

# How Well Do Strike Rates Translate from T20 to ODI Formats for Top 100 vs. Non-Top 100 players?\*

My subtitle if needed

Dhruv Gupta

December 3, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Data Selection and Sourcing . . . . .	2
2.2	Raw Data . . . . .	2
2.2.1	Match Information . . . . .	3
2.2.2	Batting and Bowling Teams . . . . .	3
2.2.3	Runs and Extras . . . . .	3
2.2.4	Wicket Details . . . . .	4
2.3	Cleaning . . . . .	4
2.3.1	Desired Structure of Analysis Data . . . . .	4
2.3.2	Filtering on Balls Faced . . . . .	5
2.3.3	Calculating Average Strike Rates, Identifying Top 100 Players . . . . .	5
2.3.4	Structuring the Final Dataset . . . . .	5
2.4	Summary of Analysis Data . . . . .	6
<b>3</b>	<b>Model</b>	<b>7</b>
3.1	Model Set-Up . . . . .	7

---

\*Code and data are available at: <https://github.com/dhruv5423/Top-100-Batters-T20-to-ODI-Strike-Rate-Mapping>

3.2	Model Justification . . . . .	8
<b>4</b>	<b>Results</b>	<b>8</b>
4.1	Fitted Values . . . . .	8
<b>5</b>	<b>Discussion</b>	<b>10</b>
5.1	First Discussion Point . . . . .	10
5.2	Second Discussion Point . . . . .	10
5.3	Third Discussion Point . . . . .	10
5.4	Weaknesses and Next Steps . . . . .	10
<b>6</b>	<b>Appendix</b>	<b>11</b>
<b>7</b>	<b>Additional Data Details</b>	<b>11</b>
<b>8</b>	<b>Model Details</b>	<b>11</b>
8.1	Posterior Predictive Check . . . . .	11
8.2	Diagnostics . . . . .	11
	<b>References</b>	<b>12</b>

# 1 Introduction

The sport of cricket has evolved greatly in recent times. Traditionally played over the course of 5 days, new, shorter formats are becoming increasingly popular due to their accessibility and faster pace of play. One Day International (ODI) and T20 formats are foremost among these formats - consisting of one innings of 50 overs and 20 overs respectively for either side. (International Cricket Council 2024) All three formats, Test, ODI, and T20 present different challenges and require different skill-sets. Lasting 5 days, test matches require endurance, patience and strategy. In comparison, T20 cricket is much more exciting, with batsmen ‘aiming for boundaries’ as the teams try to score as many runs as possible in this short format. ODI cricket is a balance of the two - while relatively fast, strategy is key and slower play is more prevalent. (Hnrd 2024)

A batter’s strike rate is a key performance indicator of their ability to score runs. It is defined as the ‘runs per 100 balls’ they face. (Sportmakers 2024) For example, if a batter scores 100 runs in 50 balls, their strike rate can be calculated to be 200. Given the differences in strategy between formats, it is common to expect differences in strike rates. Longer formats generally have lower strike rates, and shorter formats higher. But how well do strike rates translate across formats, particularly for top-performing players compared to the rest of field? This is of particular interest, not only in evaluating individual player adaptability, but in understanding how skills in one format can be extrapolated to another.

This paper aims to investigate whether how well strike rates translate from T20 to ODI formats, for top-performing players compared to their peers. Specifically, we aim to estimate the difference in ODI strike rate attributable to 1) T20 Strike Rates, 2) Player Group (whether they are a top performing batter or not), and 3) the interaction between the two factors. Our estimate is the conditional effect of T20 Strike Rate on ODI Strike Rate, adjusted for player type, while accounting for player-specific variability.

Estimand

Results

Broader Context

The rest of paper is structured as follows..

## 2 Data

### 2.1 Data Selection and Sourcing

Data used in this paper has been sourced from the reputable cricket database, CricSheet, (Rushe 2024). Created by Stephen Rushe, CricSheet is an open-source platform providing detail ball-by-ball data for 17,944 cricket matches. Contained in the database were 2,858 One Day International matches and 3,856 T20 International matches. This paper focuses on the male cricket matches, with 2,391 Male ODIs and 2,499 Male T20 match data downloaded as CSV files. This data set is particularly well-suited for this study because it ensures consistent measurement across matches, detailed coverage of player-level performance, and a large sample size, enabling robust statistical analysis of performance differences across formats.

Additional details on the data selection process and a step-by-step guide to access and download the data are provided in Section 7.

### 2.2 Raw Data

Below are samples from the raw data regarding ODI matches after all raw csv files of individual match data were combined. The data is split into logical groups for clarity, but are all part of the same data frame.

### 2.2.1 Match Information

Table 1 contains basic information about the match: match ID, season, start date, venue, innings, and ball. So, the first row would give match information of Match ID 1131279 at the ICC Academy in the 2017/2018 season, on the 1st of January 2023. The ball that the data captured is the 4th ball from the 44th over, in the first innings.

Table 1: Raw Data - Sample Match Information

Match Id	Season	Start Date	Venue	Innings	Ball
1131279	2017/18	2018-01-23	ICC Academy	1	44.4
1031673	2017	2017-09-27	Kennington Oval	2	26.6
656483	2014/15	2015-03-18	Sydney Cricket Ground	1	20.5

### 2.2.2 Batting and Bowling Teams

Table 2 is about the batting team, bowling team, and key players involved at the time of the delivery. The first row can be understood to show that the time of delivery 44.4 of Match 1131279 (corresponding to row 1 in Table 1), the batting team is India, the bowling team is New Zealand, the Striker is V Kohli, the other batsman (or the non-striker) is SS Iyer, and the Bowler is TA Boult.

Table 2: Raw Data - Sample Batting and Bowling Teams Information

Batting Team	Bowling Team	Striker	Non Striker	Bowler
India	New Zealand	V Kohli	SS Iyer	TA Boult
Sri Lanka	South Africa	I Udana	CAK Rajitha	A Nortje
Sri Lanka	New Zealand	WU Tharanga	DPMD Jayawardene	SE Bond

### 2.2.3 Runs and Extras

Table 3 contains the runs scored, extras such as wides, no-balls, byes, leg-byes, and penalties. So for the same ball in Table 1 and Table 2, 1 run was scored by V Kohli. There were no extras, wides, no balls, byes, legbyes, or penalties.

Table 3: Raw Data - Sample Runs and Extras Information

Runs Off Bat	Extras	Wides	Noballs	Byes	Legbyes	Penalty
1	0	NA	NA	NA	NA	NA
0	0	NA	NA	NA	NA	NA

Table 3: Raw Data - Sample Runs and Extras Information

Runs Off Bat	Extras	Wides	Noballs	Byes	Legbyes	Penalty
0	0	NA	NA	NA	NA	NA

### 2.2.4 Wicket Details

Table 4 contains information about wickets, including the type of dismissal and the players involved. For each of the 3 balls displayed in this sample there were no wickets or players dismissed.

Table 4: Raw Data - Sample Wicket Details

Wicket Type	Player Dismissed	Other Wicket Type	Other Player Dismissed
		NA	NA
		NA	NA
		NA	NA

## 2.3 Cleaning

### 2.3.1 Desired Structure of Analysis Data

Refer to the “Sketches” section of the repository. There was an understanding of how the dataset needed to be transformed in order to get it ready for analysis through our model. The desired structure of the data aimed to include the following key variables: player name, player type, match type, strike rate, and total balls faced. Below is an example of a representation of the desired structure:

Table 5: Desired Structure of Analysis Data

Player	Player Type	Match Type	Strike Rate	Total Balls Faced
Player A	Top 100	T20	140	1200
Player A	Top 100	ODI	90	1200
Player B	Non-Top 100	T20	120	800
Player B	Non-Top 100	ODI	75	800

### 2.3.2 Filtering on Balls Faced

The first step after downloading raw data files and combining them into large T20 and ODI ball by ball data files was to filter players who had a large enough sample size across formats. That is to say, only those who meaningfully participated. Importantly we wanted to exclude players whose sample size of balls faced were too few, as these observations could skew the findings from the paper. Therefore, players were required to meet a minimum threshold of 200 balls faced in T20 matches and 500 balls faced in ODI matches. The combined balls faced in both formats needed to be above a threshold of 700. In doing so, we avoid noise from low-sample sizes. The library dplyr, (Wickham et al. 2023), was used to filter and aggregate the data.

### 2.3.3 Calculating Average Strike Rates, Identifying Top 100 Players

The ‘mean’ function was used to calculate the average T20 Strike Rate and average ODI strike rate. Missing or null values were excluded using ‘na.rm = TRUE’.

Because part of our model is concerned with differences between top performing players and their peers, we obtain the Top 100 batsmen before ranking them based on the average of their T20 and ODI strike rates. Here, the ‘mutate’ function was used to calculate ranks and assign each player as either ‘Top 100’, or ‘Non-Top 100’.

### 2.3.4 Structuring the Final Dataset

The arrow package was used to save the cleaned data set as a Parquet file. (Richardson, Hester, et al. 2023) The final cleaned dataset now included the columns “Player”, “Player Type”, “Match Type”, “Strike Rate”, “Total Balls Faced”. Table 6 contains a sample of the first 5 rows of the final analysis data.

Table 6: Sample of Analysis Data

Striker	T20 Strike Rate	ODI Strike Rate	Total Balls Faced	Player Type
CK Kapugedera	114.60	69.77	2746	Non-Top 100
MD Shanaka	116.76	87.55	2540	Non-Top 100
EJG Morgan	132.92	89.52	9605	Top 100
SA Edwards	122.68	90.66	2457	Non-Top 100
LRPL Taylor	118.49	80.94	11668	Non-Top 100

## 2.4 Summary of Analysis Data

After getting the data ready for analysis, it may be useful to visualize the data before we run the model. Figure 1 displays the average strike rates for both match types, for both player groups. As expected, we find that Top 100 players have higher strike rates than Non-Top 100 players.

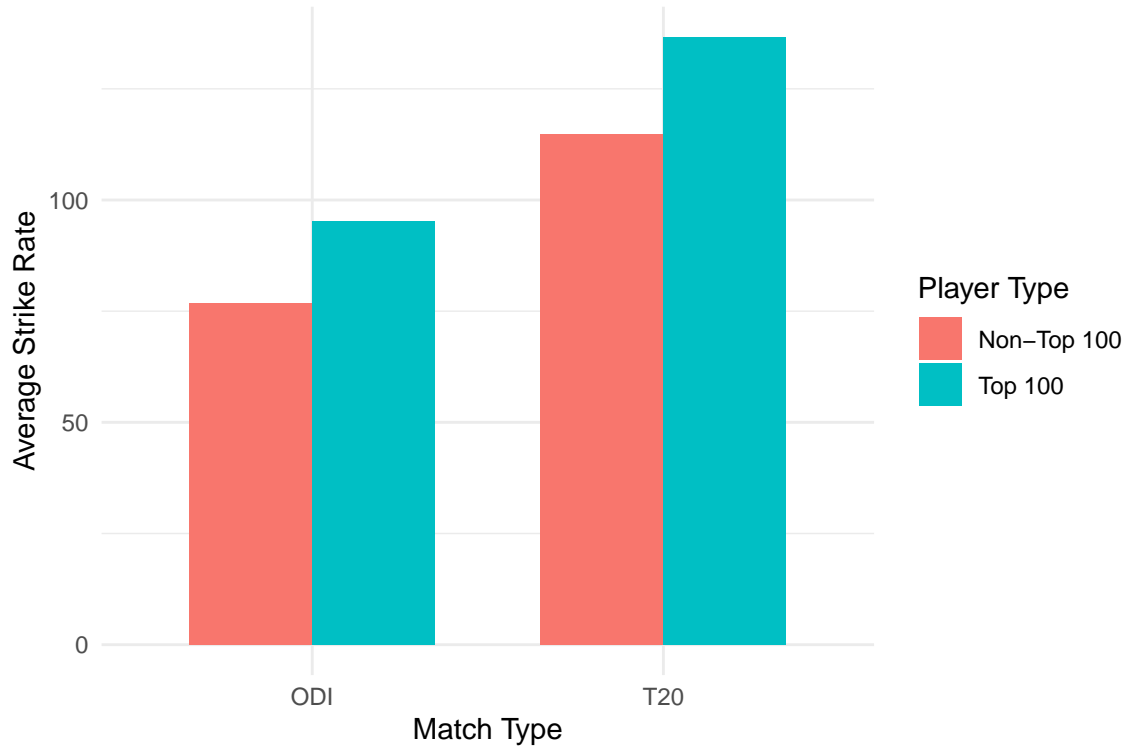


Figure 1: T20 vs. ODI Average Strike Rates by Player Group

Table 7 summarizes the average strike rates for each player type - numerically describing Figure 1.

Table 7: Average Strike Rates by Player Type

Player Type	T20 Strike Rate	ODI Strike Rate
Non-Top 100	114.7136	76.76375
Top 100	136.6294	95.16540

## 3 Model

### 3.1 Model Set-Up

In order to investigate how T20 strike rates translate into ODI Strike rates, we fit a linear regression model. Our outcome variable is  $Y_i$ , for which we attempt use 3 explanatory variables to fit: T20 Strike Rate -  $X_{1i}$ , a dummy variable for player type (whether they are Top 100 or not) -  $X_{2i}$ , and an interaction term between the two,  $X_{1i} \cdot X_{2i}$ .

The mathematical representation of the model is as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \cdot X_{2i}) + \epsilon_i$$

Where:

- $Y_i$ : The ODI strike rate of player  $i$ .
- $X_{1i}$ : The T20 strike rate of player  $i$ .
- $X_{2i}$ : A binary indicator of player type, where  $X_{2i} = 1$  for “Top 100” players and  $X_{2i} = 0$  for “Non-Top 100” players.
- $X_{1i} \cdot X_{2i}$ : The interaction term between T20 strike rate and player type.
- $\beta_0$ : The intercept, representing the predicted ODI strike rate for “Non-Top 100” players with a T20 strike rate of 0.
- $\beta_1$ : The main effect of T20 strike rate, representing the change in ODI strike rate for every 1-unit increase in T20 strike rate among “Non-Top 100”
- $\beta_2$ : The main effect of player type, representing the difference in ODI strike rate between “Top 100” and “Non-Top 100” players when T20 strike rate is 0.
- $\beta_3$ : The interaction effect, representing how the relationship between T20 and ODI strike rates differs for “Top 100” players.
- $\epsilon_i$ : The residual error, assumed to be normally distributed with mean 0 and constant variance  $\sigma^2$ .

We fit the model using the `lm()` function in R, implemented in the **Base R** package.

Our response variable is  $Y$ , ODI Strike Rate. Our predictors are the continuous variable T20 Strike Rate,  $X_1$ , the binary variable Player Type,  $X_2$ , and the interaction term  $X_1 \cdot X_2$ .



## 3.2 Model Justification

We attempt to capture the relationship between T20 Strike Rates and ODI Strike Rates, while accounting for player type using a dummy variable. Further, the inclusion of the interaction term directly tackles the research question, allowing us to test whether the relationship between T20 and ODI Strike Rates differs between groups - demonstrating potential differences in their adaptability.

In designing this model, we make the following assumptions:

1. **Linearity in Coefficients:** We assume that there exists a linear relationship between T20 Strike Rates and ODI Strike Rates. The coefficient  $\beta_1$  captures this. Additionally, the effect of player type,  $\beta_2$ , is also assumed to be linear. Furthermore, the coefficient of the interaction term,  $\beta_3$ , is also modeled linearly. Section 8 runs a diagnostic plot to check for linearity.
2. **Residuals are Normally Distributed:** This is to say that the variability of ODI Strike Rate around the predicted value we fit using the model should be normally distributed, around 0. Section 8 plots a histogram of the residuals - a bell shaped curved implies normality.
3. **Homoscedasticity:** This assumption is built into the model to imply that the variance of the residuals,  $\epsilon_i$ , is constant across independent variables. That is to say that the spread of the residuals should be homogenous regardless of whether we look at the predictor  $\beta_1$  or  $\beta_2$ . Section 8 runs a residuals vs. fitted values plot to check for homoscedasticity.
4. **No Collinearity:** This assumption states that our predictor variables are not highly correlated with each other - this can make it difficult to interpret our coefficients as we are unsure regarding the interaction between the two, and if one is influencing the other.

## 4 Results

### 4.1 Fitted Values

To understand the results of the model, we begin by examining the fitted values. These are the predicted ODI Strike Rates for each player based on the T20 Strike Rates, Player Type, and their interaction. The fitted values are computed using the following equation:

$$\hat{y} = \beta_0 + \beta_1 \cdot \text{T20 Strike Rate} + \beta_2 \cdot \text{Player Type Binary} + \beta_3 \cdot (\text{T20 Strike Rate} \times \text{Player Type Binary})$$

where:

- $\hat{y}$  is the predicted ODI Strike Rate,

- $\beta_0$  (Intercept) = 54.87,
- $\beta_1$  (T20 Strike Rate Coefficient) = 0.19,
- $\beta_2$  (Player Type Binary Coefficient) = 19.36,
- $\beta_3$  (Interaction Term Coefficient) = -0.038.

The fitted values provide insight into how well the model captures the observed data. Below is a table of a sample of the observed ODI Strike Rates and their corresponding fitted values, grouped by Player Type.

Table 8: Sample of Fitted Values and Observed ODI Strike Rates

Striker	Player Type	T20 Strike Rate	ODI Strike Rate (Observed)	ODI Strike Rate (Fitted)	Residuals
Jatinder Singh	Non-Top 100	115.02	71.37	76.82	-5.45
CK Kapugedera	Non-Top 100	114.60	69.77	76.74	-6.97
KH Prajapati	Non-Top 100	103.95	73.72	74.71	-0.99
Aqib Ilyas	Non-Top 100	118.34	72.27	77.46	-5.19
H Klaasen	Top 100	137.64	113.90	95.32	18.58
MN Waller	Non-Top 100	136.66	75.13	80.95	-5.82
JE Root	Non-Top 100	122.53	85.38	78.26	7.12
Aff Hossain	Non-Top 100	116.65	88.58	77.13	11.45
XM Marshall	Non-Top 100	97.75	55.98	73.53	-17.55
WA Young	Non-Top 100	99.71	85.83	73.90	11.93

## **5 Discussion**

### **5.1 First Discussion Point**

### **5.2 Second Discussion Point**

### **5.3 Third Discussion Point**

### **5.4 Weaknesses and Next Steps**

## **6 Appendix**

### **7 Additional Data Details**

### **8 Model Details**

#### **8.1 Posterior Predictive Check**

#### **8.2 Diagnostics**

Residuals vs Fitted Values

## References

- Hnrd. 2024. “Cricket Analysis and Insights.” <https://www.behind.hnrd.co/blog/6jmr1026fpsj9mdn2orestodtiegl>.
- International Cricket Council. 2024. “The Three Formats of Cricket.” <https://www.icc-cricket.com/about/cricket/game-formats/the-three-formats>.
- Richardson, Neal, Jim Hester, et al. 2023. *Arrow: Integration to 'Apache Arrow' Library*. <https://CRAN.R-project.org/package=arrow>.
- Rushe, Stephen. 2024. “CricSheet: Ball-by-Ball Cricket Data for Male ODI and T20 Matches.” <https://cricsheet.org/downloads/>. <https://cricsheet.org/downloads/>.
- Sportmakers. 2024. “What Is the Strike Rate in Cricket?” <https://www.sportmakers.co.uk/cricket/what-is-the-strike-rate-in-cricket/>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.