

cricWAR: A reproducible system for evaluating player performance in limited-overs cricket

Hassan Rafique
University of Indianapolis

Abstract

[Cricket statistics](#) mainly comprise simple averages (batting avg., bowling avg., strike rates, etc.) to account for player performance. For limited-overs cricket, these statistics fall short of providing a comprehensive picture of a player's performance and contribution to the team's success. Due to the rise of [T20 cricket leagues](#), there is significant interest in comprehensive statistics that capture the net value-added by an individual player. Inspired by sabermetrics [1], we develop metrics such as run value, runs above average (RAA), value over replacement player (VORP), and wins above replacement (WAR) for batters and bowlers in limited-overs cricket. These metrics are calculated using ball-by-ball data readily available through R package [cricketdata](#), co-developed by us. We associate run value with each actual run scored using estimated expected runs. Some positions/situations are more conducive for scoring (or defending) during the innings, and we adjust the run values by using a variation of the Leverage Index [6]. Run values are adjusted for the venue, bowling pace (spin vs. pace), platoon advantage, and innings (first vs. second) using regression to estimate RAA. We assess the uncertainty in RAA and WAR estimates through a resampling and simulation-based approach and present results for the IPL 2019 regular season. Finally, we discuss the reliability of these metrics, further avenues of research and comment on the possible implications of this work for the T20 teams.

1. Introduction and Motivation

Cricket is typically regarded as a [major world sport](#), considering different measures such as global fanbase (2.5 billion), digital viewership, TV rights deals, etc. Most of the fanbase resides in South Asia, the UK, Australia, South Africa, and New Zealand. Cricket as a sport is best described as a sibling of baseball; learn more about the [similarities here](#). Cricket is significantly growing due to the popularity of the new format of T20 and the money this format has brought into the game. Indian Premier League (IPL), a T20 league, sold its [streaming and TV rights](#) for the next five years for 6 billion USD. In the recent mega auction, the [IPL teams spent 116 million USD](#) on acquiring 234 players for three-month league season. However, the public statistical analysis of the sport and use of analytics in cricket appear to lag behind the other major sports. Baseball has a long history of public sabermetrics research work, and the last decade has seen an explosion of public sports analytics work in basketball, ice hockey, soccer, and American football. On the other hand, cricket is in the nascent stages of the sports analytics era, and barring a few [private T20 league teams](#) and one [national team](#), the use of analytics seems very limited.

One of the most fundamental focuses in team sports is to quantify the role played by individual players in team success. Since the goal is to win in sports, it is natural to be interested in estimating the number of wins a player contributes to their team. In cricket, like baseball, the role of players is clearly defined, and the actions are discrete, e.g., batters and bowlers (pitchers). Most of the [cricket](#)

[statistics](#) evaluating player performances are simple averages or proportions, e.g., batting avg = runs scored/ no. of innings batter got out, batting strike rate = $100 * \text{runs scored} / \text{balls faced}$, Boundary percent = percentage of balls faced hit for a boundary (4 or 6), bowling avg = runs conceded/ no. of wickets taken, etc.

These stats may provide reasonable estimates for the players' individual performances in batting/bowling and quality. However, they do not tell us about the contribution of the player's performance to the team's success. Major inefficiencies exist in evaluating players, building a roster, and valuing player contracts. With the rise of [T20 cricket leagues](#) around the globe, there has been a significant interest in a comprehensive measure of overall player performance.

Inspired by sabermetrics, we are interested in estimating the value in terms of runs provided by the player compared to an average player (runs above average: RAA) and how that value translates to team wins compared to a replacement level player (win above replacement: WAR). RAA and WAR are useful metrics with simple meanings and units. RAA tells us how much value (runs) a player produces compared to a league-average player in the same situation of an innings after accounting for some factors. Using ball-by-ball data, we can estimate RAA for each ball faced by a batter, which can then be aggregated over a part of the innings, the whole innings, or the season. RAA for the player is then compared with the RAA of a replacement level player to evaluate WAR. If Virat Kohli, an Indian cricket player, has a WAR of around one, we could expect to drop one game in the standings over a whole season if we were to replace him with a replacement-level player.

1.1. Limited-overs Cricket

A cricket playing team consists of eleven players, and a maximum of ten can bat for a team. Wickets lost represent the number of outs for the batting team. Once players get out batting, they cannot return to bat in the game. A [limited-overs cricket](#) game comprises two innings, one for each team to bat. The toss-winning captain decides whether they bat or bowl first. An [over](#) consists of six balls (pitches) and limited-overs games have a fixed number of overs for each innings.

The team batting first sets the target, equal to their innings total plus one, for the other team to score to win. E.g., if the first innings total is 149, the team batting second has to score 150 to win the game. An innings concludes if the allotted number of overs have been bowled, the batting team loses all ten wickets (outs), or the batting team has reached the target. For the rest of the paper, we will use Twenty20 (T20) games as an example of a limited-overs game. As the name suggests, there are twenty overs (120 balls) in each innings of a T20 game.

1.2. Previous Work

[5] discusses evaluating true metrics based on the average of runs scored or conceded by over number. Say, in Big Bash League, the first over goes for 5.7 runs on average. If a particular bowler concedes five runs on average in the first over, their true economy is $5 - 5.7 = -0.7$, meaning they provide a value of -0.7 runs where negative values are considered good.

[4] shows a way of estimating RAA by using the phase of the innings, wickets lost, and the venue. However, it does not account for other factors such as innings number, bowling speed, and platoon advantage. Also, grouping by phase of the innings, e.g., power play (over 1- 6), death overs (16-20), etc., is an over-simplification as we can lose the interesting dynamics of each over's play within the power play or other phases. They also present the wins above average (WAA) metric. To simplify their WAA calculations, [4] uses the estimate of the second innings total if it were to continue even after reaching the target. This simplification would produce uncertainty in the WAA estimates,

which the author does not address. WAA takes an average player as the baseline player. Since average players are still valuable, it is not easy or possible for a team to replace a player evaluated with a player of average quality. In addition, [4] lacks the uncertainty estimation for their proposed metrics.

1.3. Contributions of cricWAR

cricWAR provides a framework for evaluating overall player performance, considering the context, and provides metrics such as run value, runs above average, and wins above replacement, which are easily interpretable. The run value for each play (ball) is estimated using expected runs. We adjust the run values for factors such as venue, innings (first or second), platoon advantage, and bowling pace (fast or spin) to allow for a fair comparison across different contexts. The adjusted run values are accredited to the players (batters and bowlers) using the run conservation framework [1]. A resampling-based method is used to estimate uncertainty in the proposed metrics. We further discuss the reliability and stability of traditional and cricWAR metrics. cricWAR is fully reproducible and open source. We use the R package [cricketdata](#), which provides easy access to ball-by-ball and player-meta data. *cricketdata* is co-developed by us.

2. cricWAR Model

Note that the [scoring is more prolific in cricket](#) than in baseball. On any given ball, a batter may take a single (1 run), a double (2 runs), a triple (3 runs) by running between wickets (bases) or hit a boundary which may be a four (4 runs) or a six (6 runs). The game will progress from one state to another mainly through the change of the over, along with some state changes when a wicket falls, which are pretty significant. These aspects of the game influence the expected runs framework defined below.

Our proposed measures are based on the *runs conservation framework*, which depends on the difference between expected and actual runs scored on each ball. The first step is establishing the expected runs scored, given the game's current state. We define the states of a T20 game by the over (o) and the wickets lost (w).

We define the **expected runs** (θ) for a state as

$$\theta(o, w) = E[R|over = o, wicketsLost = w] \quad (1)$$

where R is a random variable representing the runs scored on a ball and takes integer values ranging from zero through six. We consider a negative binomial regression model to estimate θ ,

$$\log(\theta) = \beta_0 + \beta_1 over + \beta_2 wicketsLost \quad (2)$$

The poisson regression in this case leads to overdispersion, meaning the $\text{Var}[R | o, w] > E[R | o, w]$. Thus, we modeled θ with negative binomial regression to address the overdispersion.

We define the **run value** (δ) for a ball as

$$\delta = r - \theta \quad (3)$$

where r is the actual runs scored on that ball. For each ball i , run value δ_i represent how well (positive) or worse (negative) a player did, compared to an average player, given the situation (state) of the game.

The **runs conservation framework** [1] dictates that every run value δ gained by batter on a ball is accompanied by the bowler gaining $-\delta$ for that ball. The extras (wides and no-balls) are completely attributed to the bowler.

2.1. Adjusting Run Values

2.1.1. Leverage Index

T20 cricket has some fielding restrictions, which affect the run-scoring in different stages of the innings. For the first six overs of an innings, also called power play (PP), only a maximum of two fielders are allowed outside the 30-yard (inner) circle. After PP, a maximum of five players are allowed outside the inner circle. Additionally, there can only be five players at a time on the leg side of the field.

With only two fielders outside in the inner circle during PP, batters take advantage by hitting shots to cross the inner circle and score boundaries (4s or 6s). Bowlers can place more fielders outside right after PP and control the scoring. Towards the end of the innings, in the last few overs (death overs), batters are trying to score quickly and taking more risks by hitting bigger shots to score boundaries. This means that players playing in different stages (PP, death overs, etc.) can quickly accumulate (or lose) run value and hinder comparing players across different playing positions.



Figure 1 During powerplay, only two fielders are allowed outside the inner circle. After powerplay, up to five fielders can be outside the inner circle.

The weighted average of expected runs θ , of all states is evaluated, and then θ for each state is divided by this weighted average to get the Leverage Index (LI) for each state. An LI of one corresponds to a state with an average wrt scoring rate, whereas a $LI \gg 1$ means a high-scoring state, and a $LI \ll 1$ means a low-scoring state.

$$LI(o, w) = \frac{\theta(o, w)}{\frac{\sum_{(o,w)} n_{(o,w)} \theta(o, w)}{\sum_{(o,w)} n_{(o,w)}}} \quad (4)$$

$$\delta^{lev} = \frac{\delta}{LI}$$

The run value δ is divided by the corresponding leverage index (LI) and labeled as leveraged run value δ^{lev} . LI helps account for situations that might be conducive for scoring (power play with fielding restrictions, last few overs of an innings.) or defending (middle overs).

3.1.2. Batting

Different factors beyond the control of batters affect the run values and make it hard to compare them across contexts. In particular, we want to account for the venue, innings (first Vs. second), platoon advantage (right-handed batter against left-handed bowler or vice-versa), and bowling pace (spin Vs. pace). We adjust for these factors using linear regression. We fit the leveraged run values using the covariates mentioned above.

$$\delta_i^{lev} = \beta_0 + \beta_1 \text{innings}_i + \beta_2 \text{platoon}_i + \beta_3 \text{bowlingpace}_i + \beta_4 \text{venue}_i + \epsilon_i \quad (5)$$

where all the covariates are categorical. The coefficients β_1 , β_2 , β_3 , and β_4 are the effects of the innings (first Vs. second), platoon advantage, the bowling pace (spin Vs. pace), and the venue on the leveraged run values, respectively. Ordinary least square is employed to estimate the coefficients. The estimated coefficients $\widehat{\beta}_1$, $\widehat{\beta}_2$, $\widehat{\beta}_3$, and $\widehat{\beta}_4$ are evaluated by ordinary least square using all the ball-by-ball data in the season.

The estimated residuals from the regression (5)

$$\widehat{\epsilon}_i = \delta_i^{lev} - \widehat{\beta}_0 - \widehat{\beta}_1 \text{innings}_i - \widehat{\beta}_2 \text{platoon}_i - \widehat{\beta}_3 \text{bowling pace}_i - \widehat{\beta}_4 \text{venue}_i \quad (6)$$

indicate the part of the leveraged run values δ_i^{lev} that is not accredited to the venue, innings, bowling pace, and platoon advantage. We refer to them as adjusted batting leveraged run values.

3.1.3. Bowling

The runs conservation framework dictates that the bowler gains leveraged run value $-\delta_i^{lev}$ for each δ_i^{lev} leveraged run value gained by the batter. Additionally, any extras (wides, no-balls) ω is entirely attributed to the bowler. Like (5), we adjust the bowling leveraged run value $-(\delta_i + \omega_i)^{lev}$ for the venue, innings, platoon advantage, or bowling pace.

$$-(\delta_i + \omega_i)^{lev} = \alpha_0 + \alpha_1 \text{innings}_i + \alpha_2 \text{platoon}_i + \alpha_3 \text{bowling pace}_i + \alpha_4 \text{venue}_i + \eta_i \quad (7)$$

The estimated residuals from the regression (7)

$$\hat{\eta}_i = -(\omega_i + \delta_i)^{lev} - \hat{\alpha}_0 - \alpha_1 \text{innings}_i - \hat{\alpha}_2 \text{platoon}_i - \hat{\alpha}_3 \text{bowling pace}_i - \hat{\alpha}_4 \text{venue}_i \quad (8)$$

indicate the part of the leveraged run values $-(\omega_i + \delta_i)^{lev}$ that is not accredited to the venue, innings, platoon advantage or the bowling pace. We refer to them as adjusted bowling leveraged run values.

3.2. Runs Above Average

To the best of our knowledge, the fielding tracking data is not available, at least publicly. In the absence of fielding data, we accredit these adjusted batting and adjusted bowling leveraged run values to the batter and bowler, respectively.

$$RAA_i^{\text{bowl}} = \hat{\eta}_i, \quad RAA_i^{\text{bat}} = \hat{\epsilon}_i \quad (9)$$

The runs above average for each ball a player played (batted or bowled), over a season, is summed to get the total measure of RAA_X for player X, where $I(\cdot)$ is the indicator function.

$$RAA_X = \sum_i RAA_i^{\text{bat}} \cdot I(\text{batter} = X) + RAA_i^{\text{bowl}} \cdot I(\text{bowler} = X) \quad (10)$$

3.3. Value over Replacement/Readily Available Player

The usual attempts at evaluating Value over Replacement Player (VoRP) involves measuring player contribution on the scale of runs/points relative to a baseline player. A natural choice would be to choose a league-average player as that baseline for comparison. However, as discussed in openWAR [1] and nflWAR [7], the league-average players are still quite valuable, and it is unlikely for a team to be able to replace the player in consideration with another league-average player. Instead, the team is more likely to look for a replacement player from a relatively inexperienced group of players from their team roster. Hence, for the baseline comparison, we consider a replacement-level player to be one that is readily available to replace the player being evaluated. E.g., in MLB, replacement level player is more likely to come from minor leagues, and in NFL, from the bottom part of their position-based depth chart. In T20 league cricket, the replacement player is likely to be a younger or inexperienced player who is part of the team squad (roster) for the league, such as IPL, PSL, etc.

We consider an approach similar to openWAR [1] in defining a squad(roster)-based replacement level. We use the Indian Premier League (IPL) to illustrate the approach. IPL teams have a minimum of 18 players and a maximum of 25 players on their squad. On average, teams have played 19 distinct players each season over the last decade. We looked at the number of distinct batters and bowlers and how many matches they played. Considering these stats, we decided to take eight batters and eight bowlers as league-level players. We take the top 8N batters and 8N bowlers regarding the playing time to be the league-level players, and the remaining players would be considered replacement-level players, where N is the number of teams.

All the replacement level players are pooled together, and their average performance is labeled as $\text{avg. } RAA_{rep}$, which represents RAA per ball for a replacement-level player. For every actual player,

we can associate a replacement-level shadow player with RAA_{rep} by taking the *avg. RAA_{rep}* and multiplying it by the number of balls played by the actual player. This replacement-level shadow player RAA_{rep} provides a relevant baseline for comparison specific to the player being considered.

Value over replacement player (VORP) is evaluated as

$$VORP_X = RAA_X - RAA_{rep} = RAA_X - (\text{avg. } RAA_{rep}) \cdot B_X \quad (11)$$

where RAA_X is the total RAA for player X in a season, and the B_X is the number of balls player X played in (batted plus bowled).

3.4. Wins Above Replacement

Wins above replacement (WAR) are evaluated as

$$WAR_X = \frac{VORP_X}{RPW} \quad (12)$$

where RPW is the runs per win for the league. Following the approach taken by [7], we use linear regression to estimate RPW by fitting the following model

$$W_i = \beta_0 + \beta \text{ Run Diff}_i + \epsilon_i \quad (13)$$

where W_i is the team i's win total in a regular season and Run Diff_i is the run differential for the team i in that regular season. The estimate $\hat{\beta}$ represents the increase in win total for each one unit increase in the team's run differential, meaning RPW can be estimated by $\frac{1}{\hat{\beta}}$. After fitting the model, we got $RPW \approx \frac{1}{0.011839} = 84.5$.

We did fit another model with the season as the additional covariate, but the effect of the season was insignificant with the same $\hat{\beta}$ estimate. We also followed the approach in [1] to estimate RPW using the Pythagorean win expectation formula, and the results were similar to the run differential approach.

4. Uncertainty Estimation

We can characterize two significant sources of variation in cricWAR metrics for each player in a given season: model estimation variation and player outcome variation. Model estimation variation arises due to the errors made during regression models for adjusting run values and the expected runs model. These models are trained on relatively large datasets; thus, this source of variation is relatively small compared to the player outcome variability.

The inherent uncertainty in the outcomes of all events involving a particular player for a particular season is considered player outcome variability. Consider a player with fixed ability and repeat the same season for that player many times. In each of these (repeated) seasons, the player's playing events would have variation in their outcome, leading to different season-level RAA and WAR

values. We target this player outcome variability in our uncertainty estimates for RAA and WAR values.

Following the approach in [1], we use a resampling strategy to generate distributions for each player's RAA and WAR values. For a particular season, we resample (with replacement) individual balls, fit the regression model for adjusting run values, and use the adjusted run values for new RAA and WAR values for individual players. In section 5, we present the uncertainty estimates with 1000 simulated seasons. The resampling approach quantifies the player outcome variability while maintaining the inherent correlation between the individual events.

5. Results

We present the results for the Indian Premier League (IPL) 2019 regular season. For expected runs estimation, we use the data from IPL seasons 2015 through 2022 except 2020, which was played outside India. cricWAR results are also available in the shiny app

<https://dazzalytics.shinyapps.io/cricwar/>. For the 2019 IPL season, 46 players are classified as replacement-level from the 161 players. Across all players, figure 2 shows WAR has right-skewed distribution with a median of 0.18, a mean of 0.36, and a standard deviation of 0.51. The skew is due to a few excellent players accumulating large WAR values. WAR for replacement players is approximately normal, with a mean of 0, a median of -0.075, and a standard deviation of 0.12.

Figure 3 shows the RAA values for all players from the IPL 2019 season, differentiating the league-level players (blue) from the replacement-level (red) players. The gray dots represent the RAA values for replacement-level shadows associated with each actual player. Playing time for a player

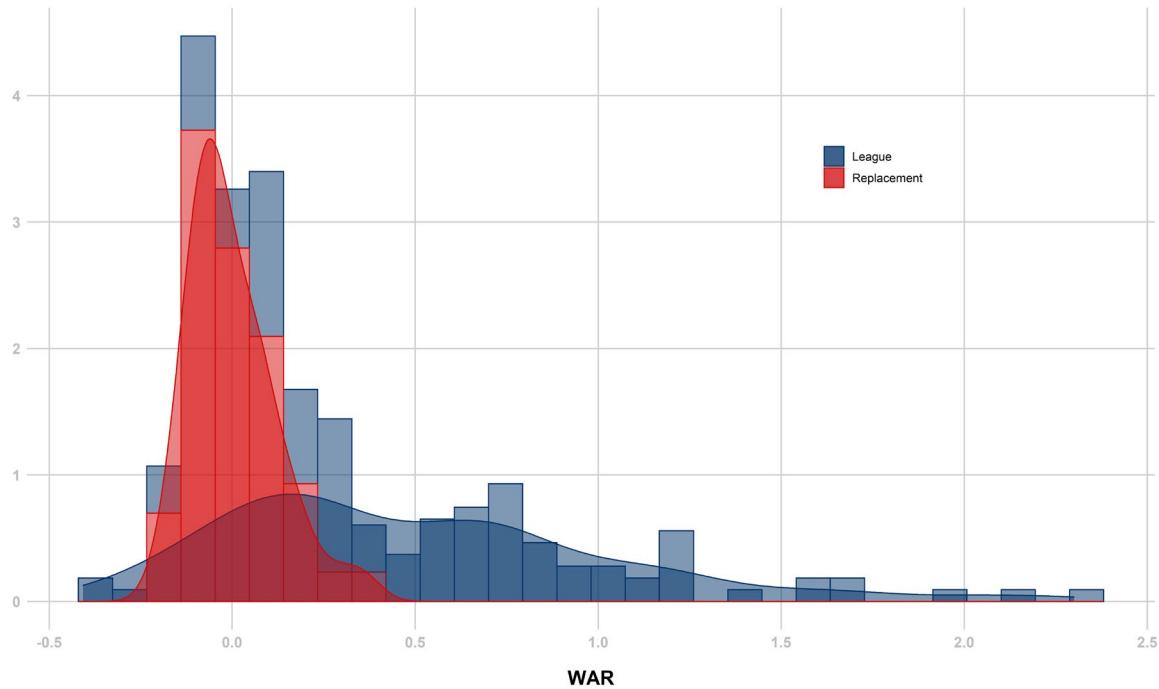


Figure 2 Distribution of WAR values for league and replacement level players from IPL 2019 regular season.

is evaluated as the number of balls faced by that player as a batter plus the number of balls bowled by that player as a bowler.

Table 1 and 2 shows the top ten best and worst batters and bowlers, wrt RAA, in the 2019 regular season. The players who both bat and bowl are assigned the playing role of all-rounders. We see AD Russell, an all-rounder who bats lower in the order, topping the batting RAA. Gayle and Butler represent the opening batters on the list. In bowling, Bumrah and Archer are miles ahead of the competition. SP Narine, an excellent all-rounder, rightfully makes an appearance in the top ten of both the batting and bowling RAA list. Access the list of all players with cricWAR stats at <https://dazzalytics.shinyapps.io/cricwar/>.

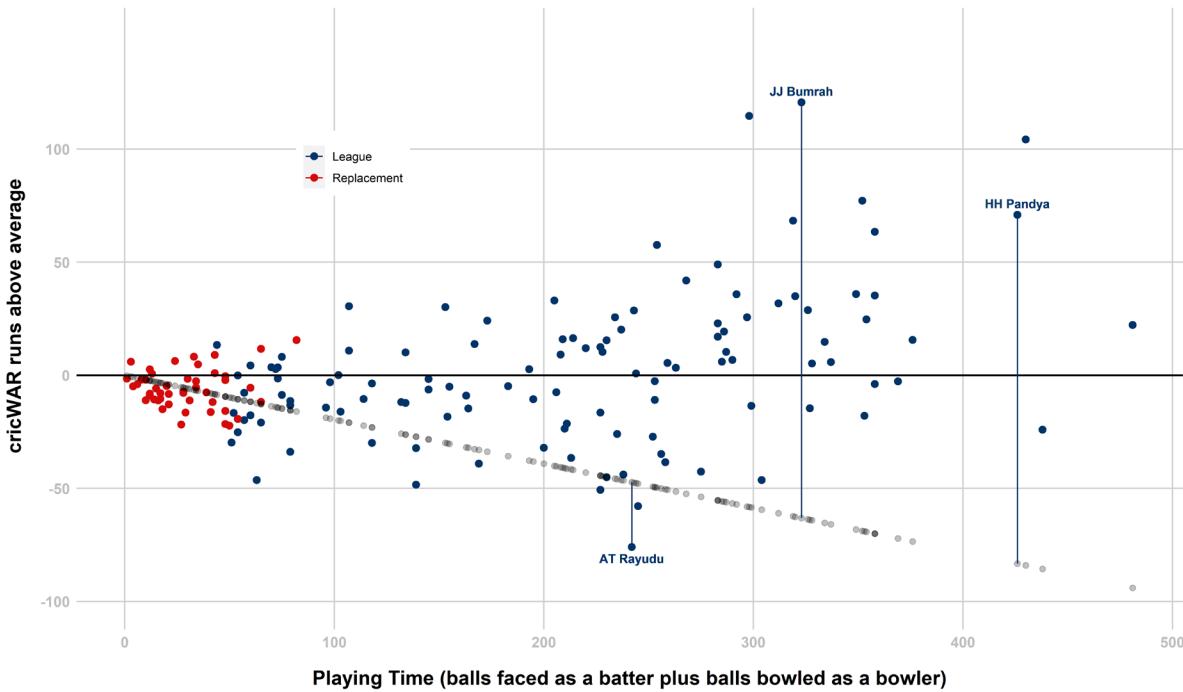


Figure 3 cricWAR RAA values for 2019 Indian Premier League (IPL) regular season. The season had 161 players play, and 46 were classified as replacement-level. Blue and red dots represent league-level and replacement-level players, respectively. Playing time for a player is evaluated as the number of balls faced by that player as a batter plus the number of balls bowled by that player as a bowler. The gray dots represent the replacement level shadow for each player. We highlight the difference between the player's RAA and the RAA for their replacement-level shadow for three players.

The results from table 1 suggest that it would be better to compare players playing the same positions, e.g., batters opening the innings plays under different circumstances (power play, new ball) than the batters playing in the middle part of the innings or the batters coming down that order who tends to be more aggressive in stroke play and score faster. We consider the top five opening batters and compare their batting RAA per ball in different phases of the innings in figure 4. CH Gayle stands out with positive RAA in each part of the innings, whereas other openers struggle in the middle or the death overs. This analysis could also be done over-by-over for an innings to investigate a batter's scoring patterns further. We present the over-by-over analysis of scoring patterns of CH Gayle and Virat Kohli in the Appendix.



Figure 4 Batting RAA per ball for opening batters wrt the phase of innings. CH Gayle stands out in all phases of the innings.

Batter Best	RAA	RAA per ball	Batter Worst	RAA	RAA per ball
AD Russell	111	0.446	AT Rayudu	-75.9	-0.314
HH Pandya	84.7	0.448	RV Uthappa	-57.8	-0.236
CH Gayle	68.3	0.214	Shubman Gill	-43.9	-0.184
RR Pant	57.6	0.227	SPD Smith	-42.6	-0.155
PA Patel	41.9	0.156	MP Stoinis	-40.6	-0.261
SP Narine	36.2	0.421	V Shankar	-40.4	0.222
JC Buttler	33.1	0.161	KM Jadhav	-39.1	-0.231
N Pooran	30.5	0.285	F du Plessis	-38.5	-0.149
KA Pollard	30.2	0.197	YK Pathan	-26.2	-0.583
MA Agarwal	25.7	0.110	RA Jadeja	-24.7	-0.297

Table 1 Batting RAA of top ten best and worst performers in the 2019 IPL regular season

Bowler Best	RAA	RAA per ball	Bowler Worst	RAA	RAA per ball
JJ Bumrah	122.0	0.379	JD Unadkat	-47.5	-0.212
JC Archer	103.0	0.400	S Lamichhane	-47.0	-0.341
Rashid Khan	61.6	0.183	TG Southee	-42.0	-0.778
B Kumar	49.0	0.148	PP Chawla	-40.2	-0.150
YS Chahal	42.0	0.142	DS Kulkarni	-39.1	-0.186
SP Narine	41.0	0.154	R Parag	-34.5	-0.411
NA Saini	37.7	0.131	K Gowtham	-28.8	-0.240
R Ashwin	36.6	0.111	VR Aaron	-28.4	-0.395
Imran Tahir	34.9	0.109	BA Stokes	-27.5	-0.273

MM Ali	31.0	0.207	S Nadeem	-25.2	-0.467
--------	------	-------	----------	-------	--------

Table 2 Bowling RAA of top ten best and worst performers in the 2019 IPL regular season

We consider the player outcome variability described in section 4 by simulating 1000 seasons of the actual 2019 IPL regular season. The simulation results are presented in figure 5, table 3, and figure 6. Note that the variation in player performance is not constant. Figure 5 shows the distribution of WAR values for the top five players by playing role and are sorted by their actual WAR point estimates. AD Russell has the highest point estimate of 2.25 WAR and has the highest variation in simulated WAR. Russell is an all-rounder, meaning he bats and bowls regularly. His batting strike rate (runs per ball) is relatively high since he hits a lot of boundaries (4s or 6s), and that could be a reason for the high variation in his WAR. Table 3 shows the quantiles and width of 95% confidence intervals of the simulated seasons for the top ten players by playing role. Top all-rounders have the highest variation, as shown by the large confidence interval width, whereas the top bowlers are among the ones with the least variation. JC Archer, a bowler, has the least simulated WAR variation among the top ten players and the top 3 cricWAR point estimate among all players, which suggests that the Archer provides relatively consistent performance on a ball-by-ball basis compared to other players.

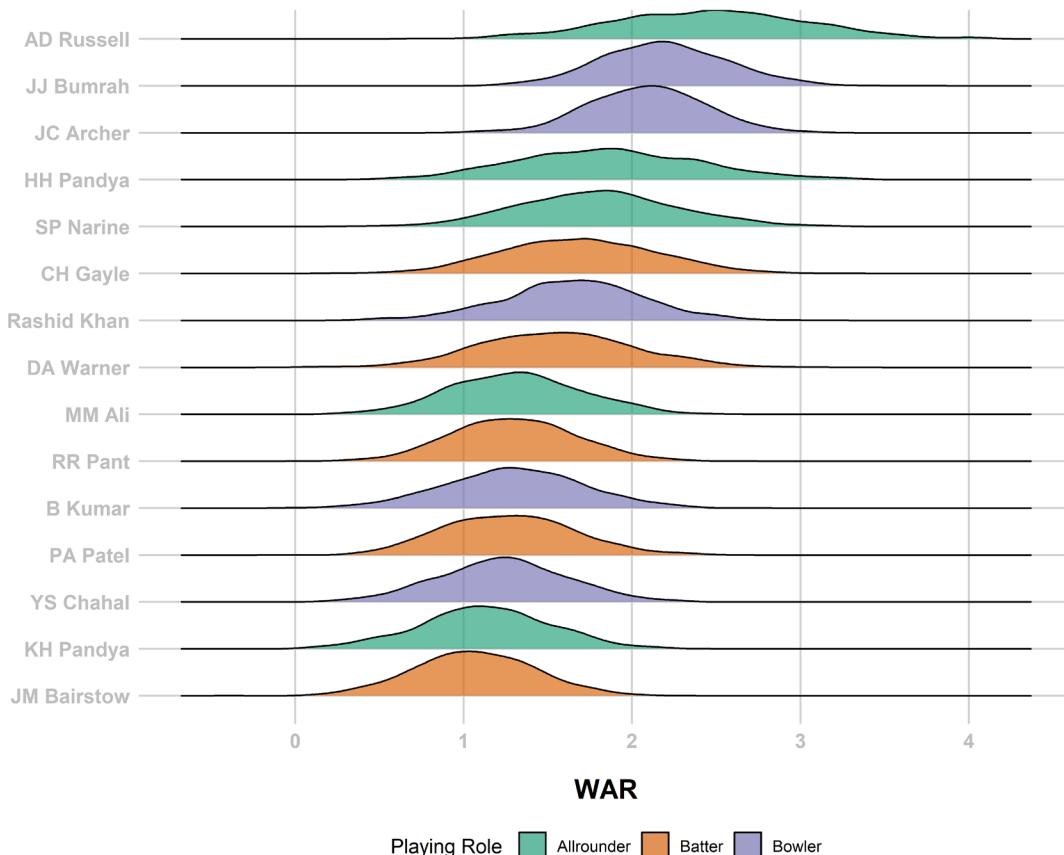


Figure 3 WAR simulation distributions of IPL 2019 regular season by playing role (batter, bowler, or allrounder) of top players. Distributions are based on 1000 simulated seasons.

Player	Role	Q2.5	Q25	Q50	Q75	Q97.5	Confidence interval width
AD Russell	Allrounder	1.28	1.98	2.38	2.77	3.51	1.11
JJ Bumrah	Bowler	1.56	2.07	2.33	2.58	3.00	0.72
JC Archer	Bowler	1.50	1.92	2.15	2.41	2.88	0.69
HH Pandya	Allrounder	0.97	1.63	2.02	2.38	3.03	1.03
SP Narine	Allrounder	0.93	1.52	1.84	2.18	2.76	0.92
Rashid Khan	Bowler	0.98	1.48	1.75	2.00	2.46	0.74
CH Gayle	Batter	0.78	1.36	1.67	1.95	2.57	0.90
DA Warner	Batter	0.78	1.26	1.56	1.88	2.55	0.88
R Ashwin	Bowler	0.61	1.10	1.39	1.65	2.16	0.78
RR Pant	Batter	0.68	1.14	1.38	1.62	2.10	0.71

Table 3 Distribution of top ten players' WAR for IPL 2019 regular season. Quantiles and 95% confidence interval are based on 1000 simulated seasons

Sports fans and media are always interested in comparing players and discussing whether one player is better. In particular, these discussions arise for awards such as MVPs (Mike Trout Vs. Miguel Cabrera[1]), Man of the Series/League, etc. cricWAR point estimates along with the WAR interval estimates (table 3) from simulated seasons allow for more refined analysis. We present, in figure6, the joint distributions of simulated WAR for a pair of bowlers and all-rounders from the 2019 IPL season. The black dots in figure 6 corresponds to the WAR point estimates of the players. AD Russell leads HH Pandya in WAR for almost 81% of the simulated seasons for all-rounders. And AD Russell had a cricWAR point estimate of 2.25 compared to 1.85 of HH Pandya. Thus, with some confidence, there is a high probability that AD Russell had a better season than HH Pandya. Bowlers, Bumrah and Archer had cricWAR point estimates of 2.19 and 2.06, respectively. From simulated seasons, we have Bumrah leading Archer in 58.1% of the seasons. Now, here we have a situation where there is a significant overlap in the interval estimates, and point estimates have a slight difference. Hence, there is enough uncertainty suggesting not to make a conclusive statement about Bumrah being better than Archer.

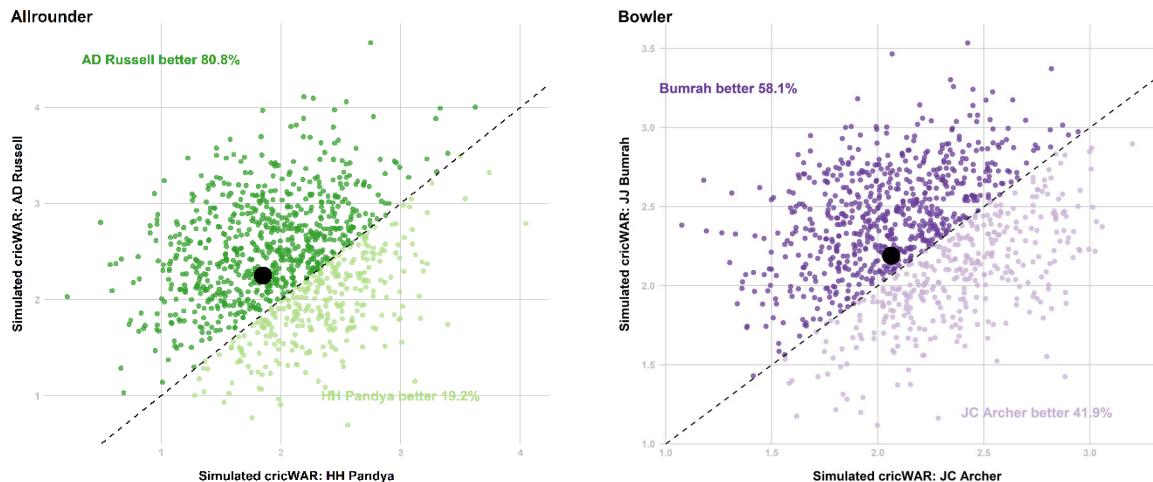


Figure 4 Joint distribution of simulated WAR for a pair of allrounders and bowlers, 2019 IPL season. An approach to compare player performance considering the player outcome variability.

5.1. Metric Reliability

We consider the stability of WAR from season to season by calculating the autocorrelation within players between consecutive seasons, presented in table 4. The correlation has been relatively stable over the years.

	`15 - `16	`16 - `17	`17 - `18	`18 - `19	`21 - `22
WAR autocorrelation	0.425	0.362	0.479	0.453	0.428
Matched Players	110	114	103	113	121

Table 4 Autocorrelation of WAR. WAR is estimated in consecutive seasons for each player, and the correlation between the players who played in both of the consecutive seasons is reported.

In [3], the authors present meta metrics to further understand the statistical properties of sports metrics. In particular, they present three meta metrics to evaluate the

- Discrimination: does the metric reliably differentiate between players?
 - It represents the fraction of between-player variance in metric m (in season s) due to true difference in player ability.
 - Discrimination meta-metric could be useful for attribution, e.g., end of year awards such as MVP, Man of the series, etc.
- Stability: does the metric measure a quantity that is stable over time?
 - It represents a fraction of the total variance in metric m, with sampling variability removed, that is due to within-player changes over time.
 - Stability meta metric could be useful for player evaluation and acquisition.
- Independence: does the metric provide new information?

of any given metric/stat.

The discrimination and stability meta-metrics are R-squared type measures extending the analysis of variance and taking values between zero and one. We evaluate the discrimination and stability of some traditional cricket stats and the cricWAR metrics. The results are presented in table 5. The discrimination is evaluated for the 2019 IPL season, and the stability is evaluated using the IPL season from 2015 through 2022, except 2020.

Batting efficiency in limited-over cricket is two-dimensional. You do not need only to score runs but score them at a fast rate. The reason is there are a limited number of balls to score, and that resource should be consumed effectively. Thus, the strike rate (SR), runs scored per ball, is a convenient stat to profile a batter. The boundary percentage (BP), the percentage of balls hit for a four or a 6 for batters, is a popular stat in T20 cricket and is among the most stable metrics considered here. This result is expected since the batters' hitting ability is usually relatively consistent and does not significantly change over seasons. Some of the cricWAR metrics, such as batting RAA and run value, are relatively more stable than traditional stats, such as total runs scored, total balls faced, and strike

rate (runs per ball). Total runs scored and balls faced are highly discriminatory, but as discussed in [3], aggregate stats can have high discrimination just by virtue of player position and corresponding playtime. The stability and discrimination values suggest that the cricWAR metric provides useful signals which are missing from the traditional stats.

Metric	Type	Discrimination	Stability
Boundary Percent	Rate	0.59	0.81
Batting RAA w/o leverage	Aggregate	0.59	0.73
Run Value w/o leverage	Aggregate	0.63	0.69
Batting RAA	Aggregate	0.55	0.67
Run Value	Aggregate	0.59	0.64
Strike Rate	Rate	0.52	0.54
Runs scored	Aggregate	0.95	0.42
Balls Faced	Aggregate	0.98	0.42

Table 5 Metric reliability for Batters w min of 60 balls faced in IPL Regular Season

6. Summary and Further Discussion

We present a novel and reproducible framework for evaluating overall player performance in limited-overs cricket. We develop metrics, such as run value, runs above average, and wins above replacement, which are easily interpretable by the public. Along with point estimates, we present the uncertainty estimates to understand the metrics better and discuss the metric reliability. The ball-by-ball data is used from the R package cricketdata, co-developed by us.

There are several exciting areas for further research. One of the main limitations is the lack of availability of fielding data. The location of the fielders at the time of the delivery of the ball and the subsequent movement by fielders can help us better allocate the run value to batters (particularly when running for runs) and bowlers (mainly when fielders make mistakes) and add the fielding contributions in the cricWAR framework as well. In the second innings, the batting team is chasing a target, and the required run rate per over affects the flow of the innings. Thus, the run values for the second innings could be considered differently after accounting for the required run rate. The ball-by-ball data from different T20 leagues can be pooled to estimate the expected runs and then used to estimate player effects across leagues. Also, by pooling the data, enough sample sizes would be available to go from (over, wickets lost) as the state of an innings to (over, ball in over, wickets lost) to capture the within-over variation.

6.1. Team Considerations

With the growing number of [T20 cricket leagues](#) around the globe, there is significant interest in evaluating the players for roster construction and drafting. These leagues draft from a pool of players who usually have played cricket at the international, league, or domestic level. And for the competitions/leagues these players have played, we can estimate these players' RAA and WAR values. Many of these players play in multiple leagues and have enough playing data.

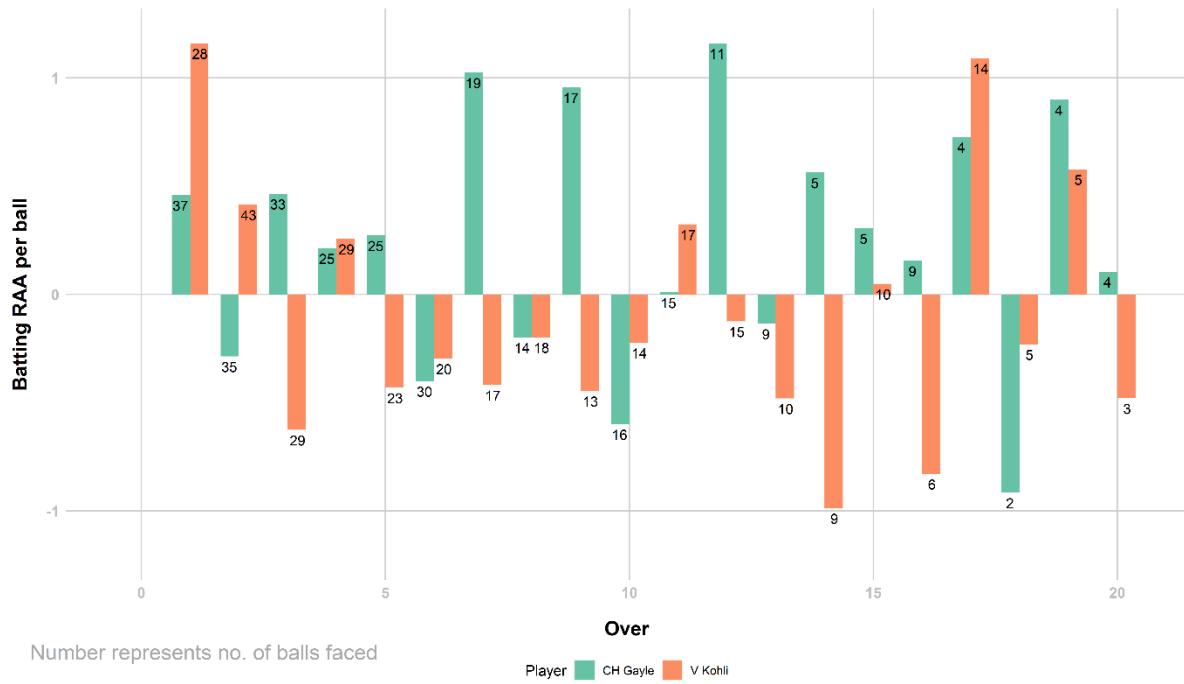
The teams could use RAA and WAR as objective measures to project the future career value of players in the T20 league drafts, similar to the work done for the NFL draft [2]. In addition, teams could create RAA and WAR based "draft pick value chart," similar to the Jimmy Johnson chart,

which has been improved by [2]. By considering the RAA of players at different phases (power play, middle, death) of the innings, team management may use players in different phases and adjust the playing order to maximize productivity. The T20 league teams draft players every year or every other year, and along with other information, they could use RAA/WAR estimates to assess the contract values better.

References

- [1] Benjamin S. Baumer, Shane T. Jensen, and Gregory J Matthews. "openWAR: An open source system for evaluating overall player performance in major league baseball". In: *Journal of Quantitative Analysis in Sports* (2015). doi: <https://doi.org/10.1515/jqas-2014-0098>.
- [2] Nick Citrone and Sam Ventura. *Statistical Methods to Maximize Value in the NFL Draft*. 2017. url: <https://ww2.amstat.org/meetings/jsm/2017/onlineprogram/AbstractDetails.cfm?abstractid=322234>.
- [3] Alexander Franks, Alexander D'Amour, Daniel Cervone and Luke bornn. "Meta-analytics: tools for understanding the statistical properties of the sports metrics ". doi: <https://doi.org/10.1515/jqas-2016-0098>
- [4] Himanish Ganjoo. *How many wins do the top players contribute to their team in an IPL season?* 2020. url: <https://www.espncricinfo.com/story/ipl-2020-how-many-wins-do-the-top-players-contribute-totheir-team-in-an-ipl-season-1236781>.
- [5] Jarrod Kimber. The new vocabulary of T20. url: <https://www.espncricinfo.com/story/jarrod-kimber-the-new-vocabulary-of-t20-1133997>
- [6] Tom Tango. *Crucial Situations*. 2006. url: <https://tbt.fangraphs.com/crucial-situations/>.
- [7] Ronald Yurko, Samuel Ventura, and Maskim Horowitz. "nflWAR: a reproducible method for offensive player evaluation in football". In: *Journal of Quantitative Analysis in Sports* (2019). doi: <https://doi.org/10.1515/jqas-2018-0010>.

Appendix



The figure shows the over-by-over comparison of RAA per ball generated by two opening batters, CH Gayle and Virat Kohli, from the 2019 IPL season.