By the Numbers

Volume 34, Number 1 The Newsletter of the SABR Statistical Analysis Committee March, 2025

Review

Academic Research, 2024

Charlie Pavitt

Charlie reviews several recent studies from the academic literature.

Guérette, Joel, Caroline Blais, and Daniel Fiset (2024), Verbal aggressions against major league baseball umpires affect their decision making, Psychological Science, Vol. 35 No. 3, pp. 288-303.

In the time since pitch location data became available, we have found a number of umpire biases on pitch calls. I refer to the most prominent of these as the "count compensation bias," a tendency to call additional balls as the number of strikes increase, and vice-versa (especially in counts with three balls).

At least two other biases have been discovered that perhaps have on occasion affected game outcomes: a home field advantage in called pitches; and different strike zones for left-handed versus right-handed batters. The latter has perhaps attenuated as umpire's called strike zones have been more closely approximating the rule book strike zone over the past few years.

In this paper, Guérette, Blais, and Fiset may have found a new bias; how significant it is on game outcomes is unclear.

For 2010-2019, they compiled a sample of 153,255 non-swinging pitches from Statcast, Baseball Reference, and FanGraphs, which they then cross-referenced to a list of Retrosheet ejections.

Then, they examined the connection between ejections and pitch calls using a series of models with a large and varying set of controls (including pitch location) and concluded the following:

- The odds of a strike call after a batter or manager/coach was ejected for arguing his batter's strike call were lower for the ejected player's or manager/coach's team, and higher for the other team following a manager/coach ejection.
- Before the ejection, the ejected player/manager/coach's batters were getting more strike calls than the opposition; after the ejection, they received fewer.
- Player status (All-Star) and fWAR had no significant effect, and ejection for other reasons had no comparable impact.

In this issue

The previous issue of this publication was March, 2024 (Volume 33, Number 1).

Patt, Emily-Anne and James Stockton (2024), Noisy judgments: A probability surface-based analysis of umpire variability, MIT Sloan Sports Analytics Conference.

The authors used 5,307,386 pitches called by umpires from 2008-2022 – using raw data from Statcast and umpire game assignments from Retrosheet – to formulate called strike zones (CSZs) for umpires, constructed as probability distributions for calls based on specific pitch location.

They determined that individual umpires' accuracy over those 15 years varied in the range of 70 to 90 percent, with most umpires clustered around 84 percent. They also found that accuracy has steadily increased over time (as we already knew).

Also not news but good to see reiterated: during the span of seasons covered, the CSZ narrowed from 63 to 54 centimeters (24.8 to 21¼ inches), with inside pitches toward righty batters responsible for most of the narrowing. The top of the CSZ went

up 2 cm (¾ inch), whereas the bottom went down 9 cm (3½ inches). The bottom narrowing in particular is one of the causes of the recent surge in strikeouts; see my 2024 BRJ article for that discussion.[1]

Other findings worth noting:

- Batters and pitchers have individual CSZ's independent of their height;
- There is now more evidence supporting the effect of pitch framing (Jose Molina still rules);
- There is a strong tendency to inaccurately call balls with two strikes on the batter. These constitute about half the overall inaccuracies. (The authors did not examine the opposite, inaccurately calling strikes with three balls.)

Brill, Ryan S. and Abraham J. Wyner (2024), Introducing Grid WAR: Rethinking WAR for starting pitchers, Journal of Quantitative Analysis in Sports, Vol. 20 No. 4, pp. 293-329.

Brill and Wyner (2024) have come up with a novel method of computing WAR figures for pitchers, which they call "Grid WAR" (gWAR).

Their measure is intended to correct two issues with bWAR (the Baseball Reference version of WAR) and fWAR (the FanGraphs version).

The first is based on their belief that WAR should be context-dependent. Their specific argument is that basing pitcher evaluation on an estimate of runs allowed computed from the likelihood of specific events ignores the fact that these events do not occur in random combinations. Pitchers can, for example, give up a lot of baserunners but not allow a lot of runs (or the opposite), a fact that is ignored in FIP and other Three-True-Outcome-founded metrics.

Put another way: in the authors' estimation, WAR should be based on runs actually given up, not estimated runs based on what would have happened with random sequencing[2] or average batted-ball results. However, the authors seem to have missed that bWAR is already based on runs rather than FIP, so their objection should truly apply only to fWAR.

In any case, the use of runs rather than FIP presents a tradeoff: it makes gWAR a better descriptive metric than fWAR, but a worse predictive one. I personally believe that WAR should primarily be a predictive metric, so I disagree with their preference for gWAR.

I am more comfortable with the authors' second issue. Pitchers who are more variable in their appearances will be unfairly penalized by averaging over their overall performance, as an occasional blow-up can cancel out several good performances. For example, a pitcher who gives up 1 run in six innings three times and 8 runs in two innings once will have given up a total of 11 runs in 20 innings, a mediocre total when in truth the pitcher did an excellent job three-fourths of the time. (This of course would be a particular problem with relievers, and perhaps is part of the reason WAR tends to value them less than you'd expect.)

The authors also make the related point that extra runs in blowouts have less of an impact on win probability than the first few given up, and so should be weighed less in WAR metrics, although their demonstration of this effect in a diagram indicates a smaller impact than I would have guessed.

Anyway, the authors argue that games should not be averaged over.

As I understand it (the complex mathematics is mostly over my head), Brill and Wyner's method begins with a computation of overall win probabilities for each combination of runs allowed per full inning pitched (assuming away extra innings), adjusted for park, to which pitcher-specific actual performance in individual games is compared, using a version of fWAR's replacement level.

In simpler terms: the authors adjust gWAR for the distribution of runs allowed over their outings, not just their average runs allowed like ERA.

They apply Retrosheet data from 2010 to 2019 to compare gWAR to the other two versions. It turns out that gWAR was indeed higher for pitchers with more variable performances compared to pitchers who allowed runs more evenly. The authors also demonstrated that one-season gWAR predicted next-season gWAR better than fWAR, which in and of itself means nothing; I would like to know which better predicts next season fWAR.

[1] https://sabr.org/journal/article/plummeting-batting-averages-are-due-to-far-more-than-infield-shifting-part-two-strikeouts/

[2] The "random sequencing" version would be one using opposition batting lines – however, I am not aware of any versions of WAR that do this.

The study also demonstrated that individual pitchers are somewhat predictably game-to-game inconsistent from year to year, which is interesting. They also showed that using one's closer as an opener would result in more victories, but my guess is that this is because normal closer usage these days is dependent on the current definition of saves and so includes three-run leads which any competent major league pitcher would successfully maintain 95 percent of the time.

Lee, Dohyun, Jeongeon Lee, Tonghoon Suk and Min Kyu Sim (2024), A measure of the importance of moment for ball-strike counts in a baseball plate appearance, Journal of Sports Sciences, Vol. 42 No. 11, pp. 959-970.

Based on all pitches thrown between 2015 and 2018, the authors compiled a Markov chain representation of the outcomes for pitches at all 12 counts, with transitions between counts plus the odds of both the pitch and the entire plate appearance ending with the batter either getting on base or making out. Of course, batters did better in counts with more balls and worse in counts with more strikes.

This sort of thing has been done before, of course[3]. The authors went beyond that by computing an index of how critical each pitch was in the final outcome of the plate appearance, sort of a leverage index for each count. Again, more balls and in particular more strikes increased this index. Then, they showed that average fastball velocity increased as this criticality measure increased, was at its highest with two strikes, and that the increase for higher velocities was more pronounced for pitchers who had received Cy Young Award votes than for all pitchers. (I wish the comparison had been only with pitchers who did not receive votes.)

Erickson, Brandon J., Peter N. Chalmers, John D'Angelo, Kevin Ma, Stephen Fealy, Frank J. Alexander, and Christopher S. Ahmad (2024), Predraft elbow magnetic resonance imaging in Major League Baseball pitchers, Journal of Shoulder and Elbow Surgery, Vol. 33 No. 11, pp. 2448-2456.

Another piece of evidence concerning the damage that youth baseball is presently causing: an examination of pre-amateur draft MRIs for 245 eligible pitchers revealed that 70 percent of them had some UCL problem, including 3 percent with full and 24 percent partial tears.

Incidentally, I recommend a 2017 book I just read, *The Arm*, by Jeff Passan. It's basically a history of pitcher UCL tears, with a lot to say about the exploitation of wannabe pitchers as young as nine years old.

Charlie Pavitt, chazzq@udel.edu ♦

[3] See, for example, Tom Tango's printing of Josh Maciel's matrix at http://www.insidethebook.com/ee/index.php/site/comments/graphical_woba_by_count .

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 110phil@gmail.com .

"By the Numbers" notifications

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 to download from the SABR website.

Pitching Pixie Dust – Success of the Rays' Imported Arms

Tom Hanrahan

The Tampa Bay Rays have been quite successful in the past decade or two, despite spending much less than other winning teams. Here, the author investigates if there is evidence that the Rays' coaching staff was able to significantly improve their pitchers' performance, as many observers – and even the Rays' own pitchers – have suggested.

Over the 16 seasons to 2023, the Tampa Bay Rays have had great success, despite a relative pittance for a budget. They won 55 percent of their games in this stretch, the equivalent of 89 wins per full season. The Rays made the playoffs in nine of those years, and won their division four times, despite competing in the perennially competitive American League East.

Yet, the Rays' player payroll was constantly near the bottom of MLB. Among the 30 teams, they never ranked higher than 21st in spending, and their average rank over those 16 seasons was 27th. Along with the Oakland Athletics and Pittsburgh Pirates, they were solidly one of 3 teams at the very bottom of MLB spending in this period[1].

In addition to outperforming their budget, the Rays also outperformed expectations – "overachieving" by finishing ahead of the preseason pundits' predictions. One source has them averaging 4.5 wins above what the betting markets expected in this period, or 72 total games above the number of wins predicted in those 16 seasons[2].

Given all of the above, one question to ask is -- How? There are many routes to success. A brief list of possibilities might look like:

- Great drafting
- Great player development and coaching
- Signing of great free agents and great international players
- Making great trades

It is not simple to identify which of those causes is responsible for success; if a player is drafted and becomes a star, was it the wisdom of the draft pick, or the player development? If a trade works out, is it because a team snookered its trading partner, or was it the coaching the player received after the trade?

In this article, I examine one possible piece of the puzzle, focusing on the Rays' pitching. My reasons for this are twofold: first, in their period of success, run prevention has been the Rays' primary strength, as opposed to run scoring. In 10 of the 16 seasons, the Rays were in the top three in the AL for fewest runs allowed. Yes, the Rays' home park has been somewhat pitching-friendly, but this is still a very impressive achievement.

Second, there have been many articles published about the Rays' success in working with pitchers. The internet is stuffed full with these stories. And so, this paper's contribution might simply be to investigate empirically if the common narrative is true.

The Question and the Method

How do pitchers fare when they get traded *to* Tampa Bay, and how do they fare when getting traded *from* Tampa Bay to someone else? Is there a pattern of improvement when the same players pitch for the Rays, compared with pitching elsewhere?

[1] source: https://www.stevetheump.com/Payrolls.htm, opening day player payroll

[2] source: https://www.sportsoddshistory.com/mlb-odds/

To check, I found all pitchers with at least 55 IP for the Rays between 2008 and 2023 who also pitched at least 55 innings for other teams in the three years immediately preceding or following their stint in Tampa. (If the pitcher was traded to or from Tampa mid-season, I used that season too, so it could be more than three years.)

Then, I compared their stats from their time in Tampa to their time with other teams.

The Results

There were 55 pitchers who qualified under the criteria above. Of those, 28 pitched for other teams preceding Tampa, and 40 pitched for other teams after Tampa. (There were 13 pitchers counted in both groups.[3])

It turns out that both groups of pitchers did better in Tampa than elsewhere:

	IP Tampa	IP elsewhere	Individual Pitchers	Average ERA difference[4]
Before	4323	4841	23 of 28 better in Tampa (23-5)	-0.76 better in Tampa
After	9903	9033	27 of 40 better in Tampa (27-13)	-0.25 better in Tampa

Raw ERA is useful, but we need to account for park effects. As per baseball-reference.com, the average (pitching) park effect for Tampa in the years 2008-2023 was .955. The composite ERA for all Rays pitchers in this study was 3.64. Accounting for park effects, their ERA would rise by .17 to be 3.81 on a neutral field (3.64/.955=3.81). Rather than attempting to make park adjustments for every individual season, I will use a blanket 0.17 earned runs per game to adjust every pitcher's statistics throughout the article here. After the adjustments, the weighted park-adjusted ERA changes were:

	Average ERA difference, park adjusted
Before	-0.59 better in Tampa
After	-0.08 better in Tampa

The first number looks impressive; allowing more than half a run per game less can lead to many wins! The second is so close to zero, it could easily be random[5]. So the rest of the study will focus on the "before" group: the pitchers who came *to* the Rays from other major league teams; via trade, free agency, or whatever other means got them to Tampa.

Table 1 lists the pitchers, alphabetically, in the "before" group. As we saw, only 5 of the 28 pitchers had a higher ERA after coming to Tampa. Those five are marked in Table 1 in red. Additionally, if the allowance is made for park effects as calculated above, there would be two other pitchers who had adjusted ERAs higher in Tampa. Those two (Kluber and Morton) are marked in blue.

This still leaves three-fourths (21/28) of the pitchers showing improvement with the Rays, as compared to elsewhere. Whether one uses a simple test for significance by assessing how likely it is for a fair coin to come up heads 21 out of 28 tries, or a more robust statistical test for the mean being non-zero assuming a normal distribution, it is quite clear that this is not random noise: as a group, these men pitched better in Tampa.

[3] Shawn Armstrong, specifically, was a difficult case to decide. In mid-2021, he came to the Rays and pitched a little. Then he became a free agent and signed with Miami, pitching a few innings in 2022, before again free-agenting over to Tampa. I chose to not include either of these brief back-and-forth stints, using simply his previous years (2019-21) as his time prior to Tampa, and his new start with Tampa in beginning May, 2022.

[4] This is a weighted average, calculated as the harmonic mean of innings in the paired set. Erasmo Ramirez had 322 IP with the Rays from 2015-17. He had thrown 206 IP the previous three seasons with the Mariners, 2012-2014. That makes 2 / (1/322 + 1/207) = 252 weighted innings. Then I summed the cross product of the weighted IP and ERA differences for each pitcher-pair, and divided by the sum of weighted IP, to calculate weighted ERA difference.

[5] The simple (unweighted by innings) standard deviation of the mean for the "after" group was .18 runs, assuming normal distribution, with the sample size of 40 pitchers. The mean difference of .07 runs is far less than even one SD.

Aside from park effects, someone might wonder if perhaps the Rays' team defense was consistently excellent in this period, thus making pitchers better by turning more batted balls into outs than expected. Classic examples of this have certainly occurred; the early-70s Orioles' and mid-80s Cardinals' success was fueled in no small part by their gloves. However, this does not appear to be a material cause in this case. First, the Rays have not had an inordinate number of Gold Glove Awards in the past 16 years. Second, a look at one statistic tells us much: these pitchers walked significantly fewer batters while pitching for Tampa, and bases on balls are unaffected by defense. In the "before" group, 21 of the 28 pitchers reduced their walks allowed while with Tampa. The weighted average difference was 0.77 fewer BB per nine innings pitched.

Table 1 – Performance of pitchers joining the Tampa Bay Rays: before vs. after

With Tampa after				Weighted	ERA	With other teams before		
Years	IP	ERA		IP	diff	Years	IP	ERA
22-23	107	2.52	Shawn Armstrong	100	-3.19	19-21	93	5.71
12-12	62	3.04	Burke Badenhop	100	-1.30	08-11	251	4.34
14-14	62	4.91	Grant Balfour	95	2.38	11-13	199	2.53
09-10	139	3.56	Lance Cormier	161	-1.53	06-09	191	5.09
19-20	67	3.63	Drake Oliver	86	-1.18	16-18	122	4.81
11-13	113	3.57	Kyle Farnsworth	133	-0.48	08-10	162	4.05
16-17	70	3.59	Danny Farquhar	100	-0.26	13-15	178	3.85
08-10	591	3.87	Matt Garza	217	-0.60	06-07	133	4.47
18-23	387	3.21	Tyler Glasnow	207	-2.58	16-18	141	5.79
13-13	151	4.89	Roberto Hernandez	221	0.32	10-12	425	4.57
21-21	95	3.88	Rich Hill	135	0.63	18-20	230	3.25
08-12	236	3.09	J P Howell	110	-3.10	05-05	73	6.19
22-22	164	4.34	Corey Kluber	136	-0.06	19-21	117	4.40
23-23	87	3.93	Zach Littell	99	-0.37	20-23	115	4.30
21-21	64	1.55	Colin McHugh	89	-1.82	18-19	147	3.37
19-20	232	3.33	Charlie Morton	273	-0.07	16-18	331	3.40
19-19	70	2.31	Emilio Pagan	86	-1.54	17-18	112	3.85
11-14	268	3.59	Joel Peralta	172	-0.90	08-10	126	4.49
15-17	322	3.99	Erasmo Ramirez	252	-0.63	12-14	207	4.62
12-13	140	1.92	Fernando Rodney	156	-2.43	09-11	176	4.35
17-19	101	3.83	Chaz Roe	86	-0.37	14-17	75	4.20
17-18	97	3.33	Sergio Romo	124	-0.30	14-17	171	3.63
14-16	289	3.96	Drew Smyly	307	0.70	12-14	328	3.26
22-22	195	2.53	Jeffrey Springs	118	-2.89	18-20	85	5.42
21-21	124	5.07	Michael Wacha	165	0.59	18-20	245	4.48
16-17	71	3.79	Chase Whitley	81	-1.23	14-15	95	5.02
21-22	73	2.22	Matt Wisler	95	-2.22	18-21	136	4.44
13-13	70	3.09	Jamey Wright	103	-0.57	10-12	194	3.66
			Overall	4007	-0.76			

So what are we left with? Perhaps the frequently heard narrative is indeed correct – perhaps the Rays' coaching has indeed helped their pitchers to a large extent. I illustrate the narrative with just a few quotes:

- "I want to sign with the Rays, just to see what they can tell me." – Ryan Thompson
- "If Tampa sees something in you, there's probably something there" – Zach Littell[6]
- "There's a lot of different ways to say, 'Trust your stuff'... it's just a matter of how you message things for people to understand it." – Kyle Snyder[7]
- "They actually have concrete data to make you good at pitching… They're telling you: 'If you do this, you will succeed.' And chances are, that usually happens." – Tyler Glasnow
- "What makes Kyle and Stan very special in that they're able to just condense everything … they value as the most important and really harp on that." – Kevin Cash[8]

Aside from the statements above, here is an important piece of information: During the years under discussion, the Rays' coaching staff has had amazingly little turnover:

Manager:	**Joe Madden**	**(2006-14)**
	Kevin Cash	**(2015-24)**
Pitching Coach:	**Jim Hickey**	**(2007-17)**
	Kyle Snyder	**(2018-24)**
Bullpen Coach:	**Stan Borowski**	**(2010-22)**

That, fellow reader, is a great example of a team's management working together for a long time!

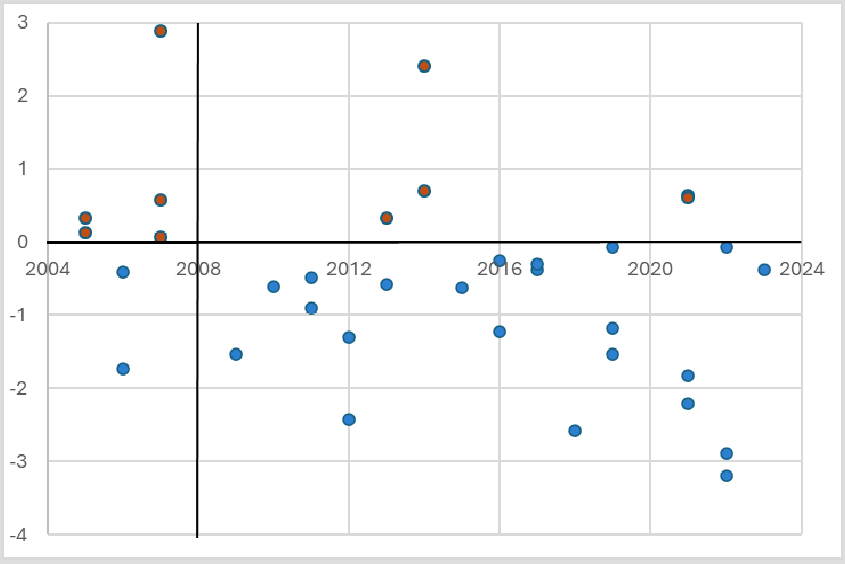
Caveat and Conclusion

Before wrapping up, I should note that by choosing the data set beginning with the season where the Rays first were successful (2008), I could be guilty of (intentionally or not) "cherry picking" the data. And it's true that, by adding pitchers into the data set from previous years, the picture gets murkier… the Rays pitchers in 2005-07 who came from elsewhere did not show the same improvement.

Figure 1 depicts the information from the previous table, adding in other pitchers from those three preceding years; the x-axis shows the first year each pitcher was with Tampa.

Is it fair to quote the conclusion that the Rays' pitchers improved by 0.59 earned runs per game (park adjusted) after coming from other teams? That's a judgment call, based on whether it is fair to reason that they "figured something out" in or about 2008.

Figure 1 – ERA Difference for pitchers with Tampa, compared to previous three seasons



[6] https://www.si.com/mlb/2023/10/03/rays-pitcher-fixing-machine-kyle-snyder
[7] https://www.si.com/mlb/2023/08/08/rays-pitching-development-kyle-snyder
[8] https://www.mlb.com/news/featured/rays-key-to-pitching-success

The extent to which you will find it persuasive enough to be predictive -- to assume going forward that such an improvement will be forthcoming – depends on your answer to whether 2008 was cherry-picked.

Regardless, this article has shown that empirically: YES much of the Rays' success – winning often, on a very tight budget, often overachieving pundits' projections – can be attributed to the improvement pitchers experience in Tampa. The theory that the difference was the Rays' coaching is certainly credible, and well-supported by the data – even if not necessarily proven.

Tom Hanrahan, han60man@aol.com ♦