A Holistic Examination of Streakiness and Consistency in Major League Baseball

Ryan Feder
The Dalton School
New York, NY

Advised by:
Luke Benz, MA
Doctoral Candidate in Biostatistics
Harvard T.H. Chan School of Public Health
Boston, MA

Abstract

Streakiness and baseball go hand in hand, but accurately measuring streakiness and consistency in sports is difficult. While studying hitting streaks is an old idea, relatively few works have examined streaks for hitters at the pitch outcome granularity, or for pitchers more generally. Furthermore, little is understood about how streaks correlate with more traditional player outcomes. In this work, we utilize permutation tests, which we use to apply two metrics to four outcomes of interest from the perspective of both hitters and pitchers in order to quantify internal player streakiness in a holistic manner. This method is used to study the streakiness and consistency of the 136 batters and 127 pitchers during the 2023 Major League Baseball season, and study the association between streakiness/consistency and traditional player statistics. Our findings suggest that league-wide trends in pitcher outcomes are slightly streakier than those of hitters, and that consistency for both player types is moderately correlated with traditional measures of player success. Finally, our approach identifies Ronald Acuña, the unanimous 2023 National League Most Valuable Player, as a model of consistency, demonstrating the utility of a holistic approach to player streakiness evaluation.

1) Introduction

There is a saying that basketball "is a game of runs," though this saying reasonably applies to all sports, especially baseball. Momentum is something that is difficult, if not impossible to quantify, with ideas like the "hot hand" in basketball still under debate today[1]. But this hard to define feeling of momentum, as all sports fans know, certainly feels tangible[2] when a player gets hot or when a team is on a roll.

In baseball in particular, streaks are a huge part of the history of the game, and they are not just special for the players who have them. Fans have gravitated towards streaks such as Joe Dimmagio's record 56-game hitting streak, the Cleveland Indians' 22 game win streak in 2017, or even Cal Ripken's 2,632 consecutive games played. Even this past MLB season, there were streaks that stood out including the Rays 13 game win streak to open the season, Luis Arraez maintaining a batting average of nearly .400 for the first half of the year, and Shohei Ohtani and Matt Olsen both homering like crazy over the summer. Simply put, streaks are as integral to baseball and its rich history as batting average, strikeouts and home runs.

Just as streaks have been a huge part of baseball lore, the study of streaks has been the focus of several statistical works. For example, (Albert, 2008)[3] presented a rigorous evaluation of hitting streaks during the 2005 MLB season across patterns of hits/outs, home runs and strikeouts, using several proposed statistical metrics to capture various notions of what it means for a player to be streaky, including metrics based on permutational inference. Albert concluded that some players during that season exhibited more streakiness than one would predict from random exchangeability alone. Noting that what it means to be streaky may differ based on relative frequency of an outcome, Albert applied different metrics in subsequent work[4-5] to analyze the gaps between consecutive home runs.

In other work, (McCotter, 2008)[6] analyzed player-seasons between 1957–2006 and found more hitting streaks at the game level than one would expect due to random chance under an independence assumption. (Albright, 1993)[7] similarly found evidence of streaky players within a given season, but found that such streakiness did not tend to last over a larger time frame of four years. Bock and colleagues suggested that hot hitting is contagious in the sense that when a player was on a multi-game hitting streak, his teammates were more likely to demonstrate better hitting performance[8]. The previously mentioned works are merely selections from a large body of work on hitting streaks in baseball. For comprehensive reviews on streakiness in baseball and the hot-hand in sports more generally, we recommend (Reifman, 2012)[9], (Bar-Eli et al, 2006)[10], and (Cohen, 2020)[11].

Comparatively fewer works have looked at streakiness for pitchers, perhaps because the notion of what is meant by pitcher streakiness is less intuitive. (Gamble, 2015)[12] studied pitcher streakiness from a fantasy baseball perspective and found that a pitcher's previous three starts had little to no predictive value for projecting the fantasy value of his next start. (Arthur and Matthews, 2017)[13] used a Hidden Markov Model to classify pitchers into states of hot/normal/cold solely based on fastball velocity. While fastball velocity is a different outcome than has previously been the focus of streakiness literature, which primarily examines binary outcomes, this work by Aruther and Matthews is noteworthy in that the outcome of interest is a pitch-level outcome, rather than an at-bat or game level outcome. More recently, (Evanko, 2020)[14] studied streakiness in a sample of 50 pitchers from the 2019 season and found little evidence of the hot hand.

To date, the majority of the literature has focused on analyzing at-bat or game level outcomes for hitters, rather than more granular outcomes at the pitch or swing level. What work does exist for pitchers does not utilize tools of permutational inference traditionally applied to analysis of hitters[3-5,15-17]. Furthermore, to our knowledge, much of the literature on streaks has been interested in classifying players as abnormally streaky, and to a lesser extent, abnormally consistent, but little work has actually assessed how such classifications are associated with statistics by which players are traditionally evaluated.

In this work, we bridge these gaps by applying streak and spacing ideas to analyze consistency and volatility of 136 hitters and 127 pitchers during the 2023 MLB season across four distinct outcomes at both at-bat and pitch level outcomes to create holistic streakiness profiles. Furthermore, we examine how these profiles correlate with traditional statistics such as runs batted in (RBI), batting average, and earned run average (ERA), to name a few, in order to better understand whether and how streaks underlie the success of the games' better players. In other words, do MLB stars players arrive at their results in a manner that is consistent or one with more extreme fluctuations?

The rest of this paper is structured as follows: Section 2 outlines the methods used to evaluate player streakiness, which are evaluated via a simulation study in Section 3. Section 4 presents the results of applying this method to the 2023 MLB season, and Section 5 concludes with some discussion.

2) Methods

2.1) Mathematical Formulation of Streakiness

Streakiness is inherently somewhat of a vague term, whose essence can be qualitatively described in several ways. Frequent clumping of successes or failures, success portending subsequent success and failure signaling subsequent failure, high variability, inconsistent, and volatile are all notions of streakiness which sound intuitive but imprecise.

Towards attempting to formulate a precise mathematical notion of streakiness, let Y_{ij} denote a binary outcome for player *i* during observation *j*, with 1 indicating a success and 0 indicating a failure. In the majority of the literature reviewed in Section 1, *j* indexes a player's at-bats with Y_{ij} denoting indicators of hits (H) or outs. In our work, Y_{ij} will denote several different outcomes for both hitters and pitchers, which we outline in Section 2.2, across various outcome granularities. For certain outcomes, *j* will index at-bats or plate appearances, while for other outcomes *j* will index pitches or swings.

Define player *i*'s rolling mean (RM) of size *m* observations beginning at observation *k* as follows:

$$RM_i(m,k) = \frac{1}{m}\sum_{j=k}^{k+m-1} Y_{ij}$$

While perhaps somewhat complex looking, this is simply the percentage of the most recent *m* outcomes which yielded successes, anchored at observation *k*. An example of 25 observation rolling means $RM_i(25,k)$ for various outcomes during Ronald Acuña's 2023 season is shown below in Figure 1.

Intuitively, players with more large fluctuations in their rolling means $RM_i(m,k)$ should be classified as more streaky, while players whose fluctuations are smaller should be classified as more consistent. Visually it's hard to inspect what is meant by large fluctuations. Whether the rolling means in Figure 1 present evidence of abnormal streakiness requires some notion about what a large fluctuation even entails. We utilize previous work by Jim Albert[3-5,15-17] as a starting point to define two metrics which capture streakiness from this point of view. Using simulation studies, we show in Section 3 that these metrics adequately describe whether or not a player is more streaky or more consistent on both common and rare outcomes.

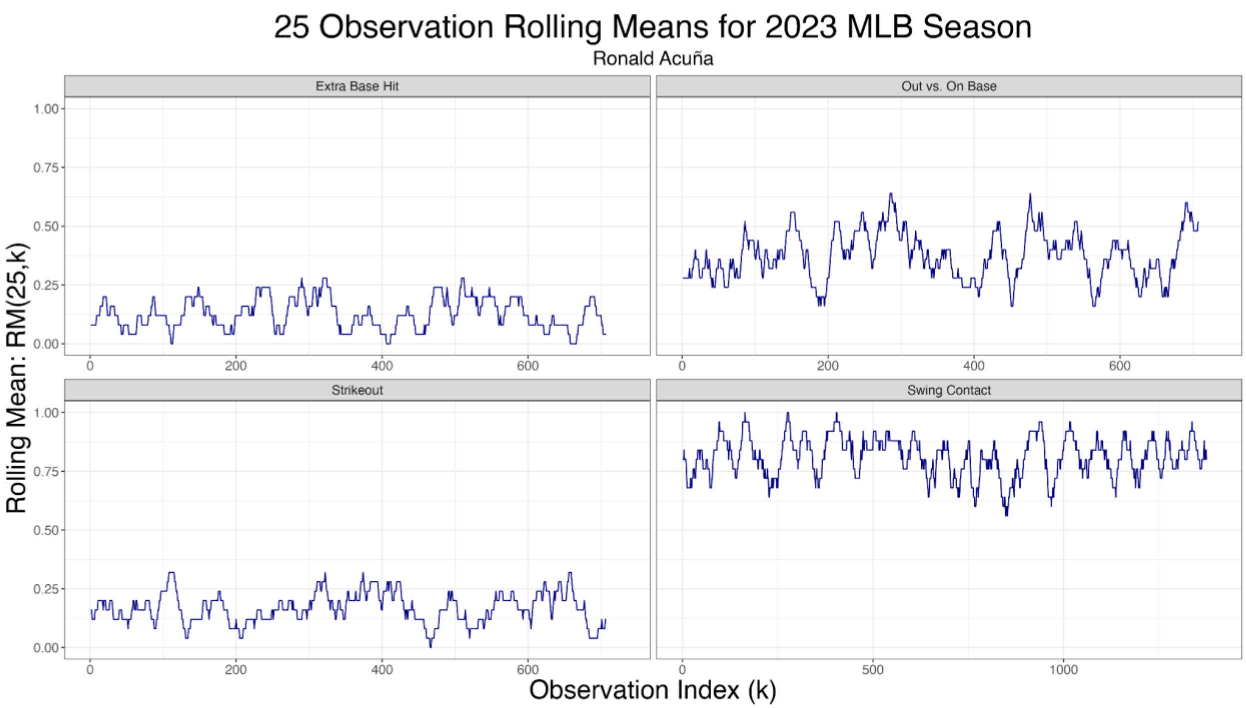

Figure 1: 25 observation rolling mean $RM_i(25, k)$ for various outcomes for Ronald Acuña during the 2023 MLB season

Before introducing these two metrics, some additional notation is needed. Let

$$S_i(a) = max_{b>a}\,(b-a)\prod_{j=a}^{b} 1[Y_{ia} = Y_{ij}]1[Y_{ia} \neq Y_{i(a-1)}]$$

$$G_i(a) = max_{b>a}\,(b-a)\prod_{j=a}^{b} Y_{ia}(1 - Y_{ib})$$

$S_i(a)$ denotes the length of a streak of either successes or failures beginning at observation a. Notice that the product term will be 0 once an observation Y_{ib} not longer equals the starting observation Y_{ia}. Additionally, the product will be 0 if $Y_{ia} = Y_{i(a-1)}$, that is observation a is not the start of a streak. This forces $S_i(a)$ to be 0 for observations in the middle of streaks of consecutive successes or failures, which will make it convenient for defining player metrics below. By a similar notion, $G_i(a)$ denotes the gap between consecutive success – that is the number of 0s between a 1. Note that if observation $Y_{ia} = 0$, then $G_i(a)$ is defined to be 0 because that observation is itself in the gap between two consecutive successes. If Y_{ia} is a success, $G_i(a)$ will be positive until the next success Y_{ib} resets the term to be 0.

Using this notation, we can define the two streakiness metrics of interest. The first metric, which we refer to as Streak Score, is defined as the sum of squared streak lengths $S_i(a)$ of both zeros and ones, for non-overlapping streaks. The second, which we call Spacing Score, is defined by taking the number of 0s between consecutive ones $G_i(a)$, squared, then summed. Spacing Score has been used by Albert[4,5] to target streakiness in rare outcomes, where the Streak Score would be dominated by streaks of consecutive failures, and consecutive successes are very uncommon. In mathematical notation, these metrics can be expressed as follows.

$$Streak\ Score_i = \sum_{j=1}^{n} S_i(j)^2$$

$$Spacing\ Score_i = \sum_{j=1}^{n} G_i(j)^2$$

These metrics give a notation of absolute streakiness. To get at relative streakiness – that is, how streaky was player *i* relative to expectation by random chance, given their own success rate – we utilize permutational inference. Because we are looking at multiple outcomes which are likely correlated, and more granular outcomes than have been the focus of previous work, where say pitches in a sequence are unlikely to be independent, more care is required when doing sampling. A full description of the permutational test procedure and sampling scheme is outlined in Section 2.3.

2.2) Outcome Definitions

We chose to study four outcomes, analyzing each outcome from the lens of both hitters and pitchers. Those outcomes were: on base events, extra base hits (XBH), strikeouts, and swing contact. Each of these outcomes was turned into a binary sequence of ones and zeros, with ones denoting the outcome of interest. The first three outcomes were analyzed at the plate appearance level (i.e. a one or zero for each plate appearance) except swing contact, which was at the swing level. Outcome definitions are provided in Table 1, below.

Multiple outcomes were chosen because there are many ways that players can have success, and failure to look at players holistically might cause one to miss the full picture. For example, Luis Arraez and Matt Olson were two of the best players in the National League this year, but Arraez was noted for his contact and plate discipline while Olson was noted for his power but also his propensity to swing and miss.

Streak Score was used to analyze each outcome with the exception of extra base hits, which we feel is more appropriately measured by Spacing Score, due to the fact that hitters may typically go long periods of time without an extra base hit. Note that while in Figure 1, the rate at which Ronald Acuña got extra base hits in 2023 was similar to the rate at which he struck out, this is generally not the case (and part of the reason he won MVP in 2023). Given that the proportion of plate appearances ending in strikeouts will generally be higher than the proportion ending in extra base hits, and the fact that we will analyze each outcome from the perspective of pitchers as well, where strikeout rate should be higher than for hitters, we feel that strikeouts are not sufficiently rare to require use of the Spacing Score metric.

Outcome	Granularity	Definition	Analysis Technique
On-base Event	Plate Appearance Level	1 if plate appearance ended in hit or walk 0 if plate appearance ended in anything else	Streak Score
Extra Base Hit	Plate Appearance Level	1 if plate appearance ended in double, triple or home run 0 if plate appearance ended in anything else	Spacing Score
Strike Out	Plate Appearance Level	1 if plate appearance ended in strikeout 0 if plate appearance ended in anything else	Streak Score
Swing Contact	Swing Level	1 if swing made contact with the ball (even fouls) 0 for swing and miss	Streak Score

Table 1: Outcome definitions and analysis methods

2.3) Permutational Inference and Sampling Scheme

To quantify whether players reached their final season statistics in a manner that was streaky, we ran a permutation test for the sequences of their outcomes. That is, we shuffled their individual sequences (the collection of zeroes and ones) 1,000 times and then computed Streak and Spacing Scores under each of these 1,000 permutations. Due to the fact that we were looking at multiple outcomes at the same time, and the fact that granular outcomes at the swing level or even at-bat level outcomes for pitchers are not independent due to their dependence on game state, we could not simply use naive shuffling methods like those used in previous works[3-6,14-17].

Instead, we used block permutation[18], permuting innings for at-bat level outcomes when analyzing pitchers (keeping the ordering of at-bats within an inning fixed) and permuting at-bats when analyzing swing-level outcomes for both batters and pitchers (keeping the ordering of pitches in with an at-bat fixed). When analyzing at-bat level outcomes for hitters, we did use naive permutation under the assumption that consecutive at-bats for a hitter are sufficiently independent. There may still be some dependence on game-state for hitter at-bat level outcomes, but we are of the opinion such dependence is far greater for pitchers, hence the different sampling scheme. A graphical overview summarizing all permutation methods used in this work is provided in Figure 2.

Figure 2: Summary of permutation scheme

Upon applying the appropriate sampling scheme, taking the percentile of where the observed Streak or Spacing score was in relation to the permutation gives rise to a notion of internal consistency/streakiness. In other words, what percentage of the 1,000 simulated sequences had a smaller Streak, or Spacing, Score than that of the actual player. Formally, let $Streak\ Score_i$ denote a players observed Streak Score for a given outcome and

$Sim\ Streak\ Score_{ik}$ denote the streak score computed on permuted sequence k. We compute the Internal Streakiness Percentile (ISP) for player i as

$$ISP_i = \frac{1}{1{,}000}\sum_{k=1}^{1{,}000} 1[Sim\ Streak\ Score_{ik} \leq Streak\ Score_i]$$

with an analogous definition for outcomes utilizing the Spacing Score.

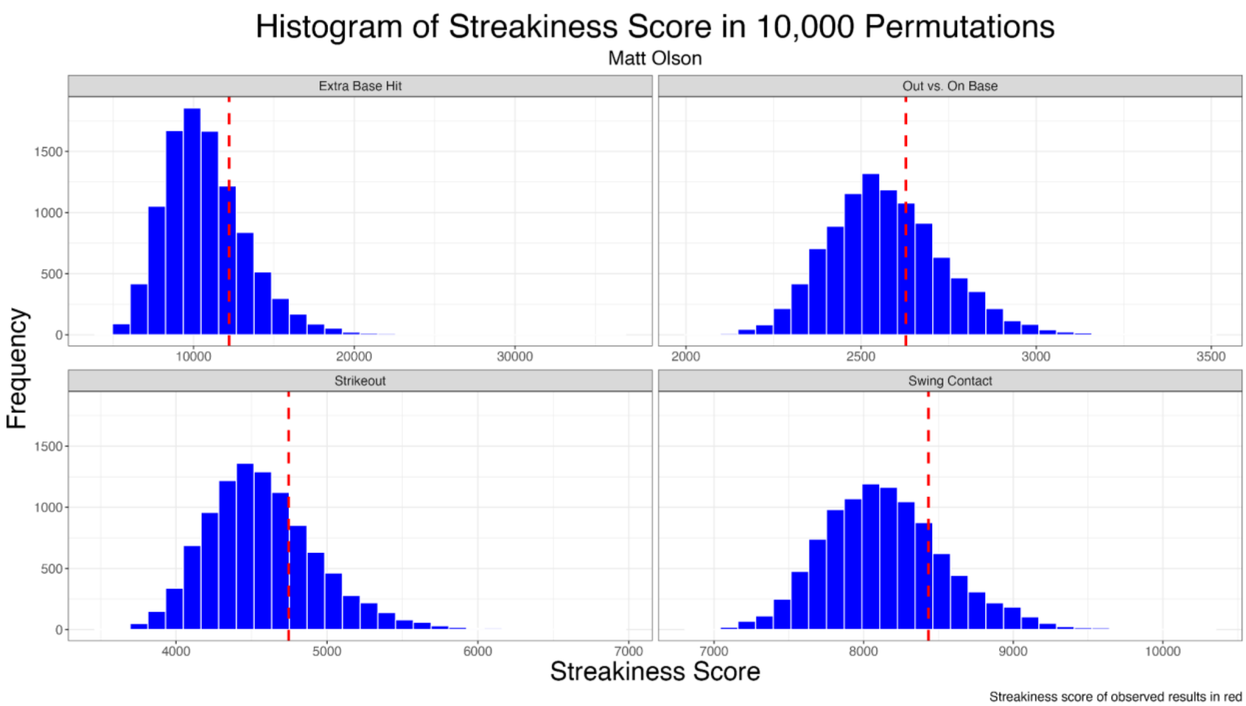

Figure 3: Distribution of Matt Olson streak/spacing score under 10,000 permutations

Percentile numbers closer to zero mean that the player is more consistent while a number closer to one (100%) would mean that the player is streakier. Note that we call this percentile internal streakiness because we are only comparing each given player to permuted versions of himself, rather than re-sampling or permuting other players' outcomes. Figure 3 shows an example histograms $Sim\ Streak\ Score_{ik}$ and $Sim\ Spacing\ Score_{ik}$ across the four different outcomes of interest, using Matt Olson as an illustrative example to give insight into the intermediate steps of how Internal Streakiness Percentile is calculated.

2.4) Data

At-bat and pitch level outcomes were scraped from MLB Statcast[19] and stats for the 2023 MLB season were scraped from Baseball Reference[20], both using the baseballr[21] package in R. Analysis was restricted to batters who compiled at least 500 plate appearances (PA) and pitchers who threw at least 100 innings (IP) during the 2023 season. After applying these cutoffs, our sample totaled 136 batters and 127 pitchers.

3) Simulation Study

In order to demonstrate that permutation-based tests of Streak and Spacing Score capture hitters when they are abnormally streaky or consistent, and don't falsely attribute those extremes due to random chance, we conducted a small simulation study simulation.

We simulated 1,000 seasons of 500 at-bats for three types of hitters (the so-called random hitter, streaky hitter, and consistent hitter), for both common (average = 0.3) and rare (average = 0.1) outcomes. We then applied permutation tests based on Streak Scores (in the case of the common event) and Spacing Scores (in the case of the rare outcome) to each of the 1,000 hitter seasons of each type, and examined the distributions of resulting Internal Streakiness Percentiles.

3.1) Simulating Hitter Types

Outcomes for each of the three hitter types were simulated as follows. For random hitters, each at-bat was independently sampled as $Y_{ij} \sim Bernoulli(0.3)$. For streaky hitters

$Y_{ij} \sim Bernoulli(p_{ij})$ where $p_{ij} = \frac{1}{2}[0.3 + RM_i(25, j - 25)]$ – that is, the hit probability during at-bat *j* was the mean of the baseline batting average of 0.3, and the hitter's rolling batting average in the previous 25 at-bats. While in expectation, the streaky hitter would still have a 0.3 average, there is a lot more variability in at-bat success probability and a high correlation between subsequent observations. Finally, outcomes for consistent hitters were sampled such that each chunk of 25 at-bats had a fixed batting average drawn from *Uniform(0.25, 0.35)* with the appropriate number of hits induced from that batting average randomly dispersed throughout the 25 at-bat chunk. That is to say, the consistent hitter's batting average never dropped below 0.25 or above 0.35 in non-overlapping sequences of 25 at-bats. An analogous simulation mechanism was used for rare outcomes as well, with baseline event rates centered around 0.1 rather than 0.3.

3.2) Simulation Results

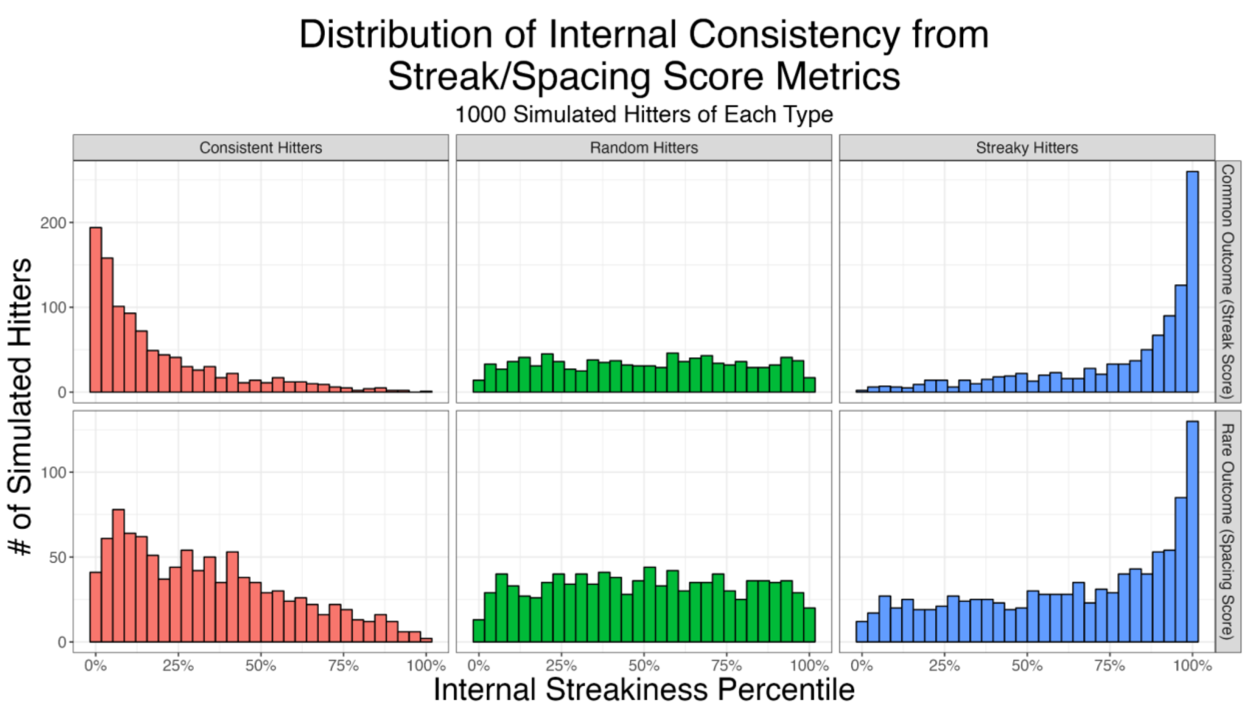

Figure 4: Histogram of ISP for 1000 simulated hitters of varying styles

Histograms of the Internal Streakiness Percentile for simulated hitters are shown in Figure 4. Notably, the distribution of ISP for consistent hitters is heavily right-skewed, with the majority of the distribution closer to 0, while the distribution of ISP for streaky hitters is heavily left-skewed, with the majority of the distribution closer to 1. Since ISPs towards 1 (100%) indicate evidence of streaky performance, and ISPs towards 0 indicate evidence of abnormal consistency, these results suggest that this method does well distinguishing between various hitter types. Moreover, the distribution of ISP for random hitters is roughly evenly spread out between 0 and 1, like a Uniform distribution, suggesting that this method isn't biased towards one extreme on outcome sequences that are truly random.

4) Results

4.1) League Wide Trends

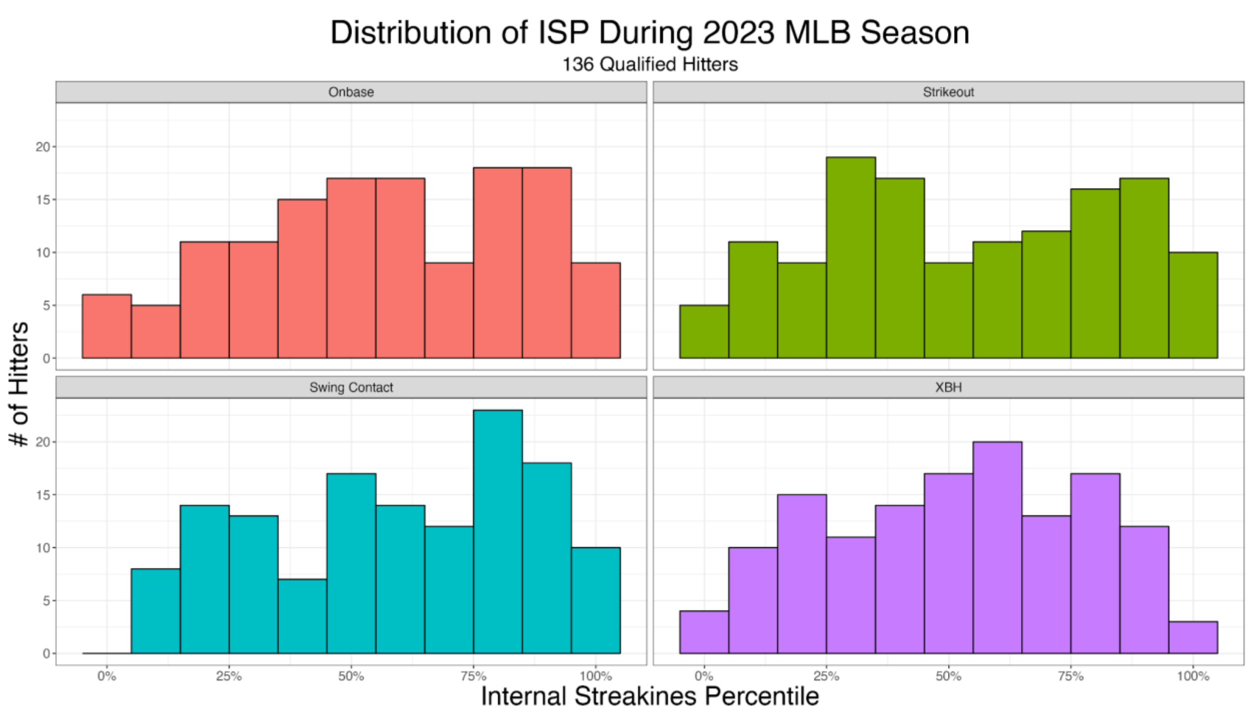

Figure 5: Histogram of ISP for hitters during the 2023 MLB Season

Figure 5 shows the distributions of ISP for all 136 qualified hitters during the 2023 MLB Season. As depicted in the figure, qualified hitters as a whole were more uniformly distributed with respect to XBH streakiness than other outcomes. When looking at the distribution of the XBH panel, and contrasting it with the distribution of the ISPs among random hitters in Figure 4, the charts are extremely similar as players range from being extremely consistent to extremely streaky—covering almost every value in between 0 and 1, with slightly greater concentration in the middle. While certain hitters may be more or less streaky in this metric, it seems that streakiness in extra base hitting from a league-wide perspective may be relatively noisy.

While none of the four metrics fully matched the distribution of the model streaky hitter chart from Figure 4, it is clear from Figure 5 that the distributions of ISP for both swing contact and on base events are more inherently streaky qualities. In both panels, around 35-40 players reside in the 75-100% ISP range. One interesting outcome of this test, however, is that none of the distributions are skewed towards more consistent hitters. This suggests that many of the MLB's most successful hitters (notably, those in this experiment are successful enough to merit playing time across an entire season) are more streaky than consistent in general. This makes sense given the current state of pitching in MLB. In recent seasons, increasing velocity and overpowering junk have made pitching dominant to the point where it is nearly impossible for hitters to succeed at a high level being completely consistent in any particular aspect of the game.

Figure 6, below, depicts the ISP for all 127 qualified pitchers during the 2023 MLB season, and the results are slightly different from the hitters. The results from the XBH and swing contact metrics are similar to those of the hitters, as swing contact seems to be inherently streaky while extra base hits seem to be inherently random. On the other hand, the biggest difference between the two charts is most notable in the distribution of strikeout ISPs.

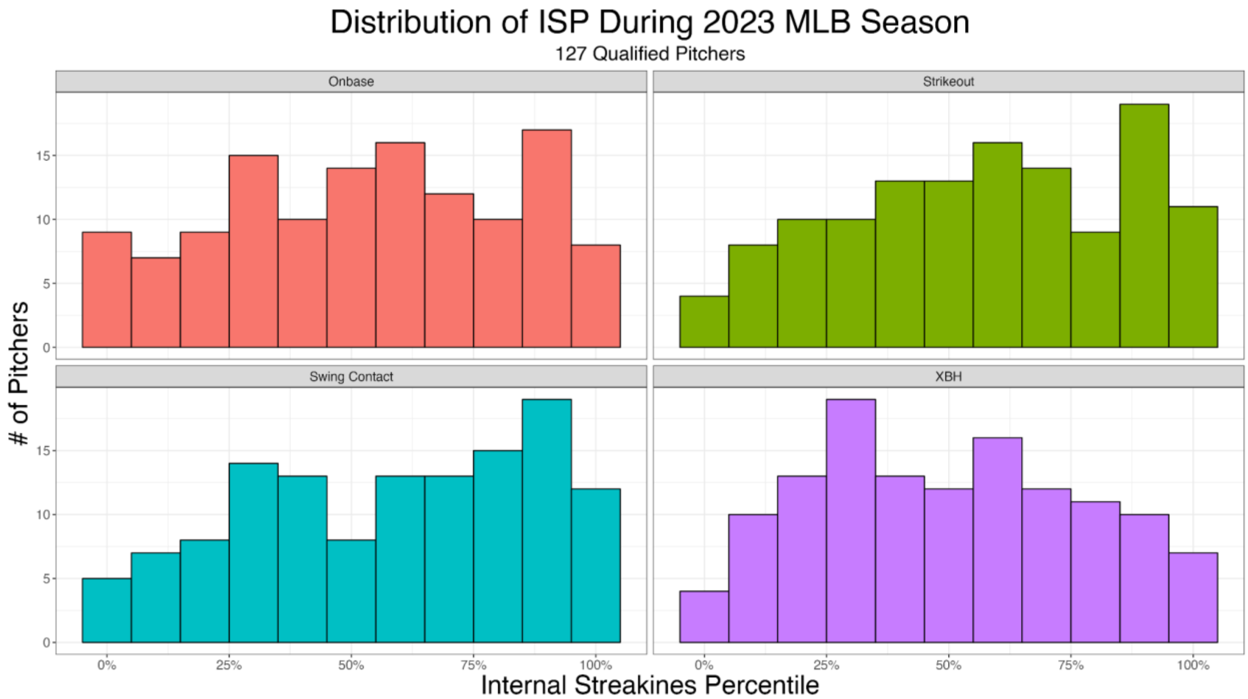

Figure 6: Histogram of ISP for pitchers during the 2023 MLB Season

The distribution of the strikeout chart in Figure 6 suggests that pitchers are more inclined to be streaky with their strikeout patterns. The mode of the distribution for swing contact, on base events, and strikeouts from the pitchers' perspective all reside above 85th percentile, something that wasn't true for any of these same outcomes from the hitting perspective. Once again, this supports basic logic when thinking about the flow of an MLB game or season, because strikeouts and swings and misses (both metrics yielded an inherently streaky distribution) are very indicative of pitcher performance.

When a pitcher is at the top of their game, or has their best "stuff," they will force more swings and misses, and get more strikeouts, while when they are not performing well they will let up more contact—resulting in less strikeouts as well. This line of thinking explains why pitching is more inherently streaky than hitting—which seems more inherently random—as the pitchers largely dictate the outcome of each play.

4.2) Correlation With Traditional Statistics

Figure 7: Correlation between ISP and observed traditional baseball statistics

Figure 7 displays the correlation between the derived ISPs for each of the four outcomes that we tested along with players' actual statistics from the 2023 season. Correlations between ISPs for pitchers were all positive, ranging from 0.05 to 0.36. For hitters, correlations between ISPs ranged from -0.12 (between swing contact and extra base hits) to 0.36, though the majority of correlations between hitting ISPs were somewhat weaker than between pitching ISPs. This is not surprising given that we thought pitching outcomes would be more correlated than hitting outcomes. Nevertheless, the fact that no correlations are too large between pitching outcomes suggests the block permutation scheme outlined in Section 2.3 is reasonable.

The fact that the only negative correlation between ISP metrics is between swing contact and extra base hits seems to get at the distinction between contact and power hitters. That is, being consistent at contact is correlated with being more streaky when trying to hit for power, suggesting that the types of swings needed to yield consistent contact come at the expense of consistent power, and vice versa.

When looking at the associations between ISPs and traditional statistics, one of the trends that immediately jumps out from the pitchers' point of view is how ground ball to flyball rates correlate with XBH streakiness (correlation = -0.25). One interpretation of this, and the most likely one at that, is that flyball heavy pitchers are much streakier with the XBH they give up. This can likely be credited to the fact that flyball pitchers give up harder contact when they don't have their best stuff, leading to more extra base hits—however the correlation to home runs per 9 innings and XBH ISP is almost three times lower than that of XBH ISP and groundball/flyball rates.

For other pitcher statistics such as ERA or XBH, larger values indicate worse performance, so positive correlations between ISP and those metrics are suggestive or worse performance. The opposite is true for hitting metrics, where in general, larger values suggest better performance.

When examining Figure 7, more of the larger correlations between ISPs and traditional statistics seemed to happen for pitchers. This especially stood out when looking at the ISP strikeouts and onbase metrics, as well as the entire XBH row for pitchers. The correlations for example of 0.12, 0.18, and 0.19 are higher than most hitting correlations (in magnitude) and indicate that the streakier pitchers may be slightly less successful than more random or consistent ones.

By contrast, hitter streakiness, in any of the metrics, does not seem to have as strong of a general trend of correlation with stats that indicate success—such as strikeouts, homeruns, extra base hits, or on-base plus slugging (OPS). The most notable part of the hitting data, does relate to XBH streakiness, however, because it seems to be that the more consistent XBH hitters get more XBH on average (as logic would suggest), which leads to both higher slugging percentages and OPS.

Though correlations were relatively moderate for both hitters and pitchers, our findings seem to suggest that in general consistency was associated with success for both player types, with relationships being slightly stronger among pitchers.

4.3) Analysis of Individual Hitters

Figure 8 below shows hitter level ISPs for the top 20 batters by OPS during the 2023 MLB season. As shown in the graph below, there is huge variability in how players reach their own results. From the previous section, we saw that swing contact and on base events tend to be the two of the streakier metrics, while strikeouts and extra base hits tend to be somewhat less streaky. Those results are even more noticeable when plotting ISPs for select hitters.

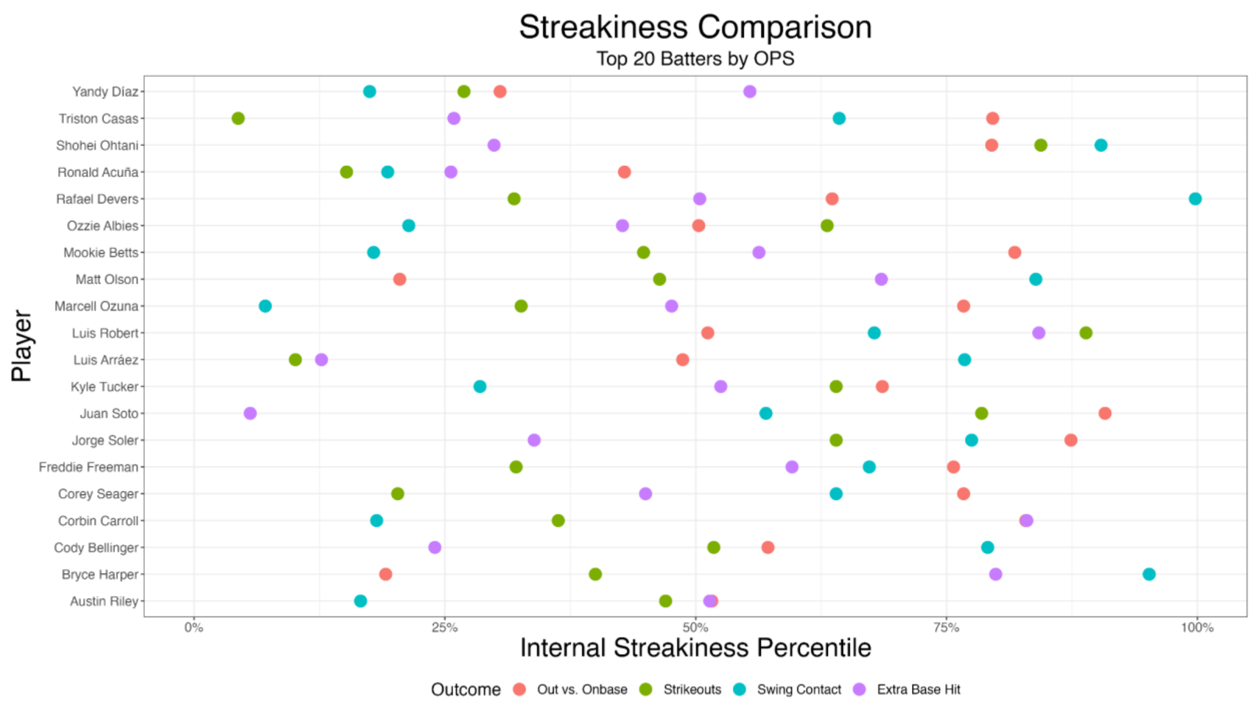

Figure 8: Internal streakiness for top hitters

Some players, like Shohei Ohtani, and Luis Robert are streakier than others as they both have at least three of their four metrics higher than the 70th percentile. Rafael Devers may be an even more interesting case considering that his swing contact is incredibly steaky all the way at 100th percentile while a seemingly related metric, strikeouts, is down just below the 32th percentile and his on-base metric lies around the 50th percentile.

On the flip side, the data indicates that Ronald Acuña’s 2023 season was abnormally consistent throughout all of his metrics as his percentile numbers rank in the top three lowest for each statistic among the best hitters analyzed. Furthermore, Acuña’s most streaky ISP of 0.429 (on-base events) was the minimum “most streaky” metric among any top hitters. Perhaps this is not surprising, as he was in the thick of the most valuable player (MVP) conversation the entire season, and ultimately was named the league's MVP[22-23]. Other notable extremes in terms of consistency include Bryce Harper (on base events), Triston Casas (strikeouts), Rafael Ozuna (swing contact) and Juan Soto (extra base hits).

4.4) Analysis of Individual Pitchers

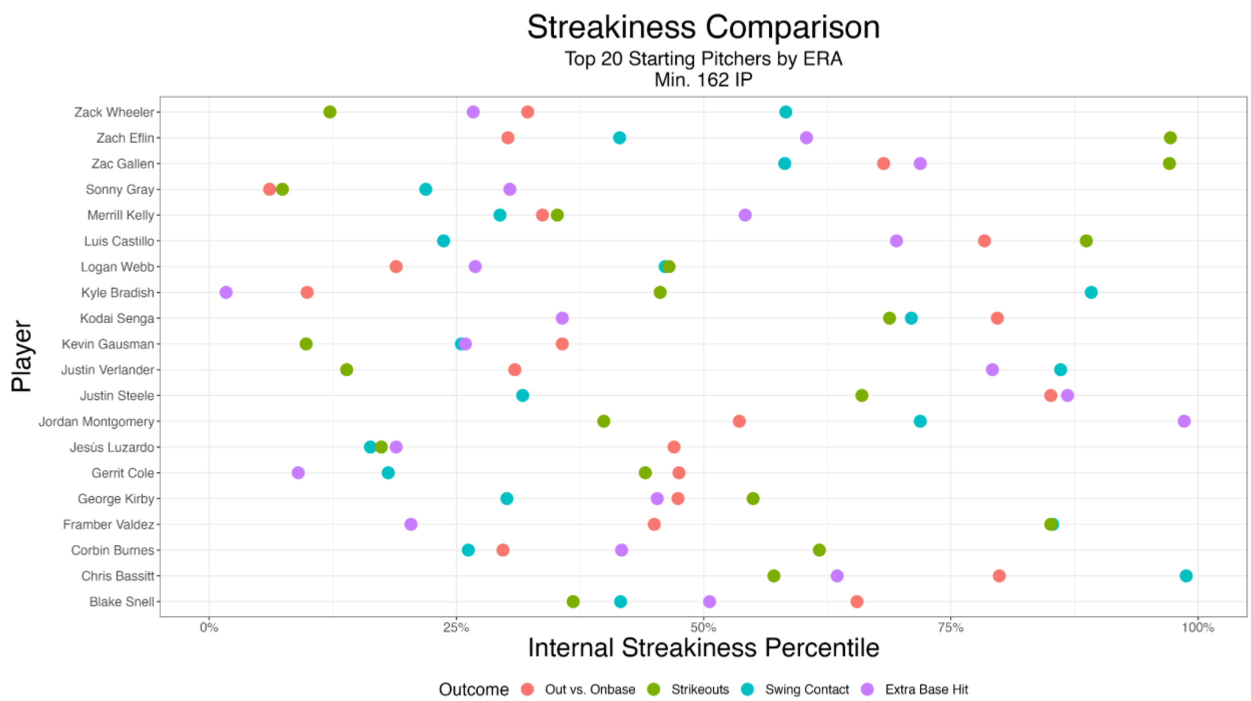

Figure 9: Internal streakiness for top starting pitchers

Like hitters, the top MLB pitchers also have very streaky aspects to their game. As with hitters, extra base hits tend to be among the least streaky outcomes for pitchers, while swing contact appears to be the most streaky. On base events appear less streaky on average for top starting pitchers than for hitters, as Figure 9 shows the ISPs for the top 20 starting pitchers by ERA during the 2023 MLB season.

One of the more interesting outcomes of this data is how XBH streakiness among top starting pitchers has an extreme case at both tails of the distribution, as Kyle Bradish is in the 2nd percentile while Jordan Montgomery is all the way up in the 99th percentile of internal streakiness. This is even more interesting considering that their GB/FB ratios—something we found to be more correlated with XBH ISP—were only .05 apart.

Another interesting datapoint from the chart above is that more top pitchers seem to be abnormally consistent vs abnormally streaky. There are only two pitchers where all four metrics are above the 50th percentile, while five different pitchers have all four below the 50th percentile. This matches the findings from Section 4.2 and Figure 7, where we found consistent pitching to be correlated with traditional measures of success, including ERA, the selection criteria used to compare pitchers in Figure 9.

Perhaps the most consistent of these select pitchers was Jesús Luzardo, who had three of his metrics come below the 20th percentile. Additional outliers for players metrics of note include Chris Bassit (99th percentile swing contact), Zach Eflin and Zac Gallen (97th percentile strikeouts), Sonny Gray (6th percentile on base events), and Kyle Bradish (10th percentile on base events).

While Shohei Ohtani did not meet the 162 innings pitched cutoff required to appear on Figure 9, he is a very interesting data point in our study because he uniquely serves as both a hitter and as a pitcher. Interestingly, Ohtani was slightly more consistent as a pitcher than as a hitter for every single outcome of interest, in general contradicting league wide findings that pitching was streakier than hitting, particularly on swing contact and strikeout metrics. Perhaps this explains why Ohtani was a serious contender for both the Cy Young award and the MVP prior to a suffering a midseason elbow injury, which prevented him from pitching during the second half of the season.

Figure 10: Internal streakiness for Shohei Othani

5) Discussion

This work sought to add to the long-standing statistical interest in streaks in baseball by examining streakiness and consistency for both hitters and pitchers simultaneously across outcomes more granular than those traditionally studied in the literature. Furthermore, a primary goal of this work was to understand the degree to which streakiness contributed to player success amongst MLB's top players. Our findings indicated great deals of heterogeneity in streakiness across individual players and outcomes, with swing contact and on base events generally being streakier events than one might expect for both hitters and pitchers due to randomness alone.

We found moderate but notable correlations between internal streakiness percentiles and traditionally studied baseball statistics, with ISPs indicating more consistency generally associated with more successful values of these canonical measures. These relationships were stronger for pitchers than they were for hitters, especially amongst top players in the league, which may speak to the fact that pitchers have more control over the outcomes we chose to study than hitters, and the current quality of pitching in modern day MLB.

Finding streaky aspects to various hitting metrics is not a novel finding, and only confirms work by Albert and others for on base events[3,6-7] and home runs[4,5] (proxied here by extra base hits). On the other hand, finding some evidence of streaky pitching is much more interesting given the current state of the literature. While (Arthur and Matthews, 2017)[13] found evidence that pitchers existed in 3 different streaky states via fastball velocity, (Evanko, 2020)[14] and (Gamble, 2015)[12] found no evidence of streakiness among pitchers. Perhaps the reason we are able to find some evidence of streakiness owes to the fact that our work looks at more granular outcomes than (Evanko, 2020)[14] and (Gamble, 2015)[12], at roughly the pitch level similar to (Arthur and Matthews, 2017)[13]. Furthermore, our improved block permutation scheme may have improved our power to detect some streakiness among pitchers. Prior versions of this work, which did not use the permutation scheme outlined in Section 2.3 did not find as much evidence of extreme streakiness on certain outcomes.

Perhaps the most noticeable aspect of this study was Ronald Acuña's remarkable consistency. Being the heavy favorite for NL MVP with the MLB's first ever 40 home run and 70 stolen base season, Acuña was far in away the most consistent hitter from a holistic perspective among the top hitters studied. His ranking inside the top 3 lowest ISP among hitters in 8 shows just how productive he has been at every step of the 2023 season, across a range of metrics. The raw numbers reflect this too as he had a batting average between .326 and .356 in five of his six months during the season (April/March and September/October are combined). Simply studying a single outcome may have missed the degree of universal consistency which made his 2023 so special.

There are a few limitations of this work worth mentioning. While block permuting outcomes surely preserves much of the dependence these outcomes may have on game state, particularly for pitchers and pitch level outcomes, there may still be some residual dependence across permuted blocks. Additionally, because we are only comparing players to permuted versions of themselves, it is somewhat difficult to draw comparisons between players. One future step that could address both of these steps would be to resample (i.e. bootstrap) outcomes from other players in similar game states, as has been used in the analysis of football[24-25] . Doing so would answer a slightly different question than studied in this paper, namely how streaky/consistent a player is relative to an average MLB player rather than to themselves. Such a question is certainly of interest but distinct from the primary questions explored in this work.

Finally, this study focused on very short term outcomes, either at the swing level or plate appearance level. These outcomes are inherently noisier, so considering additional metrics like rolling averages to better capture long term streaks is another possible extension., especially for pitchers, where much less work on streakiness has been conducted. A long term vision for this work may be the creation of some Baseball-Savant[19] style dashboard which breaks down streakiness for all players across many outcomes, utilizing both the notions of ISP considered in this work and streakiness relative to league average, as suggested in the preceding paragraph.

Overall, we feel that holistic evaluation of player streakiness offers the best way to understand how such streaks underlie player success. Identification of extreme consistency for a unanimous MVP suggests that this work is doing something right. Much work lies ahead to keep unlocking better overall understandings of why certain players are more consistent than others, but this work is an important first step.

Data and Code Availability

Data and code are made available on GitHub at https://github.com/c25rf/MLB-Streak-Project/tree/main

References

1. Miller, J., & Sanjurjo, A. (2018, March 28). *Momentum Isn't Magic—Vindicating the Hot Hand with the Mathematics of Streaks.* Scientific American. https://www.scientificamerican.com/article/momentum-isnt-magic-vindicating-the-hot-hand-with-the-mathematics-of-streaks/
2. Wagner, J. (2014, July 16). *Research supports the notion of of the 'hot hand'; baseball players always believed in it.* Washington Post. https://www.washingtonpost.com/sports/nationals/research-supports-the-notion-of-the-hot-hand-baseball-players-always-believed-in-it/2014/07/16/5a70653e-0cf9-11e4-b8e5-d0de80767fc2_story.html
3. Albert, J. (2008). *Streaky Hitting in Baseball*. Journal of Quantitative Analysis in Sports, 4(1), Article 3.
4. Albert, J. (2013). *Looking at spacings to assess streakiness.* Journal of Quantitative Analysis in Sports, 9(2).
5. Albert, J. (2014). *Streakiness in Home Run Hitting*. Chance. https://chance.amstat.org/2014/09/streakiness/
6. McCotter, T. (2008). *Hitting Streaks Don't Obey Your Rules: Evidence That Hitting Streaks Aren't Just By-Products of Random Variation*. Baseball Research Journal.
7. Albright, S. (1993) *A Statistical Analysis of Hitting Streaks in Baseball*. Journal of the American Statistical Association. 88(4), 1175-1183.
8. Bock, J. R., Maewal, A., & Gough, D. A. (2012). Hitting is contagious in baseball: evidence from long hitting streaks. PloS one, 7(12), e51367. https://doi.org/10.1371/journal.pone.0051367
9. Reifman, A. (2012). *Hot Hand: The Statistics Behind Sports' Greatest Streaks*. Potamac Books.

10. Bar-Eli, M., Avugos, S., & Raab, M. (2006). *Twenty years of "hot hand" research: Review and critique.* Psychology of Sport and Exercise, 7(6), 525-553.
11. Cohen, B. (2020). *The Hot Hand: The Mystery and Science of Streaks*. Custom House.
12. Gamble, R. (2015, April 8) *Does A Pitcher's Last 3 Start Performance Matter For Projecting Their Next Start?* Razzball. https://razzball.com/pitcher-streakiness/
13. Arthur, R. and Matthews, G (2017, August 11). *Baseball's Hot Hand is Real*. FiveThirtyEight. https://fivethirtyeight.com/features/baseballs-hot-hand-is-real/
14. Evanko, C. (2020). *Throwing Heat: Is There a "Hot Hand" Among Baseball's Top Pitchers?* Undergraduate thesis, Princeton University. https://dataspace.princeton.edu/handle/88435/dsp015h73q008w
15. Albert, J. (2016). *Dustin Pedroia's Hit Streak*. Exploring Baseball Data with R. https://baseballwithr.wordpress.com/2016/09/06/dustin-pedroias-hit-streak/
16. Albert, J. (2018). *Statcast Streakiness?* Exploring Baseball Data with R. https://baseballwithr.wordpress.com/2018/03/12/statcast-streakiness/
17. Albert, J. (2020). *Statcast Streakiness, Part II*. Exploring Baseball Data with R. https://baseballwithr.wordpress.com/2020/10/12/statcast-streakiness-part-ii/
18. Baranger, D. (2019, November 14). *A quick intro to block permutations and bootstraps for analyzing hierarchical data.* The Startup. https://medium.com/swlh/an-quick-intro-to-block-permutations-and-bootstraps-for-analyzing-hierarchical-data-d219b319ef55
19. *Baseball Savant: Statcast, trending MLB players and visualizations*. https://baseballsavant.mlb.com/
20. *Baseball Reference.* https://www.baseball-reference.com/
21. Petti B, Gilani S (2023). *baseballr: Acquiring and Analyzing Baseball Data*. https://github.com/BillPetti/baseballr.
22. Petriello, M. (2023, September 26). *The striking stat fueling Acuña's historic campaign*. MLB. https://www.mlb.com/news/ronald-acuna-jr-historic-strikeout-rate-improvement
23. Petriello, M. (2023, August 29). *Mookie's case to take the NL MVP over Acuña.* MLB. https://www.mlb.com/news/mookie-betts-case-for-2023-nl-mvp-award
24. Lopez, M. (2019). *Estimating NFL drive outcomes under rules that don't exist.* https://statsbylopez.netlify.app/post/resampling-nfl-drives/
25. Elmore, R. (2022) *NFL SimulatoR* https://github.com/rtelmore/NFLSimulatoR