By the Numbers

Review

Academic Research, 2018-19

Charlie Pavitt

Reviews of academic studies from the past year. Subjects include: effects of borderline pitches, heat-aggression induced HBPs, MLB ownership changes, post-contract shirking, and the tendency of Canadians to bat left.

Michael William Gmeiner (2019), <u>History-dependent mixed strategies: Evidence from major league baseball</u>, Journal of Sports Economics, Vol. 20 No. 3, pp. 371-378.

Based on PITCHf/x data on almost six million pitches from 2007-2015, Gmeiner examined the cat-and-mouse game between batter and pitcher. This is a very dense article with a ton of reported findings, some of which are contradictory depending on the complexity of the statistical analysis performed (i.e., number of predictor variables in the equations).

Gmeiner found that batters are more likely to swing after a borderline called strike than a borderline ball. This is true regardless of whether the next pitch is in or close to the strike zone (specifically, an area with a greater than 50 percent of being called a strike), or not. In the first case, when the pitch is likely a strike, the swing rate is 74.6 percent after a previous borderline strike call, but 63.2 percent after a previous borderline ball call. When the following pitch is likely a ball, the corresponding swing rates are 31 percent and 27.9 percent, respectively.

In response, pitchers are more likely to throw outside the strike zone after a borderline strike than after a borderline ball — 42.0 percent versus 54.5 percent, respectively. In the former of these, whether the strike is taken (42.8 percent) or swinging (38.3 percent) does not matter a lot. Incidentally, Gmeiner found that umpires don't change their strike zone after a borderline call (work by Jon Roegele back in 2013 contradicts this claim).

In addition, it is unclear whether batters are more likely to swing following a called or swinging strike, as Gmeiner's findings are dependent on the number of predictor variables included in the analysis. In any case, it is likely to be tiny, perhaps a one percent difference.

The count after a borderline strike matters. A borderline strike call resulting in a two-strike count increases the likelihood of batters swinging at obvious strikes but decreases that probability for obvious balls. However, if a borderline strike call leads to a one-strike count, the likelihood of swinging is lower no matter where the pitch is. Again, I think this only makes sense if batters are expecting that next pitch to be a ball when there is one strike. However, the tendency for pitchers to throw a ball after a borderline called strike is not dependent on number of strikes, with only a tiny one in favor of throwing balls after called rather than swinging strike. Finally, 83.7 percent of first strikes are called, but that drops to 53.7 percent for second strikes. That probably reflects the tendency for batters to take the first pitch.

Gmeiner continued with a very interesting analysis of strategic issues for hitters using run expectancy figures from Tango/Lichtman/Dolphin's "The Book."

On a one-strike count, when that first strike was a no-swing borderline call, run expectancy for swinging at an obvious ball is −.024 runs, and +.027 for an obvious strike. That means the breakeven point for swinging at that next pitch is an 8:9 ratio of strikes to balls. After a borderline pitch called a ball when there is already one strike; the analogous numbers are −.063 and +.061.

As the downside is greater than the upside, the implication is that batters should never swing on pitches after borderline balls with one strike. Two strike counts are far more conducive to swinging (not surprisingly), with the figures –.093 and +.295 after a previous borderline strike (a breakeven of 24 percent), and –225 and +.438 after a borderline ball (breakeven 34 percent). The pitcher's response concerning the value of trying to throw a strike is quite complicated, taking into consideration the run expectancy for the batters as just described plus the probability that the batter will or not swing; see the article for details.

William L. D. Krenzer and Eric D. Splan (2018), <u>Evaluating the heat-aggression hypothesis: The role of temporal and social factors in predicting baseball related aggression</u>, Aggressive Behavior, Vol. 44 No. 1, pp. 83-88.

It is becoming pretty clear that hotter weather is associated with more hit batsmen, perhaps due to discomfort increasing pitcher aggressiveness.

The previous pioneering study was by Reifman, Larrick and Fein (1991); based on 829 randomly chosen games from 1986 through 1988, these researchers discovered that pitcher wildness (both walks and wild pitches) and temperature increased hit by pitches per game. They found a J-shaped relationship, i.e., increasing ever-faster as temperature goes up.

Krenzer and Splan (2018) continued in that vein using 2000-2015 Retrosheet data, noting both temperature and, more importantly, wildness as predictors of HBPs. After dividing the season into fifths based on games played, they observed the correlation only occurs during the middle three-fifths -- the warmer months -- implying a probable threshold temperature effect before aggression steps in.

Further, HBPs were greater within division than otherwise -- which is what you'd expect, given that within division is where the best rivalries lie. They were also higher in blowouts than in one-run games. Unfortunately, the authors did not analyze winner versus losers separately, which leaves the question open as to whether the cause is the losing team's frustration.

And, for some reason, HBPs were greater against visiting teams with better records. Why, and why not better home teams also?

Heather M. O'Neill and Scott Deacle (in press), <u>All out, all the time? Evidence of dynamic effort in major league baseball</u>, Applied Economics, DOI: 10.1080/00036846.2019.1591598

The best research implies that position players facing free agency perform on average a tiny bit better – perhaps a run or two's worth – during their contract season and revert to normal after their signing. Heather O'Neill has been involved in this work, and has strived to improve her efforts over the years. She summarized one of her studies in the 2014 Baseball Research Journal, but, (as Phil Birnbaum noted in the May, 2015 issue of BTN) failed to control for careers ending at the conclusion of contracts. Her latest work took care of that problem, but is further plagued with another -- the failure to sufficiently control for player career trajectories.

The sample consisted of players eligible for free agency (6+ years) who played for at least two seasons between 2007 and 2011 -- a total of 225 players and 822 player-seasons. Without going into too much detail, here are O'Neill's estimates for differences between predicted and obtained OPS+ for position players overall, for those in the top quartile and bottom quartile, for their contract year and the subsequent season depending on the length of their new contract:

	Contract year	1-year contract	2-year contract	3-year contract	4-year contract	5-year contract	6-year contract
Overall	+6	+12	+3	0	−3	−6	−8
Top quartile	+7	+15	+6	+5	+3	+1	−1
Bottom quartile	+2	+1	−7	−14	−20	−27	−34

The overall effect remains as tiny, consistent with work. Additionally, it looks like players, during the first year of a new contract, shirk a bit more the longer that contract lasts. But check out the distinction between quartiles. It looks as if the strongest hitters are inspired by their new contract whereas the weaker ones shirk significantly that first year after signing contracts of any considerable length.

O'Neill's interprets these data along the extra effort shirking lines. That could well be correct, but I am not at all convinced, and the quartile distinction reveals why. As I noted, O'Neill failed to sufficiently control for player career trajectories. Although there are problems with

relevant existing studies, there is evidence[1] that stronger players enter the majors a couple years earlier than weaker ones. Putting the two together, I hypothesize the following alternative:

1. The top quartile players hit their contract year a couple of years earlier than the bottom quartile players. Therefore, the former are coming into their peak whereas the latter are at theirs. This is why performance is higher for the top quartile than the bottom.

2. The more that a player randomly happens to overproduce above expectation during the contract season -- in other words, has a well-timed "career year" -- the longer a contract he is able to sign. After that signing, he returns to his normal performance. The greater the overproduction, the greater the fall to normal. That is why productivity is lower the longer the contract.

Heather, if you have the opportunity to read this: you have one more study to do, this time including age and age squared as controls.

John Cairney, Danielle Chirico, Yao-Chuen Li, Emily Bremer and Jeffrey D. Graham (2018), <u>Are Canadian-born major league baseball players more likely to bat left-handed? A partial test of the hockey-influence on batting hypothesis,</u> PLOS One, Vol. 13 No. 5, Article Number e0195252.

Here's a fun one: Based on their birth place and using every player who appeared in at least one 2016 game as their sample, the authors determined that Canadian-born players were far more likely to be left-handed hitters (69.2 percent if switch hitters are considered lefty; 66.7 percent if not) than U.S., Dominican, or Asian born players (all less than 40 percent lefties).

The difference might be partly a result of small sample size for the Canadian sample, but Canadians remained relatively likely to be lefties (53.2%; 51.7%) going back as far as 1917.

There was no corresponding difference in throwing hand, making nurture a more likely explanation than nature. Their proposed reason; a greater likelihood for Canadian ballplayers to have once played ice hockey, in which the development of a lefthanded shot is encouraged. They have no evidence for their proposal, so a survey of players would be useful here.

Philip L. Hersch and Jodi E. Pelkowski (2019), <u>The consequence (and nonconsequences) of ownership change: The case of major league baseball,</u> Journal of Sports Economics, Vol. 15 No. 1, pp. 72-90.

To the best of my knowledge, this is the first published study on the topic referenced in the article title. Based on forty cases of "full" ownership change — that is, not including either minority owners or family taking over — the authors did not note any impact on team performance. They did note the change increasing the odds of "full" general manager replacement (that is, not including interims or GM promotions) by 177 percent during the first year, although that 177 percent increase still reflects out to only 17 percent of GMs leaving.

Further, there was no relationship between ownership change and GM turnover after the first year. Both team performance and GM experience with the team had bigger negative effects on GM job retention than ownership change. New ownership did lead to a first-year impact on "full" field manager replacement (again omitting interims or promotions), increasing manager change by 206%, this time adding up to a 35 percent overall chance. Not surprisingly, team performance and improvement between seasons also mattered.

Although they didn't analyze the issue explicitly, the authors interpret the results as implying that ownership-induced change on both GMs and field managers go hand-in-hand.

Incidentally, Hersch and Pelkowski referenced a senior thesis by J. Douglas Black[2] as an earlier study of GM turnover, which uncovered both overall and same-team experience as encouraging turnover and team success (both regular season record and making the playoffs) as discouraging it.

Charlie Pavitt, chazzq@udel.edu ♦

---

[1] See the Age section of my Offensive Issues chapter at https://charliepavitt.home.blog/
[2] Available at http://hdl/handle.net/10066/584

Back issues

Back issues of "By the Numbers" are available at the SABR website, at http://sabr.org/research/statistical-analysis-research-committee-newsletters, and at editor Phil Birnbaum's website, www.philbirnbaum.com .

The SABR website also features back issues of "Baseball Analyst", the sabermetric publication produced by Bill James from 1981 to 1989. Those issues can be found at http://sabr.org/research/baseball-analyst-archives.

Submissions
Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, preferably by e-mail. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to Phil Birnbaum, at philbirnbaum@outlook.com.

"By the Numbers" mailing list

SABR members who have joined the Statistical Analysis Committee will receive e-mail notification of new issues of BTN, as well as other news concerning this publication.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail to Phil Birnbaum, at philbirnbaum@outlook.com. If you don't have internet access, we will send you BTN by mail; write to Phil at 88 Westpointe Cres., Ottawa, ON, K2G 5Y8.

WAR, Win Shares, and MVP Leaders by Age

Tom Hanrahan

The author examines the performance of some of the best hittters by age, and notes that Mike Trout is an outlier in all of them. No other leading player has accomplished so much so young.

Among active players, who has had the best career? And who has had the best career through their current age? Here is a brief study, with the key points captured on one chart.

Wins Above Replacement (WAR) has become a popular, often-quoted uberstatistic. Baseball Reference[3] regularly updates its leaderboard pages so we can see at any given time who has the most of what, and WAR is one featured stat. There are other sources of WAR; Baseball Prospectus and Fangraphs each have their own variant. I will use the Baseball Reference calculation.

Win Shares (WS), created by Bill James about two decades ago, has not found the same level of popularity, for a variety of reasons. But it is another well-grounded statistic which measures a player's total contribution toward winning games for his team. The current statistics are not as easy to find, but the website baseball gauge[4] has downloads available which are updated regularly.

These two statistics are both denominated in wins (after dividing WS by three), so they are comparable. WS has a lower replacement level or "floor" than WAR, and so a variant of WS named "Win Shares Above Bench" (WSAB) was created at The Hardball Times[5]. This brings it more in line with the floor used by WAR.

Current Leaders

I recorded the WAR leaderboard for active players as of the end of April 2019, and downloaded the WSAB records in late May.
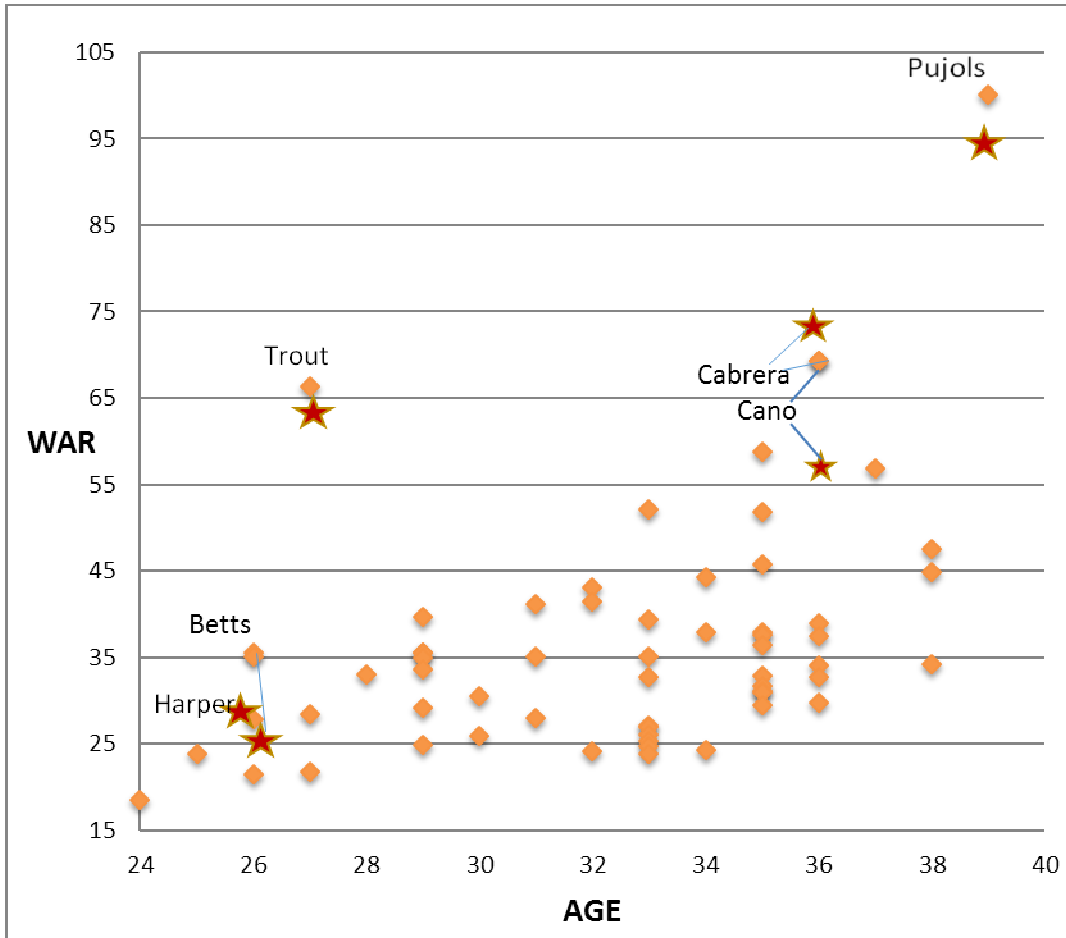
I started with the top 58 position players (not pitchers) by career WAR to date, and then added a few others in the top 70 who were younger than 28 -- Kris Bryant, Jose Ramirez, and Carols Correa. Then I found WSAB for these players.

The diamond-shaped points on the chart on the next page shows the results for career WAR vs. age. Also, the star-shaped points show WSAB (divided by 3) for those players leading in WAR or WSAB for their age.

---

[3] http://www.baseball-reference.com
[4] http://www.thebaseballgauge.com
[5] An explanation of WSAB and historical context is found at https://tht.fangraphs.com/historical-win-shares-above-bench/ .

A few notes:

- In general, correlation between WAR and WSAB is good.

- Miguel Cabrera and Robinson Cano have accumulated equal WAR in their careers to date, but Cabrera is significantly ahead by WSAB.

- Mookie Betts has the most WAR at age 26, but Bryce Harper is slightly ahead in WSAB.

- The chart shows two obvious outliers: Albert Pujols, and Mike Trout.

- Pujols has had by far the most valuable career out of all active players; Trout is a close third behind Pujols and Cabrera, despite being many years younger than either of them.

- Trout's performance has not gone unnoticed; so much has been said and written about him, the Best Player in Baseball since shortly after he arrived. Trout is *so* far ahead of the pace of every other active star that it is almost silly. This exercise shows (as if we needed another reminder) the amazing career Mike Trout has had since 2012.

Now that we've seen Trout's outlying WAR and WSAB performances, let's take a look at his career through the lens of MVP voting.

Trout and MVP

The MLB MVP award has been with us in more or less its current form only since the 1930s. So, we cannot compare old-timers like Wagner, Cobb, Hornsby, and Ruth this way. But we can make side-by-side comparisons from the 1930s on.

The number of MVP votes has changed over time with the expansion of the major leagues, but it is simple to construct a measure which levels the playing field. Baseball Reference reports MVP "award shares" on the player pages; these are calculated by awarding a full share of an MVP award to a player who wins unanimously (he received as many votes as he possibly could). Other players receiving votes are awarded shares calculated as (number of MVP points received) divided by (maximum number of points available).

Currently, the MVP candidates are awarded 14 points for a first-place vote, 9 points for second, and then 8 through 1 point for third through tenth place. There are two voters per team, totalling 30 per league. In the 2018 NL season, the top MVP award shares looked like

Player	First place votes	Total points	Share
Christian Yelich	29	415	.99
Javier Baez	0	250	.60
Nolan Arenado	0	203	.48

For a career, we just add the player's award shares for each season. So, if Javier Baez repeated his .60 each year for ten years, he'd have a total of 6.00.
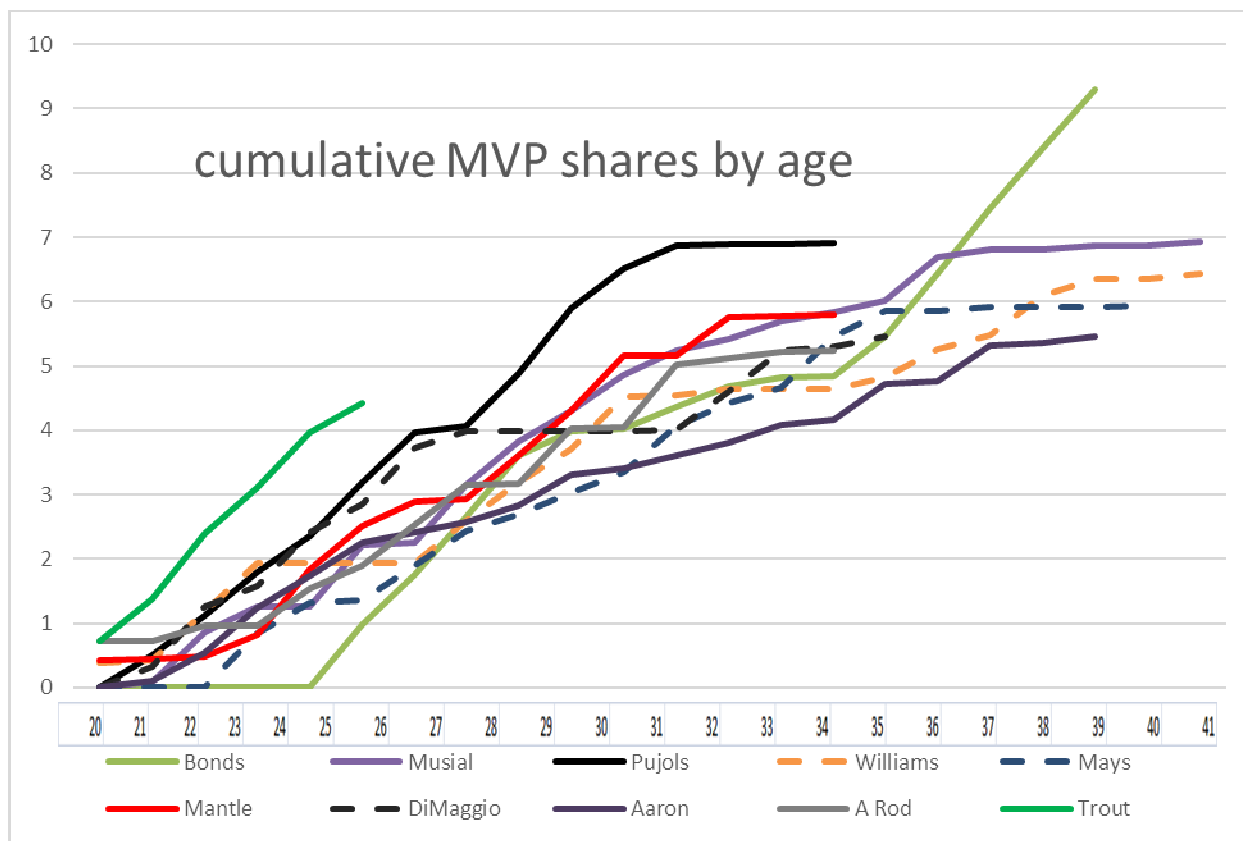
When we do that, the leaders since 1936 turn out to be[6]:

Career share	
9.30	Barry Bonds
6.96	Stan Musial
6.91	Albert Pujols
6.43	Ted Williams
5.94	Willie Mays
5.79	Mickey Mantle
5.45	Hank Aaron
5.45	Joe DiMaggio
5.23	Alex Rodriguez
5.06	Mike Trout

Of course, only two of these players are active, and it seems unlikely Pujols will add much to his career MVP record.

Some of these players picked up MVP award shares throughout their careers (Teddy Ballgame, and Stan the Man). Some were particularly great in their primes (Mantle, DiMaggio). One was famous for his late-career exploits; the likely-chemically-enhanced Mr. Bonds.

---

[6] Courtesy of Baseball Reference; see https://www.baseball-reference.com/leaders/mvp_cya.shtml

Below I show a graph of their cumulative MVP award shares earned by age (as of July 1 of the season in question):

We begin at age 20, as none of these men received any MVP votes before then. Only Williams and Musial got some award recognition at age 41.

General Observations

- Among this group, the ages where most award shares were won were 22 to 32. The true peak years were 23 to 29.

- Bonds is an outlier, both overall and late in his career. Of this group, only two players won MVP awards after age 32; Mays once, and Bonds four times.

- DiMaggio and Pujols both had gained about 5 award shares by age 27. DiMaggio of course then lost 3 seasons to WWII. Up until just the past few years, those three players had the highest MVP recognition at young ages, although Ted Williams was also on fire early in his career before losing his age 24 to 26 seasons to WWII.

Troutian observations

Once again, Mike Trout has lapped the field. Not only does he stand out among current players, he has far-and-away won more MVP recognition at every age through the current (age 26) season than any other player, including the men here who are considered among the best who have ever played the game. Is there any doubt we are witnessing something amazing?

Tom Hanrahan, han60man@aol.com ♦