Statistics and Applications {ISSN 2454-7395 (online)}
Volume 24, No. 1, 2026 (New Series), pp 79–91
http://www.ssca.org.in/journal

Competing Risks Analysis of Factors Influencing the Runs Scored by Top T20 Batsmen - A Survival Analysis Approach

M. Sathishkumar[1], M. Ramakrishnan[1] and N. Viswanathan[2]

[1]*Department of Mathematics, RKM Vivekananda College, Chennai, India*
[2]*Department of Statistics, Presidency College, Chennai, India.*

Received: 22 August 2024; Revised: 07 June 2025; Accepted: 09 June 2025

Abstract

$T20$ cricket is an exciting format characterized by explosive hitting and strategic play, engaging fans with each boundary and a wicket. A comprehensive dataset of T20 matches is analyzed to understand the factors affecting batsmen's performance in this highly dynamic format. Survival analysis approach is used to study the performance of the batsmen, measured in terms of 'number of runs' taken as the 'innings survival time'. In this context, dismissal of batsmen is taken as the 'event'. The dismissal may be due to getting Bowled, being Caught, LBW, Run out, Stumped or Hit wicket. These different forms of dismissal can be taken as 'competing risks' and this study specifically focuses on identifying factors associated with specific dismissals. In this process, Cumulative Incidence Function (CIF), Cause-Specific Hazard function (CSH), and Fine and Gray's Subdistribution Hazard function (SDH) were used. The results from this analysis offer insights into the game dynamics and aids in player's performance evaluation and strategic decision-making, such as, team composition, batting order and making the choice of 'batting first' or 'chasing'. Data for representative batsmen from the top ICC ranking of T20 game with specific inclusion and exclusion criteria was taken from www.espncricinfo.com as on 28$^{\text{th}}$ of April, 2025. The analysis was carried out using R Programming Language (R 4.5.0) with suitable packages.

Key words: Survival analysis; Competing risks; Cumulative incidence function; Cause-specific hazard; Subdistribution hazard.

AMS Subject Classifications: 62N01, 62N02

The video recording of the paper made under the SSCA's Online Lecture series is available at the Youtube channel URL `https://youtu.be/E4rnAoT1f0g`.

1. Introduction

Cricket is becoming one of the most popular sports of the today's world. International cricket games are categorized as Test Cricket, One-day International (ODI) and Twenty20

Corresponding Author: M. Ramakrishnan
Email: mramkey@rkmvc.ac.in

(T20). One-day cricket was introduced in 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 in every four years when the Men's Cricket World Cup is organized by the International Cricket Council (ICC) which is the global governing body for cricket games. Later in 2003, T20 form of the game was introduced with focus on gaining wider audience and with emphasis on power hitting. Cricket in T20 format is limited to twenty overs for each team. The current study focusses on the game of T20's. Batting is the heart of cricket and bowling is its backbone. Cricket knowledge tells us that batting is more difficult early in a player's innings but becomes easier as players familiarize themselves with the pitch conditions.

Survival Analysis is defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of a particular event of interest. In Competing Risks model, the subject is exposed to more than one possible event of interest, but only one event will occur at any given time. In cricket, a player's dismissal is taken as the event of interest and the different ways in which the dismissal occurs, namely, Bowled, Caught, LBW, Runout and Stumped can be considered as competing risks for the event 'dismissal'. Of these competing risks, one can be taken as the main event and the rest as competing events that prevent the observation of the main outcome or change the probability of its occurrence. The added feature of competing risks data is the presence of failure types, in addition to failure time or survival time.

The developments in the field of survival analysis had the most profound impact on clinical trials are the Kaplan and Meier (1958), for estimating the survival function, the Log-rank statistics by Mantel (1966) for comparing two survival distributions, and the Cox (1972) proportional hazards model for quantifying the effects of covariates on the survival time. The Cox regression model can be used to identify the variables that significantly affect the outcome of interest and present the results in terms of the hazard ratio.

Staden *et al.* (2010) developed alternative batting average measures to address issues with traditional averages, particularly when a batsman ends up with 'not out'. In most of the previous studies such as Kimber and Hansford (1993), Kachoyan and West (2016), Brown (2017) and Saikai and Bhattacharjee (2018) survival abilities of individual batsman were analyzed based on the survival function of the number of balls faced till dismissal. In this study, instead of considering the 'number of balls faced before dismissal', the 'number of runs scored before dismissal' is taken as the life span. Kachoyan and West (2016) described batsman's innings as a lifespan, in that, "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's 'death' which is the event of interest. When a batsman was not dismissed during a match, the particular observation was referred to as a censored observation. In this study, instead of considering the 'number of balls before he is dismissed', the 'number of runs scored before he is dismissed' is taken as the life span.

Survival analysis using competing risks data is widely used in medical research and is gaining increasing attention and interest in a variety of research fields. Cumulative Incidence Function (CIF) represents the cumulative probability of an event due to a particular type of cause over time and it is a useful metric for analysing competing risks data (Pintilie (2006)). By treating the Cumulative Incidence Curve (CIC) as a subdistribution hazard Function, Gray and Jason (1999) provided a methodology for computing the CIF and comparing it across different categories of a variable. Sapir-Pichhadze *et al.* (2016) observed that the Cause-specific and subdistribution hazard models provide complementary information regarding the relationship between exposures and outcomes of interest in the presence of competing events.

Shah *et al.* (2023) studied the survival probabilities of the top-10 ODI batsmen around the world and suggested that it can be used as a new measure for evaluating batsman's ability to survive on crease. Kottearachchi *et al.* (2022) studied the survival abilities of the opening batsmen in one-day international cricket. Preetham *et al.* (2023) suggested a model for predicting the outcome of the IPL matches, in particular, to forecast the score of an innings using machine learning models. Saikai and Bhattacharjee (2018) examined survival ability of batsmen in Indian Premiere League (IPL) 2012. Ramakrishnan *et al.* (2023) studied the performance of Indian Batsmen in the 2023 World Cup Squad using survival analysis approach.

In this paper, in order to identify the factors influencing the survival time of a batsman, measured in terms of the number of runs scored, the Cox Proportional Hazard model under 'single event' and 'competing events' are considered. For competing risks model, both 'Cause-specific' and 'Subdistribution' approaches are employed. Representative batsmen from Top 10 T20 batsmen, as per the ICC rankings, are selected with specific inclusion and exclusion criteria described in the 'Data structure' section. The data and the rankings pertain as on 28$^{\text{th}}$ April, 2025 taken from www.espncricinfo.com. The covariates of the models include the order of batting, the position in which the player is slotted for batting, Match venue, the time of the game, Toss Result, the stage of the match and the Tournament Type. The results from the study clearly point to using the efficacy of competing risks models in identifying the significant factors and the patterns present in them.

2. Methodology

The Cox PH model is used to identify significant factors that influence the runs scored by the batsmen. In this paper, the general Cox PH model and the one pertaining to 'competing risks' are used. Under the considered competing risks models, both Cause-specific and Subdistribution Hazard approaches are used.

2.1. Survival functions and methods

The survival function is of at most 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).

The hazard function $h(t)$ gives the instantaneous potential per unit time for the event to occur, given that the individual has survived up to time t.

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

The hazard function and the survival function have a straightforward one-to-one connection.

2.2. Cox proportional hazard model

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

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

where the survival time is denoted by t, the p covariates are denoted as $(X_1, X_2, ..., X_p)$. The coefficients $\beta = (\beta_1, \beta_2, ..., \beta_p)$ quantify the effect of these covariates on the hazard function. Additionally, the term $h_0(t)$ signifies the baseline hazard, serving as a reference line for understanding how the covariates modify the risk of the event occurring over time. This model of the hazard function is used to analyze the survival data and identify the significant covariates influencing the hazard of the event under consideration. The proportional hazards assumption requires that covariates are multiplicatively related to the hazard. This study employs Schoenfeld Residual test to verify the proportional hazard assumption.

2.3. Competing risk models

Competing risk models, in this section, predominantly use Cause-specific and Sub-distribution hazard models. These approaches employ cumulative incidence function for studying the pattern of the risks, which in turn extracts the significant factors, if present in the model.

2.3.1. Cumulative Incidence Function (CIF)

In situations where competing risks are involved, the survival curve approach by the Kaplan-Meier method may not be fully reliable due to the violation of the independence assumption regarding the competing risks. This paved the way to introduce new approaches, one of which is the Cumulative Incidence Function that uses marginal probabilities.

The Cumulative Incidence Function (CIF) for event type c at time t_j which is calculated as the cumulative sum upto time t_j of the incidence probabilities over all event type c failure times

$$CIC_c\left(t_j\right) = \sum_{i=1}^{j} I_c\left(t_i\right) \; = \sum_{i=1}^{j} \widehat{S}(t_{i-1})\widehat{h}_c(t_i)$$

where $\widehat{h}_c\left(t_j\right) = \frac{m_{cj}}{n_j}$ represents the ratio of number of events for type c that occur at t_j to the number at risk at t_j. Here, $S(t_{j-1})$ is the surviving probability of the prior time t_{j-1}, where *S(t)* represents the general survival curve instead of the Cause-specific survival curve S_c(t). Here, $I_c\left(t_i\right)$ is the incidence function for event type c at time t_i.

2.3.2. Cause-Specific Hazard (CSH) model

The Cause-Specific Hazard (CSH) Model that makes use of the Cox PH method to individually evaluate hazards for every type of failure, treating competing events in the form of censors as well as the others who are censored due to follow-up loss or due to pulling out of the study.

The Cause-specific hazard for event type c under Cox PH model with covariates $X = (X_1,\ X_2,\ ...,\ X_p)$ is defined as

$$h_c(t, X) = h_{0c}(t)e^{\sum_{i=1}^{p}\beta_{ic}X_i}$$

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

and T_c is the random variable that represents the time to failure for events of type c with $c = 1,\ 2,\ 3,\ \ldots,\ K$, β_{ic} represents the regression coefficient of X_i to the event type c and $h_{0c}(t)$ is the baseline hazard for event type c.

2.3.3. Subdistribution Hazard (SDH) function

Gray and Jason (1999) provides a regression model that applies SDH and can directly be used in CIF for computation under competing risks analysis. Under this model, In addition to the cause c's hazard, this CIF for cause c depends on hazards of all other causes, as well. For this approach, the SDH is also defined as

$$h_c^*(t; X) = \lim_{\Delta t \to 0} \left\{ \frac{P\left[t \leq T < t + \Delta t, D = c | T \geq t \cup (T < t \cap D \neq c)\right]}{\Delta t} \right\}$$

As a result, the covariate effect is proportional to the CIF. Proportional hazards assumption was imposed by Fine-Gray on the SDHs and provided estimators and their large sample properties. The SHD approach assumes that the occurrence of a competing event does not influence the rate of the event of interest. The partial likelihood estimation used in the standard Cox model is followed in the CSH and SDH models to estimate the covariate coefficients. However, the difference between CSH and SDH exists only in the risk set. The risk set for the CSH, decreases each time an event of another cause occurs. But these observations remain in the risk set for the SDH.

3. Data structure

The Data on the T20 matches for selected batsmen were taken from `www.espncricinfo.com` as on 28th April, 2025. The batsmen with top 10 ICC rankings, with the condition that they should have played at least 50 matches were selected. These 10 batsmen are Suryakumar Yadav (India), Jos Buttler (England), Pathum Nissanka (Sri Lanka), Tim Seifert (New Zealand), Babar Azam (Pakistan), Kusal Perera (Sri Lanka), Reeza Hendricks

(South Africa), Mohammad Rizwan (Pakistan), Finn Allen (New Zealand) and Kusal Mendis (Sri Lanka). Of these 10 batsmen, three were selected as a fair representation from this conditional top 10 batsmen. They constitute Jos Buttler (Rank 2), Babar Azam (Rank 5) and Mohammad Rizwan (Rank 8). The selection was made to avoid the extreme ranks, namely Rank 1 and Rank 10.

For the purpose of using competing risks model, runs scored by batsmen are considered as their survival time, represented by the variable Runs. The type of dismissals of a batsman is represented by the variable Dismissal. The dichotomous variable Dismissal_2 takes values '0' and '1', with '1' representing a dismissal, irrespective of its type such as Bowled, Caught or LBW and '0' representing the 'not out' status of the batsman. The covariates included in the Cox PH competing risk models include 'Innings' (Batting First or Chasing) represented by the variable Innings, 'Position' (Top, Middle, Low) represented by the variable Position, 'Match Place' (Home Venue, Home of Opposition, Neutral Venue) represented by the variable Venue, 'Tournament Type' (2 Team Series, 3-4 Teams, 5 or more Teams) represented by the variable Type, 'Time of the match' (Day, Day/Night, Night) represented by the variable Time and 'Toss Result' (Won, Lost) represented the variable Toss. All the included covariates are categorical in nature.

The first few rows of data for the batsman Babar Azam are presented in Table 1.

Table 1: Data for batsman Babar Azam representing first 10 of his matches

S. No	Runs	Dismissal	Innings	Venue	Type	Time	Toss	$Dismissal_2$	Position
1	86	caught	Batting First	Home Venue	2 Team Series	Night	Lost	1	Top
2	45	caught	Batting First	Home Venue	2 Team Series	Night	Won	1	Top
3	48	caught	Batting First	Home Venue	2 Team Series	Night	Lost	1	Top
4	34	not out	Batting First	Home Venue	2 Team Series	Night	Lost	0	Top
5	17	lbw	Batting First	Home Venue	2 Team Series	Day	Lost	1	Top
6	97	not out	Batting First	Home Venue	2 Team Series	Day	Won	0	Top
7	51	caught	Chasing	Home Venue	2 Team Series	Day	Lost	1	Top
8	13	caught	Chasing	Home Venue	2 Team Series	Night	Won	1	Top
9	3	bowled	Chasing	Home Venue	2 Team Series	Night	Lost	1	Top
10	27	caught	Chasing	Home Venue	2 Team Series	Night	Lost	1	Top

4. Empirical analysis

An overview of the details of the selected batsmen regarding their matches is presented in the Table 2.

Table 2: Descriptive statistics of the selected batsmen

Batsman	No. of Matches	Average Run	# Not Out	# Bowled	# Caught	# LBW	# Run Out	# Stumped	# Dismissal
Jos Buttler	123	28.74	23	14	72	7	5	2	100
Babar Azam	121	34.90	15	15	74	8	6	3	106
Rizwan	93	36.71	21	14	44	6	5	3	72

Table 2 indicates that out of 123 matches played by Jos Buttler, he reminded 'not out' in 23 matches and dismissed by 'Caught' in 72 matches followed by 'Bowled' in 14 matches. Similar information for the other two batsmen is also provided in Table 2. From this Table, it is inferred that the dismissal occurs mostly by way of Caught, followed by

Bowled. The other three categories, namely LBW, Runout and Stumped are relatively low in number. The low frequencies in these categories make us take Caught and Bowled as the two competing risks and drop the rest from the competing risks analysis.

A Cox PH model (Model 1) with 'Runs' as survival time and all types of dismissals taken as the event of interest is developed with covariates Innings, Position, Venue, Tournament Type, Time and Toss for all three batsmen separately. The results are consolidated in Table 3.

Table 3: Results of Cox PH model with all types of dismissal considered under one category -Model 1

		Jos Buttler		Babar Azam		Mohammad Riswan	
Covariates	Level	Hazard Ratio	p-value	Hazard Ratio	p-value	Hazard Ratio	p-value
Innings							
	Chasing	1.016	0.952	1.144	0.524	1.311	0.287
Position							
	Middle	1.709	**0.021***	1.083	0.920	2.802	**0.022***
	Bottom					6.275	0.093
Venue							
	Home	1.039	0.883	1.347	0.281	1.575	0.202
	Neutral	1.358	0.590	1.166	0.680	0.498	0.377
Tournament Type							
	Tournament3-4	2.382	0.132	0.749	0.596	2.288	0.195
	Tournament5+	0.595	0.341	1.435	0.319	2.659	0.211
Time							
	Day/Night	1.049	0.864	1.234	0.697	3.562	0.065
	Night	0.835	0.496	1.187	0.480	1.558	0.152
Toss							
	Lost	0.589	**0.050***	1.152	0.511	0.865	0.578

*denotes significance at 5% level

The results of Model 1 presented in Table 3 indicates that for the batsman Jos Buttler, Position and Toss turn out to be statistically significant at 5% level. For batsman Mohammad Rizwan the only covariate that is statistically significant is Position. For batsman Babar Azam, it is observed that none of the covariates is statistically significant. Going by the estimated Hazard Ratio, it is seen that for the batsman Jos Buttler, the hazard of getting out while playing in the middle-order is 1.71 times higher compared to playing in the top-order. It is further observed that, for batsman Jos Buttler, the hazard of getting out when his team loses the toss is 0.41 times lesser compared to when his team wins the toss. For the batsman Mohammad Rizwan, the hazard of getting out while playing in the middle-order is 2.8 times compared to playing in the top-order.

4.1. Schoenfeld residual test for PH assumption

The validity of estimation of Model 1 depends on how it satisfies the Proportional Hazards assumption for all the covariates under consideration. Schoenfeld residual test is carried out for this purpose and the results are presented in Table 4. The p-values from Table 4 indicates that the PH assumption is satisfied for all covariates for Jos Buttler and all but the covariate 'Position' for Babar Azam and Mohammad Rizwan at 5% level. Thus, it is seen that in most of the cases the PH assumption is well satisfied for Model 1 and estimates derived are stable and valid.

Table 4: Schoenfeld residual test results

	Jos Buttler		Babar Azam		Mohammad Riswan	
Covariates	Chi-square	p-value	Chi-square	p-value	Chi-square	p-value
Innings	1.786	0.180	3.606	0.058	0.042	0.837
Position	0.528	0.470	4.229	0.040	8.264	0.016
Venue	2.081	0.350	5.074	0.079	1.646	0.439
Tournament Type	0.976	0.610	0.104	0.949	1.124	0.570
Time	0.511	0.770	2.139	0.343	0.379	0.828
Toss	0.302	0.580	0.196	0.658	1.174	0.279

4.2. Cause-specific hazard model for 'Bowled'

The Cause-specific Hazard Model (Model 2) for the event 'Bowled' with 'Caught' as competing risks is developed and the significance of the selected covariates in terms of their p-values are presented in Table 5. This Table indicates that for batsman Jos Buttler, the Position and Venue are statistically significant at 10% level and their hazard ratios imply that the hazard of getting out while playing in the middle-order is 2.879 times higher compared to playing in the top-order. Further, hazard of getting out while playing in the Home town is 69% less compared to playing in the venue of the opposition team.

Table 5: Cause-specific hazard model for 'Bowled'

		Jos Buttler		Babar Azam		Mohammad Riswan	
Covariates	Level	Hazard Ratio	p-value	Hazard Ratio	p-value	Hazard Ratio	p-value
Innings							
	Chasing	0.638	0.483	0.852	0.769	1.902	0.263
Position							
	Middle	2.879	0.070	3.921	0.231	2.040	0.373
Venue							
	Home	0.309	0.091	0.901	0.874	0.958	0.954
	Neutral	0.546	0.343	1.000	1.000	1.009	0.990
Toss							
	Lost	0.646	0.498	1.058	0.917	1.392	0.571

4.3. Cause-specific hazard model for 'Caught'

The Cause-specific Hazard Model (Model 3) for the event 'Caught' with 'Bowled' as competing risks is developed and the significance of the selected covariates in terms of their p-values are presented in Table 6. This Table indicates that only for the batsman Jos Buttler, the Venue is statistically significant at 10% level and its hazard ratio implies that the hazard of getting out when the team loses the toss is 46% less compared to when the team wins the toss.

4.4. Schoenfeld residual test for PH assumption for Model 2 and Model 3

Schoenfeld residual test is carried out for Model 2 and Model 3 and the results are presented in Table 7. The p-values from Table 7 indicates that the PH assumption is satisfied for all covariates for the three batsmen for Model 2. In Model 3, for Babar Azam and

Table 6: Cause-specific hazard model for 'Caught'

		Jos Buttler		Babar Azam		Mohammad Riswan	
Covariates	Level	Hazard Ratio	p-value	Hazard Ratio	p-value	Hazard Ratio	p-value
Innings							
	Chasing	1.079	0.809	1.178	0.496	1.050	0.876
Position							
	Middle	1.556	0.102	0.689	0.717	2.289	0.140
Venue							
	Home	1.289	0.386	1.539	0.154	1.534	0.337
	Neutral	0.793	0.503	1.575	0.169	1.543	0.356
Toss							
	Lost	0.562	0.067	1.243	0.383	0.815	0.526

Mohammad Rizwan, all the covariates satisfy the PH assumption. For the batsman Bulter, the covariate Innings and Toss does not satisfy the PH assumption. The significance of the PH assumptions is tested at 5% level.

Table 7: p-values of Schoenfeld residual test for PH assumption for Model 2 and Model 3

	Bowled			Caught		
Covariates	Butler	Babar	Riswan	Butler	Babar	Riswan
Innings	0.602	0.17	0.35	0.005	0.17	0.455
Position	0.781	0.077	0.61	0.308	0.15	0.147
Toss	0.056	0.53	0.99	0.021	0.84	0.072
Venue	0.959	0.603	0.38	0.135	0.17	0.095
GLOBAL	0.456	0.286	0.69	0.048	0.15	0.112

The validity of PH assumption for batsman Jos Buttler in Model 2 is presented graphically in Figure 1. From this Figure, it is seen that for all the covariates under consideration, the Schoenfeld residuals mostly fall within the estimated boundaries, indicating that the PH assumption is valid for all the covariates in Model 2. Though similar Figures for the other two batsmen were verified, not presented in this text.

4.5. Subdistribution hazard model for 'Bowled'

Similar to Cause-specific Hazard model (Model 4), the subdistribution Hazard model is run for the event 'Bowled' and the associated hazard ratios and p-values for all the three batsmen are presented in Table 8. This Table indicates that only for the batsman Jos Buttler, the Venue is statistically significant at 10% level and its hazard ratio implies the hazard of getting out while playing in the Home Town is 73% less compared to playing in the Opposition Venue.

4.6. Subdistribution hazard model for 'Caught'

Similar to Subdistribution Hazard model (Model 5), the subdistribution Hazard model is run for the event 'Caught' and the associated hazard ratios and p-values for all the three batsmen are presented in Table 9. This Table indicates that only for the batsman

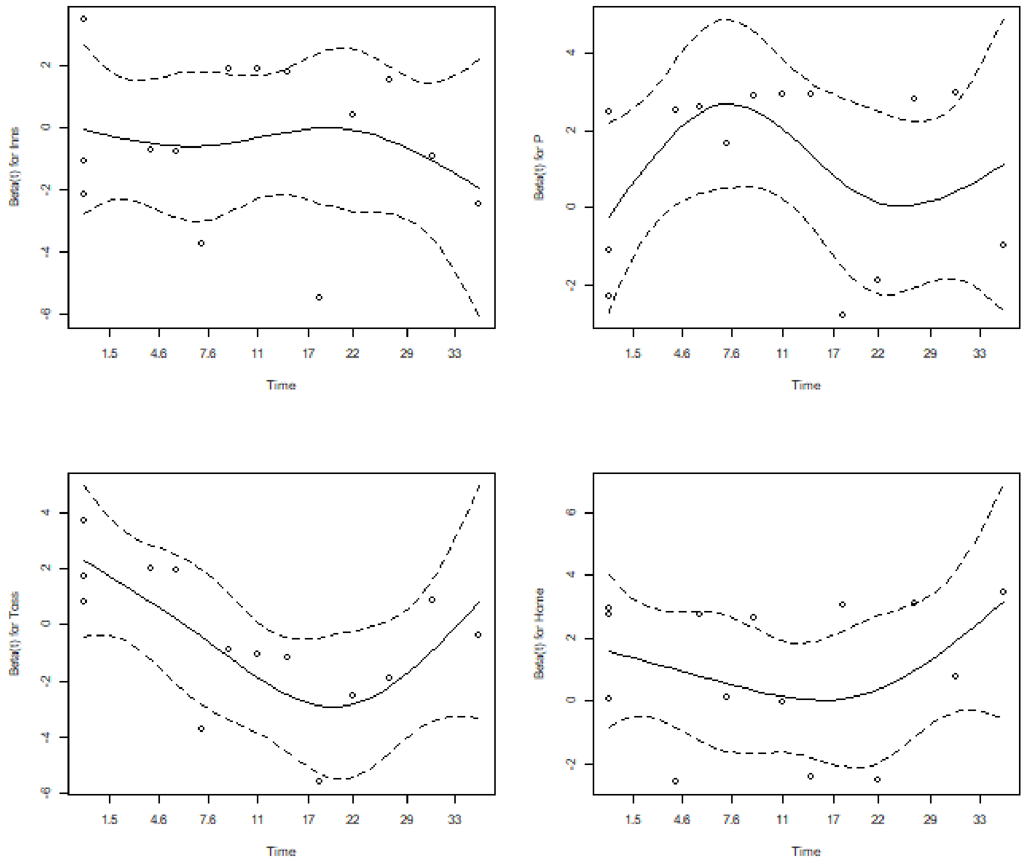

Figure 1: Pictorial representation of PH assumption verification for batsman Jos Buttler in Model 2

Table 8: Subdistribution hazard model for 'Bowled'

		Jos Buttler		Babar Azam		Mohammad Riswan	
Covariates	Level	Hazard Ratio	p-value	Hazard Ratio	p-value	Hazard Ratio	p-value
Innings							
	Chasing	0.647	0.490	0.744	0.610	1.874	0.270
Position							
	Middle	2.507	0.120	3.947	0.170	1.500	0.590
Venue							
	Home	0.276	0.062	0.754	0.670	0.875	0.880
	Neutral	0.570	0.380	0.817	0.760	0.991	0.990
Toss							
	Lost	0.916	0.880	0.921	0.890	1.589	0.480

Jos Buttler, the Venue is statistically significant at 10% level and its hazard ratio implies that the hazard of getting out while playing in the Home Town is 1.63 times compared to playing in the Opposition Venue.

The Schoenfeld residuals for the Subdistributional hazard models for 'Bowled' and

Table 9: Subdistribution hazard model for 'Caught'

		Jos Buttler		Babar Azam		Mohammad Riswan	
Covariates	Level	Hazard Ratio	p-value	Hazard Ratio	p-value	Hazard Ratio	p-value
Innings							
	Chasing	1.085	0.800	1.175	0.480	0.880	0.680
Position							
	Middle	1.026	0.920	0.484	0.450	1.304	0.640
Venue							
	Home	1.630	0.074	1.408	0.260	1.564	0.250
	Neutral	0.962	0.910	1.348	0.360	1.368	0.450
Toss							
	Lost	0.639	0.180	1.091	0.720	0.731	0.310

'Caught' satisfies the proportional hazard assumptions and as such the results from the above models can be taken as valid and stable. The Proportionality assumption for the competing risks regression log(-log(1-F)) can be plotted against log(Runs), where F is the CIF for the event of interest. The Figure 2 illustrates such a plot for the covariates in the batsman Jos Buttler for 'Bowled out'.

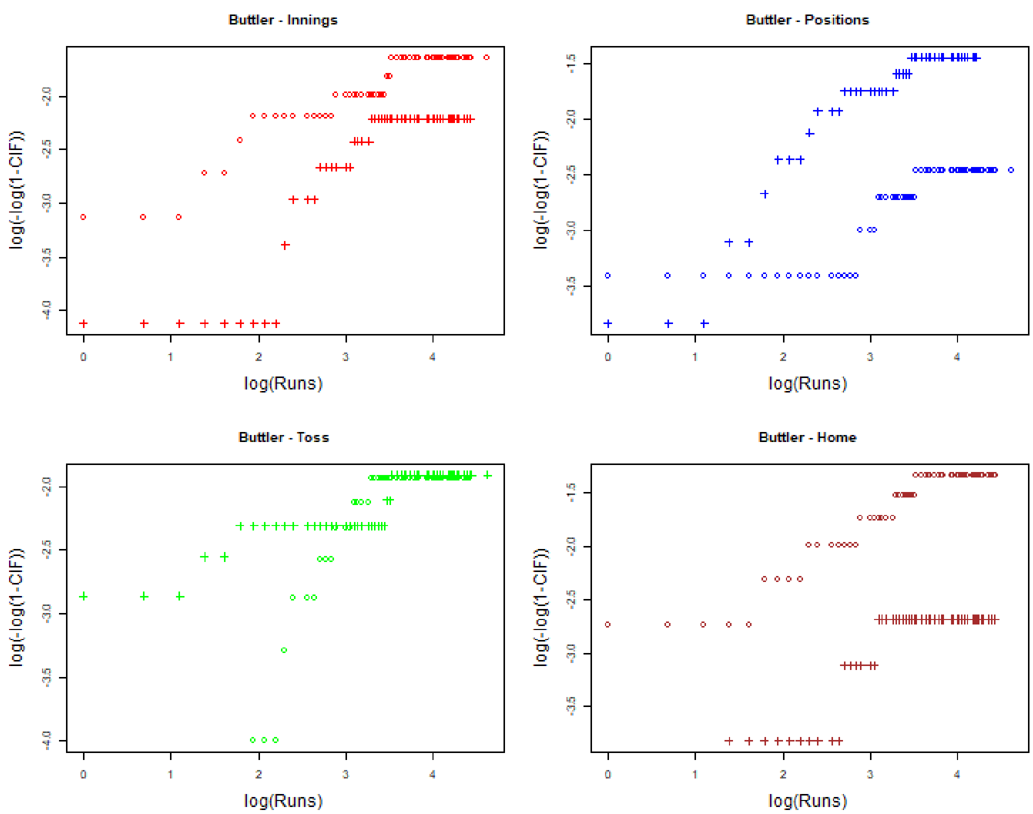

Figure 2: Proportionality of the hazard for all covariates in Model 4 for batsman Jos Buttler being bowled out

5. Conclusion

The runs scored by a batsman is taken as his 'Innings survival time' with covariates order of batting, the position in which the player is slotted for batting, Match venue, the time of the game, Toss Result, the stage of the match and the Tournament Type are considered under competing risk models. A comparison is carried out considering the usual Cox PH model with that of Cause-specific Proportional Hazards model and Subdistributional Proportional Hazards model. The Validity of the estimates derived under each model is verified by their respective Schoenfeld residuals. The General Cox PH model indicates Position and Toss as significant variable, whereas in the Cause-specific Models for the event 'bowled', Position and Venue are significant covariates. In the Cause-specific Model for the event 'caught' toss is a significant covariate. Under Subdistributional model for 'Bowled' and 'Caught', 'Venue' is statistically significant.

On comparing the significant covariates under different models, it is observed that if competing risks are not taken into account certain covariates tend to appear significant. But, when the competing risks are properly accounted for, the real scenario emerges and all the evidences point to the significant covariates 'Position', 'Toss' and 'Venue'. Thus, it is seen that the factors that influence the runs-scored by a batsman are the 'Position' in which he is slotted to play in the game, the result of the Toss and the venue in which the match is played. In particular, batsmen who play in the top order score considerably more runs compared to those who bat in the middle order slots. Further, playing the home venue decreases the hazard of getting out when the event is 'bowled' and the same increases the hazard of getting out when the event considered is 'caught'. Further, it is seen that the covariates 'Innings', 'Match Time', and 'Tournament Type' have no say in the run-scoring pattern of a batsman. This provides useful information about selection of batsmen for a particular venue and that of his batting-order in the game, which in turn increase the team's chances of winning the match.

6. Limitations of the study and future work

This study, though includes a number of covariates, some of them are dropped from the model, mainly because they have a highly imbalanced distribution in their levels. Future studies can include more batsmen in the analysis and in some of them, the distribution may be even and this will help to study the effect of the covariates that are dropped in this paper. Further, additional covariates such as the level of the match, indicating whether it is at league stage or knockout stage could also be included in the study and this may throw more light on the pressure exerted on the batsmen at different stages of the tournament.

Acknowledgements

We are very much grateful to the editor for providing consistent support to bring out tbe article in a compact, complete and readable format. We further express our high degree of appreciation to the reviewer, who had gone through the crude version of this article and helped us to transform it to a more refined structure in content and presentation, by providing finer details for great improvement in its current version.

Conflict of interest

I hereby declare that, to the best of my knowledge, I have no actual, potential, or perceived conflicts of interest related to this article. I understand the importance of objectivity and integrity in my work and will strive to fulfill my duties with the highest standards of professionalism.

References

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`.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, **34**, 187–202.

Gray, R. J. and Jason, P. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 469–509.

Kachoyan, B. and West, M. (2016). Cricket as life and death. in proceedings of the 13th australian conference on mathematics and computers in sport. *ANZIAM MathSport*, Melbourne, Victoria, Australia, 85–90.

Kaplan, E. L. and Meier, P. L. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53**, 457–481.

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.

Kottearachchi, S. S. S., Jayasinghe, C. L., and Silva, R. M. (2022). An investigation of survival abilities of opening batsmen in one-day international cricket. *Studies of Applied Economics*, **40**.

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy Reports*, **50**, 163–170.

Pintilie, M. (2006). *Competing Risks, A Practical Perspective*. John Wiley & Sons Ltd. Chichester, West Sussex, England.

Preetham, HK., Prajwal, R., Prince Kumar., and Naveen Kumar. (2023). Cricket score prediction using machine learning. *International Journal of Innovative Research in Technology*, **9**, 109–114.

Ramakrishnan, M., Viswanathan, N., and Ravanan, R. (2023). Performance of Indian batsmen in the 2023 world cup squad – a survival analysis approach. *Statistica*, **83**, 271–290.

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**, 113–116.

Sapir-Pichhadze, R., Pintilie, M., Tinckam, K. J., Laupacis, A., Logan, A. G., Beyene, J., and Kim, S. J. (2016). Survival analysis in the presence of competing risks: The example of waitlisted kidney transplant candidates. *American Journal of Transplantation*, **16**, 1958–1966.

Shah, P., Chaudhari, R. D., and Patel, M. N. (2023). Evaluating batsman using survival analysis. *Statistics and Applications*, **21**, 51–62.

Staden, P. J. V., Meiring, A. T., Steyn, J. A., and Fabris-Rotelli. (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.*