. (2013). Non-Parametric estimations of career paths of special educators. *Frontiers of Statistics and its Applications*, *Bonfring.* (ISBN: 978-93-82338-78-9).

[5] Bracewell, P. J. and Ruggiero, K. (2009). A parametric control chart for monitoring individual batting performances in cricket. *Journal of Quantitative Analysis in Sports*, 5.

[6] Kimber, A. C. and Hansford, A. R. (1993). A statistical analysis of batting in cricket. *Journal of the Royal Statistical Society.* Series A (Statistics in Society), 156, 443–455.

[7] Saikai, H. and Bhattacharjee, D. (2018). Survival ability of Indian and overseas batsmen on the cricket pitch in Indian premier league. *MOJ Sports Medicine*, 2(4), 113-116.

[8] Swartz, T. B. (2017). Research directions in cricket. Handbook of Statistical Methods an analysis in sports. Chapman and Hall.

[9] Van Staden, P. J., Meiring, A. T., Steyn, J. A., et. al. (2010). Meaning batting averages in cricket. *South African Statistical Journal Proceedings*: Peer-reviewed Proceedings of the 52nd Annual Conference of the South African Statistical Association for 2010 (SASA 2010): Congress.

[10] Sachin S.S. Kottearachchi, Chathuri L. Jayasinghe, Rajitha, M. Silva (2022). An Investigation of Survival Abilities of Opening Batsmen in One-Day International Cricket. *Studies of Applied Economics, Volume 40-1*

[11] Brown, P. (2017). Optimising batting partnership strategy in the first innings of a limited overs cricket match. [dissertation]; Available from $https : //researcharchive.vuw.ac.nz/xmlui/bitstream/handle/10063/6871$
$/thesis_access.pdf?sequence = 1$.

[12] Kachoyan, B. & West, M. (2016). Cricket as life and death. *In Proceedings of the 13th Australian Conference on Mathematics and Computers in Sport.* (85-90). Melbourne, Victoria, Australia: ANZIAM MathSport.

[13] Kaplan, E. L and Meier, P. L. (1958). Non-Parametric estimation from incomplete observations. *Journal of the American Statistical Association.* 53, 457-81.

[14] Cox, D.R. (1972). Regression models and life Tables. *Journal of the Royal Statistical Society, Series B*, 34:187-202.

[15] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 50(3), 163-70. PMID: 5910392.

[16] NanamiTaketomi, et al., (2022). Parametric Distributions for Survival and Reliability Analyses, a Review and Historical Sketch. *Mathematics.* 10, 3907.

[17] ShaojunGuo and Donglin Zeng. (2013). An overview of Semi-Parametric models in survival analysis. *Journal of Statistical Planning and Inference.* $http : //dx.doi.org/10.1016/j.jspi.2013.10.008i$

[18] Mukesh Kumar, et al., (2019). Parametric survival analysis using R: Illustration with lung cancer data, *Wiley Cancer Reports*.