? (2009). Looking for park effects that make sense.
https://statspeakmvn.wordpress.com/2009/01/page/2/

This was probably posted by Brian Cartwright, although it could have been Pizza Cutter (in actuality Russell Carleton). Using Retrosheet 1993-1999 and 2003-2008 data, the author compared player road HR per fly ball and K per PA to see when each effect stabilizes. The argument is that this would represent performance consistency in the average ballpark. Here is split half reliabilities for HR/FB for
500 FB: .703
1000 FB: .711
2000 FB: .864
3000 FB: .876
4000 FB: .864
The last dip probably due to smaller a sample size. Anyway, to put this into context, the average ballpark gets about 1800 fly balls a year.
Here is K/PA
1000 PA: .608
2000 PA: .785
4000 PA: .757
6000 PA: .813
8000 PA: .845
10000 PA: .847
12000 PA: .869
14000 PA: .905
16000 PA: .898
18000 PA: .935
The average ballpark gets about 6000 PAs a year.

Acharya, Robit A., Alexander J. Ahmed, Alexander N. D'Amour, Haibo Lu, Carl N. Morris, Bradley D. Oglevee, Andrew W. Peterson, and Robert N. Swift (2008). Improving major league baseball park factor estimates. *Journal of Quantitative Analysis in Sports,* Vol. 4 Issue 2, Article 4.

There are at least two obvious problems with the original Pete Palmer method for determining ballpark factor: assumption of a balanced schedule and the sample size issue (one year is too short for a stable estimate, many years usually means new ballparks and changes in the standing of any specific ballpark relative to the others). A group of researchers including Carl Morris (Acharya et al., 2008) discerned another problem with that formula; inflationary bias. I use their example to illustrate: Assume a two-team league with Team A's ballpark "really" has a factor of 2 and Team B's park a "real" factor of .5. That means four times as many runs should be scored in the first as in the second. Now we assume that this hold true, and that in two-game series at each park each team scores a total of eight runs at A's home and two runs a B's. If you plug these numbers into the basic formula, you get

1 – (8 + 8) / 2 = 8 for A; (2 + 2) / 2 = 2 for B
2 – (2 + 2) / 2 = 2 for A; (8 + 8) / 2 = 8 for B
3 – 8 / 2 = 4 for A; 2 / 8 = .25 for B

figures that are twice what they should be. The authors proposed that a simultaneous solving of a series of equations controlling for team offense and defense, with the result representing the number of runs above or below league average the home park would give up during a given season. Using data from Retrosheet data from 2000 to 2006 for each league separately (despite interleague play, mudding the waters) and, based on 2006, a 5000-game simulation, the authors find their method to be somewhat more accurate and, in particular, less biased than the basic formula. They note how their method also allows for the comparison of what specific players would accomplish in a neutral ballpark and how a given player's performance would change if moving from one home ballpark to another.

Adler, Joseph. (2006). *Baseball hacks.* O'Reilly Media: Sebastopol, CA.

This is a book explaining how to download and analyze baseball data from public sources, including MySQL and R code exemplars. Retrosheet is one of the public sources featured prominently. To name some examples: Chapter 2 describes the organization of the event files and how to use them to make box scores and data bases; and also how to work with game logs. Chapter 3 includes a summary of how to turn MLB.com Gameday play-by-play description into event file format. In Chapter 5, Retrosheet data was used to demonstrate an index (Save Value) intended to describe the extent to which closers' saves typically occurred in high versus low leverage situation.

Albert, Jim (2001). Using play-by-play baseball data to develop a better measure of batting performance.

Jim Albert's work, unfortunately no longer available, is an analysis of Gary Skoog's (RIP) *Value Added* approach to measuring offensive production describe in Bill James's 1987 *Baseball* Abstract. He used Pete Palmer's run expectancy table as its basis, but the method would work just as well with an alternative. Basically, one takes the run potential at the end of a plate appearance, subtracts from it the run potential at the end of the plate appearance, and adds any runs that scored during the PA. If the result is positive, the player has contributed to run scoring, and if it is negative, the player has damaged run scoring. Each inning, the lead-off hitter is changed with .454 for before the PA, which is the mean run potential for no baserunners/no outs. The batter making the third out in an inning ends with a 0, meaning that they cannot have a positive contribution unless a run scored during the inning-ending event. It is important to remember that one cannot simply use the run potential at the beginning of a plate appearance when making these calculations, because various events can occur during

a PA that change the base-out situation (SB, CS, WP, PB, Balk). Instead, one must use the run potential just before the "event" (e.g., walk, hit, out) that ends the PA. Stolen bases and caught stealing are credited to the baserunner. Getting on base on an error is not credited to the batter. The batter does get credit for baserunners getting extra bases on hits (e.g., first to third on a single), which Skoog was not comfortable with and invited discussion by interested analysts. Jim Albert (2001) recreated the Skoog method using 1987 National League data gathered by Project Scoresheet and available at Retrosheet, used it to estimate team run scoring per game, and then compared those estimates to actual team runs per game using the root mean square error (RMSE) as a goodness of fit measure. Its RMSE was .067, compared to .121 for Batting Runs, .202 for Bill James's Runs Created (described later), .212 for OPS, and .242 for OBA.

Albert, Jim (2013). Length of baseball games.
https://baseballwithr.wordpress.com/2013/11/22/length-of-baseball-games/

Based on 2012 Retrosheet data, a regression equation showed that each additional pitch added 36 seconds to the game (pre pitch clock). Based on a diagram, the equation appears to account for a lot of variance, at least 70 percent.

Albert, Jim (2014). Chance of a hit during different pitch counts.
https://baseballwithr.wordpress.com/2014/09/17/chance-of-a-hit-during-different-pitch-counts/

Used 2013 Retrosheet data to compose diagrams of hit type proportions at each ball-count count.

Albert, Jim (2014). Exploring types of batted balls.
https://baseballwithr.wordpress.com/2014/12/02/exploring-types-of-batted-balls/

Jim showed how higher flyball and lower groundball rates equated to greater average run values for batters with 2013 Retrosheet data. Two examples from quotes: "A batter with a flyball percentage of 10% and a groundball percentage of 60% is predicted to have a run value close to 0. In contrast, a batter with a flyball percentage of 40% and a groundball percentage of 30% is predicted to have a run value around 0.10." [that would be per PA]

Albert, Jim (2014). Number of pitches in a plate appearance.
https://baseballwithr.wordpress.com/2014/11/03/number-of-pitches-in-a-plate-appearance/

In 2013, the number of pitches per plate appearance was a smooth curve, with (all approximations read off graph) about 20,000 with one, 30,000 with rwo, about 33,000 with three, 35,000 with four, 30,000 with five, 20,000 with six, maybe 9,000 with seven, maybe 4,000 with 8, maybe 2,000 with nine, 1,000 with ten, and relatively few with more

than ten. The run values per pitches were again approximately from diagram 0.04 for one, 0.03 for two, then way down to –0.02 for three through five, then up to 0 for six, 0.01 for seven, maybe 0.025 for eight, maybe 0.035 for nine, and 0.04 for ten. I assume walks were likely outcome for long PAs. If you exclude strikeouts they are actually quite low at one and two (and three, below 0.05, but go up to 0.13 or 0.14 for nine and ten. Balls in play also rise but slowly, from about 0.04 to 0.06; much less effect than K's. Clearly what matters here are strikeouts and walks. See Dave Smith, 2007, for a similar study with analogous findings.

Albert, Jim (2014). Calculation of in-game win probabilities.
https://baseballwithr.wordpress.com/2014/12/29/calculation-of-in-game-win-probabilities/

For 1993 and 2013 seasons, plots of odds of winning given inning and score differential between teams. Obviously, the odds of winning and losing both increased with leads/deficits of given number of runs as innings went on. Tne home team advantage for tie games decreased a bit as innings progress.

Albert, Jim (2015). Sequence of pitch counts.
https://baseballwithr.wordpress.com/2015/07/20/sequences-of-pitch-counts/

This webpost, based on 2014 Retrosheet data, includes a few interesting graphs. The first shows the probability of adding a strike to the count for each of the initial counts. We see the probability of adding a strike is relatively low for 0-1, but relatively high for a 3-0 count. The second shows the probability of adding a ball to the count. Here we see that it is more likely to add a ball on a 0-2 count, but less likely to add a ball to 2-1 and 2-2 counts. The third shows the probability of keeping a two-strike count (with a foul ball). As you might expect, the probability of keeping the count is highest for 3-2, followed by 2-2, 1-2, and 0-2.

Albert, Jim (2015). Historical look at pitcher and batter hands.
https://baseballwithr.wordpress.com/2015/06/04/historical-look-at-pitcher-and-batter-hands/

Comparison of six seasons ending in 4's, from 1964 through 2014, read off diagrams. The proportion of right-handed pitched innings increased; in the first inning aka starters, from about 66 percent in 1964 to maybe 73 percent in 2015; in the ninth inning aka closers, from around 73 percent in 1964 (with the proportion going up gradually throughout the game) to almost 80 percent in 2014 (with it actually going down stating in the 6^{th} inning, as low as 70 percent in the 7^{th} inning, and shooting up afterward). In contrast, right-handed hitter plate appearances in 1964 was around 58 percent (its lowest) in the first inning and 68 percent (its highest) in the second, the drifted down to about 61 percent in the ninth. In 2014, itb was only about 52 percent in first, 58 percent in the second, about 55 in the eighth, and 52 percent in the ninth.

As for the same-handed matches in which the pitcher has the advantage, for righties in 1964 these were maybe 49 percent in the first, 61 percent in the second, and 58 percent in the ninth. For lefties, the proportions were 25 percent in the first, 20 percent in the second, and up to 30 percent in the eighth when, even then, lefty specialists apparently were prevalent. In 2014, about 24 percent in the first and third, 21 percent in the second and fourth, 41 and 42 percent in the seventh and eighth, and 34 percent in the ninth.

Albert, Jim (2016). Fouls with two strikes.
https://baseballwithr.wordpress.com/2016/06/07/fouls-with-two-strikes/

Based on 2015 data (likely Retrosheet), Jim looked at individual percentages of fouls per pitch on two-strike counts. Read off diagrams, the highest were Victor Martinez and Juan Lagares at over 37 percent, the lowest was Andrelton Simmons at 13 percent. No clear relationship with hitting ability (Buster Posey was fourth lowest at about 17¼ percent).

Albert, Jim (2016). Making sense of a .300 midseason batting average.
https://baseballwithr.wordpress.com/2016/03/03/making-sense-of-a-300-midseason-batting-average/

Based on 1960-2015 Retrosheet data for all 2619 batters with at least 100 AB and BA at or above .300 July 1st. 53 percent of them ended at that level. Jim divided BA into two components representing strikeout rate and batting average on balls in play. The relationship between strikeout rate and finishing above .300 was quite strong, with the odds of .300 at about 42½ percent with a K rate of 30 percent going up approximately linearly to about 62½ percent with a K rate of 15 percent. In short, fewer strikeouts meant more likely hitting .300.

Albert, Jim (2016). Graph of pitch count transitions.
https://baseballwithr.wordpress.com/2016/05/25/graph-of-pitch-count-transitions/

A count transition matrix for 2015 using Retrosheet data.

Albert, Jim (2016). Graphing pitch count effects.
https://baseballwithr.wordpress.com/2016/05/09/graphing-pitch-count-effects/

Graphs of 2011 run values per count by Max Marchi, and the same for 2015 by Jim.

Albert, Jim (2016). Situational home run hitting.
https://baseballwithr.wordpress.com/2016/12/06/situational-home-run-hitting/

1980 to 2015 Retrosheet data. The run value of homers bounced around a lot, but went from about 1.41 on average in 1980 to 1.385 in 2015 (read off diagrams), This is largely

because the proportion of solo homers (also quite variable across seasons) went up from about 56½ percent to almost 59 percent. Grand slams were at about 2.2 percent in 1980, went up to about 2.6 percent in the late 1990's through 2010, but then dropped to about 2.0 in 2015.

Albert, Jim (2016). A historical look at bunts.
https://baseballwithr.wordpress.com/2016/03/25/a-historical-look-at-bunts/

Retrosheet data 1960 to 2015, percentage of PA's that were bunts, read off diagram. Percentage marked as sacrifices: About 1.2 percent 1960 to 1980, then sharp drop in just four years to 0.9 percent, up to a bit over 1 percent in early 1990's, by 2000 down to 0.85 percent through about 2010, then collapse to about .065 percent in 2015. Percentages marked as attempted hits: Around 0.2 percent 1960, down to about 0.17 in the 1970s surrounded by short sharp drops below 0.12 percent in mid 1960's and early 1980's, then vacillating cyclically between 0.175 and 0.2 percent in the 1990's and 2000's, with an upswing slightly over 0.2 in 2015 (did it continue afterward?). Percentage of homers hit at home was under 50 percent more than over it 1980 to around 2005, and then went mostly over 50 percent through 2015.
1988 to 2015 – the percentage of homers hit in pitcher's counts went up from about 27 to about 31; in hitter's counts down from about 25 to 22½ (the remainder in neutral counts). Not sure, may have considered 2-1 and 3-2 as neutral.

Albert, Jim (2016). SB insights from 2016 Retrosheet data.
https://baseballwithr.wordpress.com/2016/12/26/sb-insights-from-2016-retrosheet-data/

2016 Retrosheet data. SB success rate for second base about 70 percent with only a runner on first but about 90 percent with two base runners (the second either on second or third). Success rate for third base around 80 with one or two runners on. One out consistently had the worst and two outs the best success rate. A plot of attempt and success rates shows at least a 0.7 correlation between the two; more successful stealing teams try it more often.

Albert, Jim (2017). History of intentional walks.
https://baseballwithr.wordpress.com/2017/03/09/history-of-intentional-walks/
NOT IN BIBLIOGRAPHY, IN REFERENCES, NEEDS RETROSHEET ENTRY

All based on Retrosheet data from 1960 to 2016, estimated off diagrams. IBB rate was around 0.7 percent of PAs in the early 1960's but by 1970 were up to around 1 percent. About 1975 to 1995 they vacillated between .8 and .9 percent, down to .6 percent late 1990's but back to .7-.8 percent early 2000's. After that until 2016, down to about .5 percent.

Average per season per player who got them has gone up and down in tandem with the overall rate, as high 4.5 (late 1960's) and as low as 3 (2016). The leader has generally gotten about 25 (with the Barry Bonds exception during the Steroid Era).

Albert, Jim (2017). Game duration exploration with game log data.
https://baseballwithr.wordpress.com/207/01/04/game-duration-exploration-with-game-log-data/

Based on 2016 Retrosheet data, game length was predicted by

95.77 + (2.71 X [hits + walks + HBP]) + (35.32 X [left on base divided by {hits +walks + HBP}])

with a standard deviation of 13 minutes (meaning two-thirds of games are within +/– 13 minutes).

Albert, Jim (2017). Platoon effects in baseball.
https://baseballwithr.wordpress.com/2017/01/11/2016-platoon-advantage/

In 2016, according to Retrosheet data, the average wOBA platoon effect for left-handed batters was in the low 80's up to about 400 PAs, in which it started decreasing at an acccelerating pace, below 75 at 600 PAs and dropping toward 70 above that. For right-handed batters, it was around 35 through about 300 PAs, down to about 25 at 500 PAs, and leveled off above that. Switch-Hitters were at just about 100, which calls the very idea into question.

Albert, Jim (2017). Looking deeper at strikeouts.
https://baseballwithr.wordpress.com/2017/07/24/looking-deeper-at-strikeouts/

The following is based on Retrosheet data from 1988 to 2016: Walk rate was about (estimated off diagram) 6.9 percent in 1988, at 8 percent in 1990, then generally down to around 7.2 percent during the mid-2010's. Strikeout rate was about 14 percent in 1990, reached 16 percent during the 2000's, then up to 20 percent in 2015 and still increasing.

Albert, Jim (2017). Predicting second half team results.
https://baseballwithr.wordpress.com/2017/07/05/predicting-second-half-team-results/

The 1967-2016 correlation between first and half season differentials between wins and losses (wins minus losses) was negative, as the extreme teams on both ends tended to regress toward the mean in the second half. Jim's formula for predicting second half from first half was

(First half wins minus first half losses) – (0.44 X first half wins minus first half losses)

Albert, Jim (2019). Leverage of win probabilities.
https://baseballwithr.wordpress.com/2019/02/18/leverage-of-win-probabilities/

Using 2018 data, Jim constructed charts showing leverage figures for innings, base-out situations, and different score differentials.

Albert, Jim (2019). Historical look at pitcher usage.
https://baseballwithr.wordpress.com/2019/01/28/historical-look-at-pitcher-usage/

Starting 1960 and up to about 1980, average 2½ pitchers per game, but after that approximately linear increase to 4½ in 2018.

Albert, Jim (2019). 2018 Retrosheet data and length of a PA.
https://baseballwithr.wordpress.com/2019/01/01/2018-retrosheet-data-and-length-of-a-pa/

These are 2018 run values from Retrosheet by number of pitches, all approximate figures read off of a diagram:

Pitches	1	2	3	4	5	6	7	8	9
RV	.056	.042	–.021	–.035	–.016	–.002	0.01	.034	.044

The webpost also includes a chart of run values and number of pitches further subdivided into various counts; for example, the RV for 3-2 for the 6th, 7th, 8th, and 9th pitches.

Albert, Jim (2019). A comparison of two seasons: 1968 and 2018.
https://baseballwithr.wordpress.com/2019/01/14/a-comparison-of-two-seasons-1968-and-2018/

Jim Albert (2019) compared the year of the pitcher with 50 years afterward based on relevant Retrosheet data. In 2018, as compared with 1968:
Games were about a half hour longer
There were four more runs scored per game
Margins of victory were generally greater
There were more homers, hit by pitches, strikeouts, doubles, and walks
There were fewer singles, triples, outs om batted balls, fielder's choices, errors, and intentional walks
There were far more pitchers used per game
See the webpost for additional details on each comparison

Albert, Jim (2020a). The 100 pitch limit and pitcher efficiency.
https://baseballwithr.wordpress.com/2020/01/14/the-100-pitch-limit-and-pitcher-efficiency/

According to Jim Albert (2020, Retrosheet data 2000-2019), for pitchers who pitched into the fifth inning, in other words did not come out early due to injury or ineffectiveness, the mean pitch count when the starter came out was at about 102.5 in 2000, at or just above 100 from 2005 to 2014, and then dipped to 95 in 2019. The analogous figures for mean number of batters faced was 28, 27, and 25 ½; the proportion who made it into the seventh inning was about .54 in 2000, .46 to .48, and around .27.

Albert, Jim (2020). Why are baseball games so long?
https://baseballwithr.wordpress.com/2020/01/01/why-are-baseball-games-so-long/

Using most likely 2000-2019 Retrosheet data, Jim Albert (2020) determined that mean number of plate appearances per game did not increase, but mean number of pitches per PA went up from about 3.7 to 3.9 and mean time per pitch from about 31.5 to 33.9 seconds.

Albert, Jim (2020). MLB's new extra inning rule.
https://baseballwithr.wordpress.com/2020/07/13/mlbs-new-extra-inning-rule/
Greenberg, Morris (2020). Optimal extra inning rules.
https://baseballwithr.wordpress.com/2020/07/20/optimal-extra-inning-rules/

According to Jim Albert (2020), the probability of runner on second/no out was about .775 from 2000 to 2008, and afterward dropped approximately linearly to a little over .7 by the late 2010's. The probability of that runner scoring was about .63 in the same 2000-2008 stretch, then down to around .6 in the 2010's. A simulation using 2019 (real extra inning baseball) and 2020 (ghost runner) data showed the change lowering the odds of an extra inning game going into the eleventh, twelfth, and thirteen innings to be .525, .278, and .151 with real baseball and .284, .084, and .024 with the ghost runner. Morris Greenberg corrected two errors Jim made due to assuming randomness in pitcher and hitter appearances in extra innings; pitchers who allow the leadoff hitter to get on base are not randomly distributed but are worse than average; and batters are most likely the best hitters. He saw the rule as increasing the odds of a game ending in the tenth from .55 to .72. The actual odds are dependent on team offensive and defensive strength and overall run environment.

Albert, Jim (2020). Introduction – home advantage in sports.
https://baseballwithr.wordpress.com/2020/02/11/home-advantage-and-margin-of-victory/

The following are approximate home field advantages categorized by margins of victory, as read off of a diagram, and based on 2000 to 2019 Retrosheet data.

Margin	1	2	3	4	5	6	7	8	9	10
Percent	61	52	52	51	50½	52	48½	53	52	53

Albert, Jim (2020). Is pitching 75% of the game?
https://baseballwithr.wordpress.com/2020/09/28/is-pitching-75-of-baseball/

Using 2019 Retrosheet data, Jim Albert (2020) estimated the relative amount of variability in batter and pitcher figures, in so doing providing evidence of which was more responsible for the result. Batters contributes a little less (0.9) to launch angle, but about 1.15 as much for walks, 1.25 as much for whiff rates, 1.4 as much for hits on balls in play, 1.5 as much for hit by pitches, 1.6 as much for strikeouts, 2 times as much for exit velocity, and 2.5 as much for homers.

Albert, Jim (2021). Visualizations of pitch values.
https://baseballwithr.wordpress.com/2021/01/18/visualizations-of-pitch-values/

Here is Jim Albert's determination of run values for plate appearances that either passed or ended at various counts, from 2019 Retrosheet data. The values for were computed by subtracting the run value of the previous count from the run value for the present count; see the webpost for details.

Albert Jim, (2021). Count transitions: A historical look.
https://baseballwithr.wordpress.com/2021/01/04/count-transitions-a-historical-look/

Jim constructed a transition matrix between counts using Retrosheet data from 2001 to 2020. During those years, the probability of a strike on the next pitch increased on every count by between three and six percent, with (except for 3-2) the probability of the plate appearance ending with a batted ball decreasing to compensate. This is another way of looking at the strikeout increase.

Albert, Jim (2023). Situational runner advancement.
https://baseballwithr.wordpress.com/2023/02/22/situational-runner-advancement/

2000 to 2019 data from from Retrosheet data, with figures here read of charts. Runner advancement from first on outfield singles year-to-year ranged from around 29 to 32 percent at home, 27 to 30 percent away. With run margins of one or no runs, between 30 and 35 percent; more than one run, between 24 and 28 percent. Number of outs; two, between 32 and 36 percent, between 26 and 30 percent; none, 24 and 28 percent.

Location; left field, 12 to 15 percent; center, 25 to 35 percent; right field, 40 to 45 percent.

Albert, Jim (2024). Historical look at batting order production. https://baseballwithr.wordpress.com/2024/10/29/historical-look-at-batting-order-production/

All based on Retrosheet data: 1980 through early 1990's, OBP highest for batting order position #3, followed by #4. #5 and #6 as high as #1 and #2, then #7, #8, and #9. By the early 2020's, the top four batting positions were pretty close to one another; the next four were also pretty close to one another. Number nine was way behind, but since the universal DH it was just behind.
SLG – early 1990's, #3 and #4 10 or so points higher than #1 and #2, but the four had become quite close by the early 2020's. Also early on, #5 and #6 were also higher than #1 and #2, but now are close to the top 4, #7 and #8 lagging throughout (#9 as before).
Run Value – early 1990's, #3 and #4 highest followed by #5, then #1, #2,and #6 pretty close to one another, then in order #7, #8, #9. In early 2020's, the top four had coalesced, followed by #5 and #6, then #7, #8, and #9 pretty close to one another.
See the webpost for details.

Albert, Jim (2025). Who controls the count? https://baseballwithr.wordpress.com/2025/01/06/csv-downloads-of-retrosheet-data-who-controls-the-count/

See this Jim Albert post using 2024 Retrosheet data for a chart of the relative variance pitchers have relative to batters in determining the outcomes of plate appearances ending at each count. Note that overall more pitches shifts majority from pitcher to batter; pitchers were responsible for about 60 percent of variance for pitches ending either 1-0 or 1-1 versus about 40 percent for 3-2.

Alcorn, Michael A. (2018). (batter|pitcher)2vec: Statistic-free talent modeling with neural player embeddings. *MIT Sloan Sports Analytics Conference.*

Inspired by Bill James's concept of Similarity Score, Alcorn (2018) presented a sophisticated method for judging similarity among pitchers and among position players, using Retrosheet data on the outcome of all 2013-2016 Plate appearances.

Andrecheck, Sky (2009). A study in home field advantage – Will the new stadiums be friendly to NY teams? http://baseballanalysts.com/archives/2009/04/a_study_in_home.php
Andrecheck, Sky (2009). Does a quirky home field cause a road disadvantage? http://baseballanalysts.com/archives/2009/04/home_field_caus.php

Sky Andrecheck used a regression equation to see which variables affected idiosyncratic ballpark home field advantage for 1901 to 2008. His original study (2009) turned out to be biased by Coors Field's extremities. A re-analysis without Coors (2009) noted the following variables to encourage a large home field advantage: a “quirky” ballpark, which he described in 2009 as having “odd dimensions, wild wind, strange ground rules, etc.,” a ballpark favoring a lot of doubles, a dome, and a pitcher's park. The first 2009 regression accounted for 39 percent of variance in home field advantage. Sky did not say how much the second one did; without Coors, probably less.

Andrecheck, Sky (2009). How to best measure a team's true talent.
http://baseballanalysts.com/archives/2009/08/how_best_to_mea.php

Based on Retrosheet data 1960 through either 2008 or 2009 (he didn't say which), Sky Andrecheck (2009) first took a random 40-game sample of team games, and then compared which of the following three predictors was the better predictor of seasonal team W-L: W-L for those 40 games, the Pythagorean formula with an exponent of 1.81, or a “component Pythagoras” based on Runs Created? Yhere was not much difference, with Component Pythagoras slightly better and W-L slightly worse, and a combination of all three a bit better yet but not worth the extra effort.

Andrecheck, Sky (2009). Do pitchers and hitters work the count efficiently?
http://baseballanalysts.com/archives/2009/07/do_pitchers_and.php

From Retrosheet 2007 data, here are the outcomes of pitches at each count.

Of course, more balls(strikes) with counts favoring pitchers(batters).

“I ran a simulation to simulate whether each at-bat turned out to be a walk, strikeout, or a ball in play, assuming that each pitch had the same ball/strike/foul/inplay probability regardless of count. I then compared this to the actual outcomes.

The following chart shows the difference between the BB/K/In-play rates in the simulation (where it is assumed both the batter and pitcher are blind to the count) and the real outcomes.”

“Major differences occur mainly in hitters counts, such as 2-1, 3-1, and 3-2, where the true propensity to put the ball in play is higher than the simulated results, and the walk rate is much less than the simulated results. This also is no surprise, seeing that the pitcher is more likely to throw a hittable pitch when he is down in the count, and hence the batter is more likely to put it into play and less likely to walk.”

“the run value outcome in various counts and comparing the simulation to the real data. Below is a chart doing just that.”

“The chart above gives the BAV, OBP, and SLG percentages for both the simulation and the real data. It also gives the wOBA for each. The last column uses Pete Palmer's Linear Weights and shows the difference between the simulation and the real data over the course of 600 PA's at each count. This is perhaps the most useful column. Those with a positive value indicate that the batter is creating more runs than would be expected in that count via the simulation, while a negative value indicates that the batter is under-performing relative to the count.

In many counts, the simulation and the real data both produce about the same amount of runs. For instance, on the 2-0 pitch, the hitter's batting average is higher than we would expect if the players were blind to the count (.292 vs. .279 in the simulation) and the batter hits the ball much harder (.497 SLG vs. .440 SLG in the simulation), but it comes at the expense of fewer walks, with the OBP falling from .519 to .487. The overall difference in production is less than a run over 600 PA's with a 2-0 count. As we would expect, the net result is that the increased power is offset by fewer walks, and neither the pitcher nor the hitter gain an upper hand. This indicates both players are likely using an efficient strategy and not allowing the opposition to use the knowledge of the count to their advantage.

However, this offsetting does not occur at every count. The 3-0 count is obviously a hitter's count - the simulation predicts a .290/.714/.457 line from hitters with a 3-0 count. However, the true data shows hitters taking an even *greater* advantage of the count. The real line is .295/.720/.516 (intentional walks are removed from the data), indicating that hitters hit the ball with much greater power without sacrificing walks. The result is an advantage for the hitter above and beyond what we would expect a 3-0 count to entail if players were not strategizing about the count. This indicates that the pitcher is not able to effectively counter the batter's 3-0 strategy.

Is this true for other highly favorable hitter's counts? Looking at the 3-1 count, we see that in fact the opposite is true! In this case, the pitcher takes the strategic edge. The expected hitting line is .252/.656/.397, while the true line is .292/.590/.500. Here we again see a big increase in power, with both the slugging average and batting average increased, but it comes at a cost of fewer walks. The result is a loss of about 6 runs due to the strategizing about the count. Pitchers are more likely to throw a strike to avoid a walk, but unlike on the 3-0 count, the batters are unable to generate enough power to offset the loss of walks. The result is that the pitchers are gaining an edge.

The 0-2 count is another extreme count, in which either the pitcher or batter may take a strategic edge. In this case, it's a pitcher's count and the expected line is .192/.207/.302. This is better however, than the real line of .180/.201/.265. In this case, the batter is

sacrificing power and average, without increasing his on-base percentage. The result again is a loss of 6 runs by the batter over the course of 600 PA's. The pitcher is throwing more balls dancing around the plate, and the result is less power for the hitter, but without a corresponding increase in walks. Thus the pitcher is gaining an upper-hand on the hitter.

The result of the other counts don't show either the pitcher or hitter gaining a huge edge. 3-2 and 2-2 show the batter gaining a slight strategic edge, and 2-1 and 0-1 show the pitcher taking a slight strategic edge, but the effect is small. These results don't account for the fact that potentially stronger batters are more likely to reach 3-0 counts, and weaker batters are more likely to reach 0-2 counts, which may somewhat explain the advantages seen in the real data vs. the simulation. However, this does not explain the fact that 3-1 counts seem to show the pitcher gaining an edge.

Andrecheck, Sky (2009). Staying alive: Who has the advantage after fouling off multiple 3-2 pitches? http://baseballanalysts.com/archives/2009/08/staying_alive_w.php

Based on 2007 Retrosheet data, this is outcome after fouling off 3-2 pitches after "removing intentional walks, at bats with bunts attempts, pitch outs, etc."

"The chart below shows virtually identical overall batting averages for the hitters in each of the situations."

0 Foul Balls: .269
1 Foul Balls: .270
2 Foul Balls: .271
3 Foul Balls: .270
4+ Foul Balls: .271

Which suggests that there are no differences in batter ability affecting number of foul balls and resulting production.

Andrecheck, Sky (2009). Do catchers wear down in September? http://baseballanalysts.com/archives/2009/09/do_catchers_wea.php

No, according to a study of all (97) catchers "in the Retrosheet era" (early 1950's to late 2000's) who caught at least 800 games. To be sure, their OPS in September averaged 11.4 points lower than their overall (should be rest of the season) OPS; but this was no different than the drop for all batters. Incidentally, the difference between September and full season performance decreased from about 20 points in the early 1950's to less than 5 points in the late 2000's.

Apostoleris, Lucas (2019). Pitchouts are going extinct.
https://www.baseballprospectus.com/news/article/48740/prospectus-feature-pitchouts-are-going-extinct

In 2003-2005, pitchouts averaged 0.14 per team per game, went up to 0.16 2007, then down to 0.02 in 2018. Managerial success rate at calling pitchouts on steal attempts did not change, 0.15 to 0.20 from 2003 to 2017 with a probably fluky 0.32 in 2018. Baserunners who were relatively aggressive (averaged steal attempts at least 20 percent of the time on first and first-and-third situations) saw pitchouts in such situations 20 percent of the time 2002-2003, up to 40 percent by 2011, back down to 20 percent by 2015, back up again to over 40 percent in 2018. Lucas uncovered the same trends with an increased sample (steal attempts at least 10 percent of the time). Those to runners with no attempts actually averaged 2-3 percent 2003-2017, and finally down to close to 0 in 2018.

Apostoleris, Lucas (2020). Should pitchers still bunt?
https://www.baseballprospectus.com/news/article/58384/high-and-tight-should-pitchers-still-bunt/

Non-pitcher sacrifice bunts, above 0.7 per game in 1920, went down below 0.1 per game by 2019. However, pitcher sac bunts went up from 1920 (about 1.2 per game) to over 2 per game from the 1970s through the mid 2010s, then back down a bit below that afterward. The mean OPS for non-pitchers was in the .700s most years 1920-2019, with a few below; for pitchers, it was around .500 in the 1920s, went down about linearly to just over .300 by the late 2010s. The following detailed sacrifice bunting figures are based on 2003-2019 Retrosheet data:

	ATTEMPTS	PERCENT	IN PLAY	FOUL	MISSED
Position Players	70,596	0.6%	48.7%	43.4%	7.9%
Pitchers	29,077	8.6%	51.2%	37.3%	11.5%

	IN PLAY	SACS	HITS	OUTS
Position Players	34,364	40.4%	25.7%	33.8%
Pitchers	14,874	64.5%	1.8%	33.7%

Note the basically identical out rate. Here are seven situations in which pitchers normally bunt:

BASE STATE	OUTS	PA	BUNT RATE
1__	0	5,439	81.1%
1__	1	6,571	73.9%
_2_	0	1,611	72.9%
12_	0	1,374	81.2%

12_	1	2,396	71.4%
1_3	0	474	67.1%
1_3	1	1,005	57.0%

In other situations, pitchers bunted 6 percent or less of the time, so there was a sharp distinction between bunt and non-bunt pitcher PAs. In the seven bunt situations:

	ADVANCE	OUTS PER PA	RE24 PER PA	STRIKEOUTS
Bunt	67.4%	0.98	-0.253	15.6%
Swing	34.3%	0.89	-0.177	36.6%

Note that outcomes were worse when bunting than swinging, as pitchers had a greater chance to get on base when swinging away.

Arthur, Rob (2017). The fly ball revolution is hurting as many batters as it's helped. https://fivethirtyeight.com/features/the-fly-ball-revolution-is-hurting-as-many-batters-as-its-helped/

Carleton, Russell A. (2017). The fly ball...revolution? https://www.baseballprospectus.com/news/article/32057/baseball-therapy-the-fly-ball-revolution/

The groundball/flyball ratio dropped from 1.34 to 1.25 between 2015 and the first month of 2017. Rob Arthur (2017) graphed the relationship between changes in fly-ball rate and changes in wOBA between 2015 and 2016 and uncovered no overall relationship. For those who increased their rate, 49.3 percent saw a higher wOBA but 50.7 suffered from a lower one. Looking at his graph, it appears that an about 50/50 split also occurred for those who decreased their flyball rate. Russell Carleton (2017) noted that BABIP on non-homer fly balls is only .150 as most are caught. Looking at 2003 to 2016 Retrosheet data for batters with at least 250 PAs in consecutive seasons, Russell extended Rob's work as follows:

Change in Outcome	Correlation with Change in FB Rate
Contact Rate (per swing)	-.114
Strikeout	.093
Walk	.054
Single	-.286

Double/Triple	.093
HR	.341
Out in Play	-.047
OBP	-0.01

Arthur, Robert (2014i). How quickly do team results stabilize? https://www.baseballprospectus.com/news/article/23423/moonshot-how-quickly-do-team-results-stabilize/

Based on 2000-2013 Retrosheet data, Robert Arthur (2014i) determined that starting at about the 30th game, team runs scored and given up is predictable to an average of about ½ run per game, which is the best that Baseball Prospectus's projection tool PECOTA was capable of at that time.

Bain, Derek (2018). Ball-strike outcomes: Gaining the upper hand. http://www.tuatarasoftware.com/baseballanalytics/2018/11/16/ball-strike-outcomes-gaining-the-upper-hand/

In another such analysis using Retrosheet data, Derek Bain (2018) presented BA, SA, and HR/AB for at bats ending on every count plus overall figures between 1998 and 2017. Overall, hitter's counts (more balls than strikes) revealed increases; the overall numbers in 1998 were .309, .484, and 3.2; by 2017 they had gone up to .353, .631, and 6.4, with much of the rises occurring by 1994 but further jumps starting about 2014. The remaining neutral counts, 0-0 and 1-1, basically mirrored hitter's counts. In pitcher's counts (more strikes than balls, plus 2-2), the overall trajectory has been a bit down for BA (a bit over .200 to about .196), well down for SA (about .550 to about .475), but up for HR/AB (about 1.4 to 2.3, with the bulk of the increase again starting in 2014. This latter generalization hides variation among very specific counts; for example, all three rose for 0-1 counts.

Baumer, Ben S., James Piette, and Brad Null (2012). Parsing the relationship between baserunning and batting abilities within lineups. *Journal of Quantitative Analysis in Sports,* Vol. 8 No. 2, Article 8.

Beyond base-out situation, the risk of attempting a steal (along with other speed-related moves such as taking extra bases on hits) depends on the specific abilities of the player making the attempt. Obviously, some players are better basestealers and/or baserunners than others, and the risk is lower the better the player is on the basepaths. Through a simulation based on the "team based on a given player" method for

evaluating offense and using 2007-2009 Retrosheet data, Baumer, Piette and Null (2012) examined the expected outcomes of such attempts for 21 players purposely chosen for their variety of capabilities as hitters and baserunners. Their results suggest that taking the risk of the steal or extra base is more productive long-term to the extent that the player is a good baserunner and a less productive hitter. This is because the cost of an out on the attempt is unsuccessful is greater for a better hitter than a poorer one. Although they interpret this in the context of the chosen individual players, the real implication is that attempting the steal or extra base makes more sense when the next batter is weak, as that next batter could use the help of the extra base for driving the baserunner in.

Baxamusa, Sal (2006). The memory remains. https://tht.fangraphs.com/the-memory-remains/
Baxamusa, Sal (2007). More on pitch sequences. https://tht.fangraphs.com/more-on-pitch-sequences/

Not only does the count, matter, but here is evidence that the order in which a specific count is reached may matter also. The following is 2005 from N.L. Retrosheet data for 1-1 counts:

Situation	AVG	OBP	SLG
After 1-1 pitch (entire PA)			
First pitch strike	.257	.314	.402
First pitch ball	.243	.312	.378
Ball in play on 1-1 pitch			
First pitch strike	.336		.528
First pitch ball	.299		.472

and

Ball in play on 1-1 pitch	AVG	SLG
First pitch swinging strike	.303	.486
First pitch called strike	.338	.532
First pitch foul strike	.346	.528
Second pitch swinging strike	.261	.367
Second pitch called strike	.290	.452
Second pitch foul strike	.328	.472

However, in 2007c, Sal tried the same analysis with 3-2 counts for two very different sequences (two called strikes and three balls versus three balls and two called strikes), and uncovered either nothing or inconsistencies across leagues with the exception of strikeouts (more for the second sequence), calling into question the generality of the 1-1 findings.

Baxamusa, Sal (2007). Can't find the strike zone? https://tht.fangraphs.com/cant-find-the-strike-zone/
NOT IN BIBLIOGRAPHY, IN REFERENCES

Batter's responses on the fourth pitch when in 3-0 counts, using 2006 Retrosheet data for 8049 plate appearances.

Called strike 59.6%
Ball four 33.8%
In play 3.1%
Foul ball 2.5%
Swinging strike 0.8%

As Sal notes, keeping the bat on the shoulder is generally a good strategy here as the odds are one-third of getting a walk whereas a strike still leaves the batter well ahead in the count. Those who chose to swung were generally more powerful batters (slash line of .278/.349/.475) than those who did not (.270/.338/.437).
Here are 2006 outcomes for fifth pitches after three balls and a strike:

Result of 3-1 pitch in play after count started 3-0

Sequence	AVG/SLG	N
BBBCX	.358/.496	1280
BBBFX	.340/.510	47
BBBSX	.272/.272	12

As before, swinging strikes resulted in worse outcomes than fouls or called strikes, but as the sample size was tiny this result could have been a fluke.

Baxamusa, Sal (2007). The long and short of plate appearances. https://tht.fangraphs.com/the-long-and-the-short-ofplate-appearances/

This piggybacks on Tom Tango's pitch count estimator:

$$(3.3 \times PA) + (1.5 \times SO) + (2.2 \times BB)$$

which implies that the average batted ball should occur after 3.3 pitches, average strikeout after 4.8 pitches, and average walk after 5.5 pitches. Sal Baxamusa (2007) used Retrosheet data to determine that in 2006, the actual figures were 3.3, 4.8, and 5.5, so the formula did well that season. Interestingly, every hit type also averaged 3.3 pitches with the exception of triples (3.5, but with a small sample size).
Looking at a graph, approximately 18 percent of plate appearances lasted either 3 or 4 pitches, with about 17 percent going 2 pitches, 16 percent 5 pitches, 12 percent 1 pitch, maybe 10½ percent 6 pitches, 4½ percent 7 pitches, 2 percent 8 pitches, 1 percent 9 pitches, and progressively less often afterward. Sal also showed a graph for the

outcomes of different numbers of pitches, but these were not as interesting as strikeouts cannot appear until the third pitch and walks until the fourth. The length of a PA had no effect on the type of batted ball.

Baxamusa, Sal (2007). Strikethrowers and control freaks. https://tht.fangraphs.com/strikethrowers-and-control-freaks/
Baxamusa, Sal (2007). In search of efficient pitchers. https://tht.fangraphs.com/in-search-of-efficient-pitchers/

Continuing this line of inquiry, Sal (2007) noted that in 2006 it took an average of 5.79 pitches to get a walk from pitchers with strikeout/walk ratios of better than 3. This makes sense because it implies that it takes longer than average to get a walk from a strike thrower. Looked on analogously, for those with walk/PA rates of less than 5 percent, the average was 5.92. PAs ending with strikeouts, hits, and outs on balls in play did not differ from the averages from the first of this webpost sequence. The second of these webposts had similar findings for strikes per pitch.

Baxamusa, Sal (2007). It's up to the hitter. https://tht.fangraphs.com/its-up-to-the-hitter/

In the last webpost in this series, using 2006 Retrosheet data for batters with 200 or more PAs (sample size of almost 300), Sal compared the top 20 and bottom 20 to see if the length of their PAs differed systematically. Contact was measured by percentage of PAs ending with balls in play, and passivity by percentage of strikes that were called rather than whiffs or fouls,

		----Contact-----		-----Passive----	-	--Pitches/PA---	
Result	Average	Top 20	Bottom 20	Top 20	Bottom 20	Top 20	Bottom 20
BB	5.67	5.44	5.67	5.56	5.58	5.74	5.32
K	4.81	4.71	4.98	4.91	4.71	5.05	4.62
Hit	3.34	3.13	3.53	3.70	2.89	3.75	2.87
Out	3.32	3.19	3.46	3.67	3.05	3.71	2.98
All	3.75	3.38	4.16	4.07	3.37	4.29	3.23

We see that PAs for top 20 contact hitters were shorter and top 20 pitches per PA longer than their bottom 20 counterparts no matter the ending, and that top 20 passive hitters were the same as top 20 pitch/PA hitters except for walks.

Belleville, Gary (2021). Who threw the greatest regular-season no-hitter since 1901? *Baseball Research Journal*, Vol. 50 No. 1, pages 60-68.

Although not accompanied by a formal analysis, it is pretty clear from a diagram offered by Gary Belleville based on Retrosheet data that there is a pretty sizable negative correlation between league batting averages and the number of no-hitters in a season.

Beltrami, Edward and Mendelsohn, Jay (2010). More thoughts on DiMaggio's 56-game hitting streak. *Baseball Research Journal,* Vol. 39 No. 1, pages 31-34.

This is one of several attempts to estimate the probability of occurrence of Joe DiMaggio's 56 game hitting streak. Beltrami and Mendelsohn used the number of hits per game DiMaggio averaged in 1941 (1.39), simulated the expected number of games in a 56 game stretch with hits given that figure and an otherwise random process (about 45), and determined that 56 is significantly more than that at better than .01. An analogous study of Pete Rose's 44 game streak using Retrosheet data had similar results.

Bendtsen, Marcus (2017). Regimes in baseball players' career data. *Data Mining and Knowledge Discovery,* Vol. 31, pages 1581-1620.

Bendtsen (2017) defined a *regime* as a phase in a position player's career within which offensive performance is relatively consistent for a significant period of time, but distinctly different than beforehand and after wards. The author evaluated a model for determining regimes and the boundaries between them using 30 seemingly randomly-chosen players whose careers began no earlier than 2005 and who had at least 2000 entries in Retrosheet, the source of study data. The number of regimes for the chosen players ranged from 3 (with one exceptional 2) to 6 and averaged 4.36; and the sample includes quite a few who were still playing when the data ended, meaning this average is almost certainly an underestimate of the number of regimes the sample will accumulate in their careers. Only forty percent of the boundaries between regimes could be accounted for by reported injuries, changes in teams, or a new season; the other sixty percent occurred within-season for no discernible reason. In addition, all but two had separate regimes that were statistically analogous. A detailed examination of two of the sample (Nyjer Morgan and Kendrys Morales) shows that differing regimes generally reflect obviously different OPS values for substantial periods of time.

Biolsi, Christopher, Brian Goff, and Dennis Wilson (2022). Task-level match effects and work productivity: Evidence from pitchers and catchers. *Applied Economics*, Vol. 54 No. 25, pages 2888-2899.

Biolsi, Goff and Wilson used Retrosheet data from 2000 to 2017 to examine another possible defensive interdependence, that between pitchers and catchers regarding getting outs of all types and strikeouts. There were a total of 5519 pitcher-catcher matches between 2000 and 2017; the authors used the 75 percent most active of those for each analysis. There was a lot of variation across seasons, but in general for individual seasons, pitchers had the most impact on both outs and strikeouts, then catchers, and finally the specific pitcher-catcher match the least. However, when combined across seasons, the match had more impact than the catcher and, in the case of outs, almost as much as the pitchers. The authors' proposed explanations for the difference between within and across season findings were that (1) the increase in sample size obtained from combining seasons reduced noise that appeared in the yearly individual pitcher and catcher coefficients, and that (2) good pitcher-catcher

matches take time to develop and the development time was reflected in the single season data; this latter proposal was supported when examining factors potentially affecting the overall results. In addition, pitcher-catcher matches were slightly more influential when the two came from the same country and, more strongly, spoke the same first language, and when their MBL debuts had been closer together in time.

Birnbaum, Phil (2000). Run statistics don't work for games. *By The Numbers,* Vol. 10 No. 3, pages 16-19.

The value of offensive indices such as Pete Palmer's Batting Runs and Bill James's Runs Created is that they represent the impact of offense on team run scoring over a season. But they do not work well for predicting team run scoring in individual games. As Phil argued, this is because run scoring is <u>not</u> a linear function of hitting. For example, it would not be surprising for a team to score one run if it got five hits. But maintaining that five-to-one ratio quickly becomes absurd. Two runs scored on ten hits does happen, but is noticeably underproductive. How about three runs on fifteen hits? Four runs on twenty hits? Runs happen when hits (and walks, and extra bases) do not occur randomly over innings but are bunched together. After making this argument, Phil shows that Batting Runs, Runs Created, and his own Ugly Weights are unsuccessful at predicting run scoring in games.

Birnbaum. Phil (2000). The run value of a ball and strike. *By The Numbers,* Vol. 10 No. 1, pages 4-6.

Phil used 1988 Retrosheet data to compute the average linear runs relative to zero that a plate appearance ends up producing for each count passed through on the way to the plate appearance's completion. The data was as follows:

	0 strikes	1 strike	2 strikes	3 strikes
0 balls	.0000	-.0365	-.0874	-.2736
1 ball	.0288	-.0119	-.0680	-.2734
2 balls	.0829	.0290	-.0306	-.2732
3 balls	.1858	.1252	.0578	-.2733
4 balls	.3137	.3137	.3135	

Not surprisingly, the better the count for the batter, the better the outcome. Phil also computed the average value of the strike (-.0829) and ball (+.0560), and noted that the sum of the absolute values of these (.1389) would be the value of a catcher framing a pitch successfully, such that a "true" ball is called a strike.

Birnbaum, Phil (2000). Does a pitcher's "stuff" vary from game to game? *By The Numbers,* Vol. 10 No. 4

There is not much evidence that a bad first inning is indicative of an off-day for a pitcher, such that the manager should pull him quickly and tax his bullpen for the rest of

the game. Phil Birnbaum (2000), using Retrosheet data from 1979 to 1990, examined the subsequent performance of starters giving up three, four, and five first-inning runs. Overall, starters averaged an RC/27 (see the Batting Evaluation chapter for that) of 4.30. Starters who gave up three first-inning runs averaged an RC/27 of 4.51 for the rest of the game; but their overall RC/27 for the season was almost the same, 4.54. In other words, they were not having a particularly bad game for them as overall they were somewhat worse pitchers than average. The same for four runs in the first; 4.56 the rest of the game, 4.57 overall. In contrast, five runs might be an indication; 5.58 the rest of the game versus 4.67 overall. However, Phil warns us of some potential problem with this data. First, the multiple-run innings are included in the seasonal figure but not the after-the-first innings. If the multiple run innings were subtracted from the overall, as they really should be in this study. it might be noticeably lower than this study's findings and from the after-the-first performance. Second, some pitchers are removed after the first and so are not represented in the after-the-first data, and these might just be the pitchers who really are having an off-day which is recognized as such by the manager or pitching coach.

Moving to the other end of the game, a lot of baserunners allowed in the late innings might well be an indicator of a tiring pitcher. Three baserunners in the first (I assume this includes more than three) resulted in a 4.35 RC/27 when it was 4.07 overall; in the eighth, 4.50 versus 4.00; in the ninth, 4.37 versus 3.89.

Birnbaum, Phil (2003). Applications of win probabilities. *By The Numbers,* Vol. 13 No. 1, pages 7-12.

Using Retrosheet data from 1974 to 1990, Phil covered the value of intentional walks and relief pitching as examples of, as he titled the article, applications of win probabilities. Most importantly, in the relief pitcher section, Phil defined a measure of "clutchiness" that he called "relative importance" of a given situation. Tom Tango was working on the same idea about that time, and Tango's label (leverage) is the one that stuck.

Birnbaum, Phil (2005). Do some teams face tougher pitching? *By the Numbers,* Vol. 15 No. 1, pages 9-12.

In the 1986 *Baseball Abstract* (pages 238-239), Bill James did a quick-and-dirty examination of a claim made by Garry Templeton that the Padres had faced an inordinate number of front-line pitchers the previous year. Phil Birnbaum (2005) decided to examine the question in detail, using Retrosheet data from 1960 to 1992. He used Component ERA as it is less impacted by luck than regular ERA, and adjusted for ballpark and overall team pitching quality, plus a shrinkage of variation from the mean for pitchers with fewer than 50 innings to correct for extreme random aberrations. The largest difference between opponent and league CERA was about 0.15, translating to about 25 runs a year, which makes Bill's estimate of 2½ games to be sensible as an extreme case. However, the standard deviation of differences was .043, or seven runs

per season, which means that for most teams quality of opponent pitcher might account for one game a season.

Birnbaum, Phil (2008). Clutch hitting and the Cramer test. *Baseball Research Journal,* No. 37, pages 71-75, and *By the Numbers,* Vol. 15 No. 1, pages 7-13.

The first serious attempt to evaluate whether there is such a thing as a clutch hitter was a study by Richard Cramer in the 1977 *Baseball Research Journal* showing very little relationship between a measure of clutch hitting for players in two consecutive seasons. Phil's work is a response to Bill James's claim in the 2004 *Baseball Research Journal* that this type of study is fundamentally flawed, because the comparison of measures across seasons multiplies the measurement error of each measure to the point that finding no difference is just as likely due to that error as the absence of clutch hitting as a skill. Phil first used Retrosheet data to correlations between the differences between clutch and non-clutch batting averages (defined as Elias LIP) for players with at least 50 clutch ABs in every pairing of two seasons from 1974-1975 to 1989-1990.(excluding the two pairings including the 1981 strike season). Interestingly, 12 of the 14 correlations were positive, but all of these positives were less than .1, and the overall average correlation was .021. Second, Phil simulated what the distribution of these clutch-non clutch differences would have been if clutch hitting is a randomly distributed skill, such that about 68% of the players had a difference between 1 and -1 s.d.'s from mean, 28% had a difference either between 1 & 2 s.d.'s or -1 and -2 s.d.'s from mean, and 5% more extreme than either 2 or -2 s.d.'s. In this case, the mean correlation across two-season pairings was .239 and was likely to occur by chance less than five percent of the time for 11 of the 14 seasons. Thus it was likely that if clutch hitting was a randomly distributed skill, Cramer would have evidence for it. Third, Phil computed the statistical power for such correlations, and noted that if clutch hitting was a skill but weak enough such that the season-by season correlation was only .2, the odds of Cramer's method would still have a 77 percent chance of finding it. Statistical power for a correlation of .15 was still slightly in Cramer's favor (.55) and finally drops below that (.32) with a correlation of .10. The conclusion we must reach is that if clutch hitting actually exists, its impact on performance must be extremely small, less than would have any appreciable impact on what occurs during a game, because if there was any appreciable difference between clutch and choking players it would have been revealed in these tests.

Birnbaum, Phil (2011). Scorecasting review. https://www.baseballprospectus.com/news/article/13003/baseball-proguestus-scorecasting-review/

Phil Birnbaum (2011), in response to the claim by Moskowitz and Wertheim (hereafter MW) in their book *Scorecasting* that pitch calls favor the away team in low-leverage situations, argued that this implies that the home team scoring advantage over away teams should be highest when leverage is highest, which tends to be in the last innings.

Using Retrosheet data from 1957 to 2007, here are the inning-by-inning differences in run scoring, contrary to MW.

Inning	Runs	Percent
1	61872-52071	+18
2	46823-42539	+10
3	53590-48188	+11
4	53357-49593	+8
5	53203-48448	+10
6	54401-50603	+8
7	52231-48641	+7
8	50451-47781	+6

To try and concentrate on low-leverage situations, which focuses on the MW claim more directly, Phil restricted the following to four run leads by either team-based

Inning	Runs	Percent
2	2543-2139	+19
3	4583-4176	+10
4	8817-7801	+13
5	10940-10057	+9
6	14371-13279	+8
7	15698-14583	+8
8	16935-16180	+5

And just to away teams ahead by four or more

Inning	Runs	Percent
2	957-1022	-6
3	1974-1799	+10
4	3609-3355	+8
5	4435-4645	-5
6	6269-5705	+10
7	6627-6562	+1
8	7309-7179	+2

Phil's conjecture concerning the second inning; if the visitors had scored four more runs than the home team in the first inning, it is likely that, more than not, their lineup is at the least productive bottom whereas the home team is in the productive middle. Anyway, the evidence points to the advantage being greater in the early innings when leverage is usually lower, contrary to what Phil thought the MW claim implies.

Bond, Brittany, and Ethan Poskanzer (2024). Striking out swinging: Specialist success following forced task inferiority. *Organization Science*, Vol. 35 No. 2, pages 698-718.

Based on 1999 to 2018 data from Retrosheet, the authors uncovered evidence that when pitchers batted and made out, they were slightly more likely to get the next half inning's leadoff hitter out. This effect decreased with subsequent batters and was gone by the fourth. It was also greatest with a tied score, also decreases with differences in score and disappearing with a four run margin. Pitchers were also more likely to throw strikes and walked fewer leadoff hitters after making out at the plate. The effect added up to 0.018 runs. There was no impact for previous pitching performance on pitcher batting. When interviewed on the topic, several MLB pitchers reported that making out at the plate motivated them to pitch more aggressively.

Borowsky, Larry (2005). Let 'er rip.
http://baseballanalysts.com/archives/2005/08/let_er_rip_1.php

2004 Retrosheet data, N.L.:

when batter...	avg	obp	slg	rc/27
takes ball on 0-0	.270	.382	.452	6.4
swings on 0-0	.270	.300	.439	4.9
takes strike on 0-0	.233	.275	.353	3.4
total takes on 0-0	.253	.337	.407	5.0

Note that RC/27 was almost the same for swings and total takes.

Borowsky, Larry (2006). Let 'er rip II.
http://baseballanalysts.com/archives/2006/07/let_er_rip_ii_1.php

Retrosheet data, 2005:

	2004				2005			
when batter . . .	avg	obp	slg	rc/27	avg	obp	slg	rc/27
takes strike on 0-0	.233	.275	.353	3.4	.237	.271	.357	3.4

swings on 0-0	.270	.300	.439	4.9	.269	.296	.440	4.8

Very consistent across seasons.

Boynton, Bob (1999). Umpire bias revisited. *Baseball Research Journal,* No. 28, pages 96-100.

This piece followed up on two earlier *BRJ* articles, by Richard Kitchin in No. 20 and Willie Runquist in No. 22, in which Kitchin presented data implying that when assigned to home plate specific umpires were biased either for or against the home team in their pitch judgments. Such bias resulted in large differences in walks and strikeouts, which filtered through to runs scored and home team winning percentages. Runquist countered with evidence that such differences were statistically insignificant. Using a much larger sample of at least eight seasons per umpire over the 1988-1996 interval with data from Retrosheet (which he mistakenly referred to as Project Scoresheet), Bob Boynton (1999) noted some ten umpires that were either above or below league mean in walks (Bob labeled his measures that way: I hope he analyzed all of them at per game rates) in every or all but one season. Although walks correlated with runs scored at .72 in the A. L. and .57 in the N. L., only three umps were as consistently above or below mean in runs scored, and none were consistently above or below mean in home team winning percentage. The implication is that there indeed are hitter umps and pitcher umps, but they call them consistently for both home and away teams, so such biases are harmless in their outcome.

Bradbury, John Charles and Douglas Drinen (2006). The designated hitter, moral hazard, and hit batters. *Journal of Sports Economics,* Vol. 7 No. 3, pages 319-329.

Bradbury, John Charles and Douglas J. Drinen (2007). Crime and punishment in major league baseball: The case of the designated hitter and hit batters. *Economic Inquiry*, Vol. 45 No. 1, pages 131-144.

Baldini, Kevin, Mark T. Gillis and Mark E. Ryan (2011). Do relief pitching and remaining gams create moral hazard problems in major league baseball? *Journal of Sports Economics,* Vol. 12 No. 6, pages 647-659.

There is a surprisingly large literature on whether hit-by-pitches are the result of strategic choice on the part of the pitcher and manager of the opposing team. The impetus of this work was the substantial increase in HBP in the American League after the appearance of the designated hitter, implying that pitchers may be more willing to hit someone when retaliation against them personally will not occur. An alternative hypothesis has been that when retaliating, pitchers are more likely to throw at good batters than poor because the former are more likely to get on base anyway, so pitchers, as generally the poorest hitters on a team, are the least likely targets. Bradbury and Drinen performed two studies that provided far better examinations of the retaliation hypothesis than those previous through use of Retrosheet 1973-2003 data.

Based on game-by-game information, they first (2006) noted evidence for both hypotheses in predictive model allowing for determination of the order of importance of associated variables. The variable most strongly associated with hit-by-pitches was whether the game had designated hitters, with this effect occurred in interleague games including NL teams, evidence against the idea that HBPs are just idiosyncratic to the AL but perhaps due to pitchers not batting. However, the difference between leagues largely disappeared in the 1990s. On the other side of the dispute, the second most associated variable was total runs scored, evidence that when teams are hitting well the other side finds less reason not to hit batters. Further, home runs by the other team were also associated, more evidence that a HBP against a powerful batter would be considered less harmful. Finally, and not surprisingly general pitcher wildness was also correlated. In their second (2007) paper, Bradbury and Drinen determined whether a hit-by-pitch in one half inning increases the odds of retaliation in the next. According to two analyses, one for 1969 combined with 1972 through 1974, the other for 1989 through 1992, it does, as does a home run by the previous batter in the more recent data set; both of these findings support the retaliation hypothesis. Consistently with the second argument, higher OPS was positively associated with HBP whereas pitchers were less likely to be plunked than everyone else; both of these results suggest the "less harm" hypothesis. In addition, large score differentials increase HBP, likely because there is less harm when such a differential leaves less doubt concerning which team will probably win the game. Again, wilder pitchers are, not surprisingly, more likely to hit batters.

Bradbury and Drinen also replicated an earlier finding that HBP exploded during the 1990s, particularly in the case of the National League, whose numbers came to approximate that of the American despite the absence of the DH. The authors believed it to be a perverse result of the rule change authorizing umpires to warn both teams not to retaliate, as it lowers the chance that pitchers will be plunked, thus leading them to feel free to throw at hitters and consistent with the first hypothesis.

Baldini**,** Gillis, and Ryan (2011) replicated the Bradbury/Drinen method (extending the Retrosheet data set through 2008) with two additional variables. First, as previously hypothesized by Stephenson (*Atlantic Economic Journal,* Vol. 32 No. 4, page 360), as relievers almost never come to bat in the National League, their plunking tendencies would not differ from American League relievers as it would for starters. Second, as the number of games left in the season decreases, the opportunity for retaliation is less likely, so HBPs should increase as the season goes on. There are a number of interesting findings relevant to the general idea. First, relievers hit more batters than starters across leagues, probably due to poorer control in general, but the difference is greater in the N.L., which the authors argued is due to their not being as concerned at being hit themselves as would A. L. relievers. Second, the more relievers in a game, the more HBPs, perhaps analogously due to the additional relievers being wilder, but the difference between leagues becomes smaller as the number of relievers per game (disappearing at five), again perhaps implying that more relievers decreases the odds that any of them would bat and so again lowering their concern. Third, HBP in general slightly increase as the season progresses, less so in the National League, but

decrease between specific teams, which is not at all consistent with expectation. The authors conclude with the interesting speculation that the reason that the overall league difference in HBP has disappeared may partly be due to the fact that the number of relievers used in a game has increased markedly.

Bradbury, John Charles and Douglas J. Drinen (2008). Pigou at the plate: Externalities in major league baseball. *Journal of Sports Economics,* Vol. 9 No. 2, pages 211-224.

John Charles Bradbury and Douglas Drinen (2008) is oen of several studies that punctures the myth that fielding a lineup with two good hitters in a row "protects" the first of them, meaning that the pitcher is more willing to chance getting him out (and so perhaps give him hittable pitches) than pitching around him (making it likely it he walk and thus be a baserunner for the second to drive in. They contrasting the "protection hypothesis" with an "effort" hypothesis in which pitchers put more effort into retiring the first hitter to try and ensure that he won't be on base for the second. The protection hypothesis implies that a good on-deck hitter will decrease the walks but increase the hits, particularly for extra bases, for the first hitter; the effort hypothesis predicts decreases in all of these indices. Retrosheet data from 1989 to 1992 supported the effort hypothesis; on-deck batter skill as measured by OPS was associated with decreased walks, hits, extra-base hits, and home runs, with the association increased by a standard platoon advantage for the on-deck hitter. This support, however was weak, as a very substantial OPS rise of .100 for the on-deck hitter amounted on average to a drop of .002 for the first hitter. The authors mention an additional and important implication; contiguous plate appearances appear not to be independent, contrary to so many of the most influential models for evaluating offense. However, if their data is representative, the degree of dependence may be too small to have a practical impact on these models' applicability.

Bradbury, J. C. (2011). *Hot Stove Economics.* New York: Copernicus Books.

In his book, Bradbury used 1989-1992 data to examine differences in overall hitting and pitching between situations with runners in and not in scoring position as a proxy for clutch hitting. The effect was statistically significant due to sample size but tiny in practical terms.

Bradbury, John Charles (2019). Monitoring and employee shirking: Evidence from MLB umpires. *Journal of Sports Economics,* Vol. 20 No. 6, pages 850-872.

John Charles Bradbury (2019) used 2000 to 2009 Retrosheet data to examine the impact of QuesTec on ball/strike calls. In short, 11 ballparks were equipped with QuesTec systems between 2001 and 2008 that allowed for the evaluation of home plate umpire calls. In short, the ballparks with QuesTec had a smaller proportion of called strikes than the ballparks without it, to the tune of .016 per PA or .81 per game on

average. This impact was overwhelmed by other factors, most notably a directive to umpires to be more accurate, leading to the called strike rate to increase by two percent between 2000 and 2001 (the year of the directive) and another ½ percent in subsequent seasons. As for the effect of control variables: Consistent with past research, there were fewer called strikes for home team batters, which is part of one of the research-supported explanations for home team advantage, crowd noise; yet more called strikes due to the attendance/home team batter interaction, which is inconsistent with that explanation. In addition, there was deference for experienced batters and pitchers (consistent with past work) and more called strikes for catchers (inconsistent with the literature).

Breunig, Robert, Bronwyn Garrett-Rumba, Mathieu Jardin and Yvon Rocaboy (2014). Wage dispersion and team performance: A theoretical model and evidence from baseball. *Applied Economics,* Vol. 46 No. 3, pages 271-281.

Matching 1985-2010 Retrosheet data with salary figures, Bruenig et al. replicated earlier findings by several other researchers in noting improved team performance with payrolls that are higher and more equal among players.

Brill, Ryan S., Sameer K. Deshpande, and Abraham J. Wyner (2023). A Bayesian analysis of the time through the order penalty in baseball. *Journal of Quantitative Analysis in Sports*, Vol. 19 No. 4, pages 245-262.

Is there really a times through the order penalty, or instead is there a steady degradation of pitcher effectiveness as the game progresses. These authors' work, based on 2012 to 2019 Retrosheet data, supports the latter. They controlled for batter and pitcher quality (vua wOBA), handedness, and home versus away team. Very importantly, they limited their sample to starts in which the pitcher did not get through the second time through, in so doing protecting their work from the selection bias that most TTOP studies have suffered from. Using an expected wOBA for each plate appearance through the 27th as estimated by their models resulted in a linear degradation of starter performance across the game. Batter and pitcher quality were stronger, and handedness and home-away status about equivalent predictors of expected wOBA compared with this in-game performance drop.

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, pages 293-329.

Brill and Wyner (2024) have offered a novel approach to computing WAR figures for pitchers, which they call Grid WAR (gWAR). They raise two issues with the bWAR and fWAR approaches to pitcher evaluation. The first is based on their belief that WAR should be context-dependent. Their specific argument, which is in truth only relevant to fWAR for pitchers as it is FIP-based, 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. In their estimation, WAR should be based on runs actually given up (which bWAR does, with adjustments), not estimated runs based on event likelihood. This would make Grid WAR a better descriptive metric than fWAR, but a worse predictive one, and I personally disagree as I believe that WAR should be a predictive metric.

I am more comfortable with the authors' second issue. Pitchers who are more variable in their performance will be unfairly penalized by averaging over their performance, as an occasional blow-up can cancel out several good performance. For example, a pitcher who gives up 1 run in 6 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 a good 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 de-value them). The authors also make the excellent 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 the method, which, as it is highly mathematical, 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) with a ballpark adjustment, to which pitcher-specific actual performance in individual games is compared, with a replacement level based on fWAR's. Their analysis applies Retrosheet data from 2010 to 2019. (Incidentally, in 2019, Justin Verlander and Gerrit Cole led the way with over 7.5.) Comparing pitchers, gWAR was indeed higher for pitchers with more variable performances when fWARs were equivalent; in other words, gWAR was higher for pitchers with a higher proportion of both well- and poorly-pitched games than pitchers who were more consistent. The authors also demonstrated that one-season gWAR predicted next-season gWAR better than fWAR, which in and of itself means nothing but does demonstrate some across-year consistency in within-year inconsistency, which is interesting. They also demonstrate 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.

Bruschke, Jon (2012). The Bible and the Apocrypha: Saved Runs and Fielding Shares. *Baseball Research Journal,* Vol. 41 No. 1, pages 12-19.

Bruschke (2012) offered a fielding metric based on a completely different logic than zone approaches. In his own words, “In a nutshell, zone approaches carefully measure individual performance, but estimate productivity [by that, he means total team

success at saving runs via fielding). My approach measures productivity directly but estimates individual performance" (page 14). He called it Fielding Shares, and that is an apt title, as, analogously with Bill James's Win Shares, it begins with team performance and divides it among the players responsible for it.

began by regressing defense-independent pitching indices (strikeouts, walks, and home runs per plate appearance and infield popups per batted ball) on runs per game for 2008 and 2009. These indices combined, the pitcher's share of defense so to speak, accounted for 64 percent of the variance in runs scored; the remaining 36 percent is the fielder's share. He then transformed each team's regression residual (which correlated .64 with batting average on balls in play, an indicator that the two are likely measuring related phenomena) and BABIP into scales ranging from 50 to 100 and summed the two transformed figures, resulting in somewhere between 100 and 200 total fielding points for each team. This measure correlated much more closely with team wins (.44) than Dewan's plus/minus measure (.185), which should not be a surprise given the respective logics mentioned earlier. Next, using 2008 Retrosheet data as the basis, he assigned every out on balls in play to the responsible fielder, crediting putouts to the player making it on unassisted plays and assists to those making it (.5 if two players, .3 if three) on assisted plays. Finally, he calculated the proportion of these for each fielder, and then assigned that proportion of total team fielding point to that player as his Fielding Shares, after correcting for how much that fielder played.

This last move, in my opinion, a mistake given what this index is intended to indicate, as players who play less make a smaller contribution to total team fielding performance, as is recognized in Win Shares. The method also presumes that every fielder has an equal opportunity to make plays, which is obviously wrong given that the number of batted balls differs substantially among positions. This would be a fatal flaw if the intention was to actually evaluate fielders rather than determine responsibility for overall team fielding performance.

Burnson, John (2007). Tug of war. In David Studenmund (Ed.), *2007 Hardball Times Baseball Annual* (pages 161-164). Skokie, IL: ACTA Sports.

To what extent is the batter and the pitcher responsible for the outcome of a plate appearance. John Burnson (2007)'s very interesting take on this matter was based on analysis of batter decisions during at bats. Based on Retrosheet data from 2003 to 2005, the following tables began his demonstration:

The odds of a swing on a pitch for a given count

		Balls			
		0	1	2	3
	0	28%	41%	40%	8%
Strikes	1	46%	40%	59%	56%
	2	49%	58%	65%	74%

Batters are most likely to swing with two strikes. Are they trying to protect themselves from the embarrassment of being called out on strikes?

The odds of a called strike if no swing

		Balls			
		0	1	2	3
	0	42%	40%	47%	63%
Strikes	1	20%	23%	27%	36%
	2	8%	10%	13%	17%

Pitchers are least likely to throw a strike with two strikes. Is it because they realize that batters are likely to swing anyway, so they might as well make it hard for the batters to hit?

Now, let us break down the 3-2 count. Overall, as noted above, batters swing 74 percent of the time and pitchers throw strikes 17 percent of the time. However, as the number of pitches with a 3-2 count increases from 5 to 12 given foul balls continuing the plate appearance, the batter swinging percentage rises fairly steadily from 73% to almost 80% whereas the percentage of called strikes with no swing falls just as steadily from about 17½% to about 14½%. Again, batters seem to lose their patience and pitchers seem to take advantage of that loss.

In the rest of Burnson's essay, based on pooling specific batter/pitcher pairings that occurred at least 30 times between 2003 and 2005, he concluded that hitter ground ball rate accounts for 65%, batter strikeout rate 69%, and batter walk rate 63% of the odds that grounders, whiffs, and walks would occur on a given at bat.

Bushek, Nathaniel and Scott Erickson (2025). Weighted dyadicity in major league transaction networks. *Journal of Sports Analytics,* Vol. 11, Article 22150218251326427.

This is a study of 21,318 total player transactions (trades and purchases) between 1922 and 2021 as listed at Retrosheet. The transaction year was defined as November 1 of the previous year to October 31 of the relevant year, and teams defined as competitive if their winning average was .520 or greater. Early on, most transactions were within each league, but by the early 1960's the number of inter- and intra-league transactions became about even and remained so afterward. Transactions between competitive teams was less likely than would be expected, transactions between non-competitive teams and between one competitive and one non-competitive team was about as expected. Whether a team was above or below median payroll did not have an impact.

Callahan, Eric, Thomas J. Pfaff and Brian Reynolds (2006). The interleague home field advantage. *By The Numbers,* Vol. 16 No. 2, pages 9-10.

Data from both Retrosheet and mlb.com revealed that between 1997 and 2005, home field advantage in interleague games was .556 in American League home parks and .559 in National League, more than .02 higher than in intraleague games. The authors, Callahan, Pfaff, and Reynolds (2006), made the reasonable argument that the use of the home team's league's rules (DH in the AL, pitcher bats in the NL) and resulting differences in roster design provide an extra advantage to the home team.

Cartwright, Brian (2008). What run estimator would Batman use? (Part II).
https://statspeakmvn.wordpress.com/2008/09/page/3/

In the second part of a four-part series on offensive metrics, Brian Cartwright (2008) used 1956-2007 Retrosheet data at the level of the inning and showed that BaseRuns was a more accurate predictor than a later version of Runs Created and a linear weights formula based loosely on Extrapolated Runs.

Cartwright, Brian (2008). What run estimator would Batman use? (Part III).
https://statspeakmvn.wordpress.com/2008/09/page/2/

Based on Retrosheet 1956-2007 data, here are linear weight estimates of the overall value of events.

Name	Abbr.	LWTS	LWTS_RC
Generic Out	O	-0.234	-0.072
Strikeout	K	-0.277	-0.116
Stolen Base	SB	0.195	0.195
Defensive Indifference	DI	0.129	0.129
Caught Stealing	CS	-0.525	-0.365
Pickoff	PK	-0.217	-0.109
Wild Pitch	WP	0.276	0.276
Passed Ball	PB	0.270	0.270
Balk	BK	0.265	0.265
Other Advance	OA	-0.471	-0.334
Nonintentional Walk	NIBB	0.304	0.304
Intentional Walk	IBB	0.173	0.173
Hit By Pitch	HBP	0.329	0.329
Interference	XI	0.354	0.354
Error	ROE	0.495	0.497
Fielder Choice	FC	-0.164	-0.056
Single	1B	0.462	0.465
Double	2B	0.762	0.765
Triple	3B	1.035	1.036
Homerun	HR	1.404	1.404
Double Play	DP	-0.611	-0.449

Cartwright, Brian (2008). What run estimator would Batman use? (Part IV).
https://statspeakmvn.wordpress.com/2008/09/

This is a run expectancy chart for 1956-2007 from Retrosheet data, broken done from left to right (columns 3 to 6) to batter reaching base, baserunner advancement, baserunner out on base, and the effect of making an out on existing baserunners.

EVENT	COUNT	RUNNER ADVANCE		OOB	OUT	LWTS
Out	3819401	0.013	0.026	-0.013	-0.050	-0.024
Strikeout	1161343	0.001	0.002	0.000	-0.055	-0.053
Stolen Base	114587	0.000	0.180	0.000	0.000	0.180
Defensive Indifference	2839	0.000	0.120	0.000	0.000	0.120
Caught stealing	48906	0.000	0.010	-263	-0.015	-0.268

Pickoff	24346	0.000	0.095	-0.197	-0.017	-0.119
Wild Pitch	56520	0.000	0.265	-0.001	0.000	0.263
Passed Ball	15238	0.000	0.259	-0.001	0.000	0.257
Balk	9624	0.000	0.253	0.000	0.000	0.253
Other advance	2502	0.000	0.063	-0.298	-0.040	-0.276
Foul Error	3284	0.000	0.000	0.000	0.000	0.000
Walk	607110	0.244	0.061	0.000	0.000	0.305
Intentional Walk	59403	0.185	0.004	0.000	0.000	0.189
Hit By Pitch	49877	0.251	0.078	0.000	0.000	0.329
Interference	918	0.254	0.109	0.000	0.000	0.364
Error	90717	0.288	0.205	-0.002	-0.001	0.490
Fielder's choice	26606	0.304	0.181	-0.371	-0.152	-0.037
Single	1252776	0.260	0.207	-0.003	-0.002	0.461
Double	314183	0.415	0.332	-0.002	-0.001	0.745
Triple	44499	0.590	0.430	0.000	0.000	1.020
Home Run	178776	1.000	0.404	0.000	0.000	1.404
Double play	192350	0.002	0.023	-0.325	-0.041	-0.341
Triple play	210	0.000	0.003	-1.015	0.000	-1.012
Total	8076015	0.114	0.083	-0.018	-0.034	0.1

Cartwright, Brian (2008). Monkeying with Marcel.
https://statspeakmvn.wordpress.com/2008/08/page/3/

This was an attempt to figure out how much to weigh past seasons relative to one another when trying to project future season performance. When doing projections, one should regress past performance to the mean based on sample size aka number of plate appearances, which projection methods then in use did not do. Brian regressed BB and K rates for 1999 to 2007 toward the mean based on reliabilities previous computed for each, giving him a regression equation for predicting each rat for 2002-2007 using the previous three seasons of data for players with at least 250 PA. Beginning with walk rates, using actual past BBs accounted for 59 percent of variance in "current" season BB being predicted; using regressed rates did better at 61.4 percent. Also as expected the highest weighting in the equation was for the previous season and the lowest for three years previous, respectively accounting for 53, 26, and 21 percent of the 61.4. This implies that instead of the (for example) 5/4/3 in Tom Tango's Marcel projection method (and in Bill James's work), relative weights in projection systems would make it about 6.5/3/2.5 using the same sum of 12.
Brian did strikeout rates similarly. In this case, 69.4 percent of variance was accounted for by real K rates but a better 73.5 percent by regressed rates, with 66, 18, and 16, or relative projection weights of 8/2/2. Note that not only are both the walk and strikeout seasonal weights markedly different from 5/4/3, they are also quite different from one another.

Cartwright, Brian (2008). What run estimator would Batman use? (Part III).
https://statspeakmvn.wordpress.com/2008/09/page/2/

This is Brian Cartwright's version of BaseRuns, which is theoretically the best method for devising an offensive evaluation metric ever devised. Two good citations for learning about it are http://tangotiger.net/wiki_archive/Base_Runs.html and a description by Brandon Heipp from *By the Numbers,* Vol. 11 No. 3, pages 18-19, which is available through http://philbirnbaum.com/. In short, the point of BaseRuns is to measure offense based on the number of baserunners aboard during a player's plate appearances (labelled "A" below), the proportion of them driven in by the player ("B"), the outs made by the player ("C"), and the runs driven by that player by own effort ("D"). Including the number of runs scored by that player would be an error as, with the exception of taking extra bases on hits, other players have done the work. Using Retrosheet 1956-2007 data, Brian's version of BaseRuns, which is more complicated than most others, is

A: (1B + E + 2B + 3B + BB + HBP + IBB – CS – DP)
B: .397 X ([.466 X 1B] + [.493 X E] + [.748 X 2B] + [1.02 X 3B] + [.404 X HR] + [.30 X BB + [.189 X IBB] + [.329 X HBP] + [.038 X SB] + [.01 X CS] + [.39 X O] + [.002 X K] + [.025 X DP])
C: O + K + DP + CS
D: HR

Cartwright, Brian (2008). Error: scorekeeper?
https://statspeakmvn.wordpress.com/2008/10/

This is based on Retrosheet data from 1954 and 1956 through 2007. During that time period, the proportion of batted balls resulting in errors decreased, from 1.8 or 1.9 percent through 1970 to 1.3 percent 2005-2007. Although this could be a signal of more short-handed fielding, the more likely explanation for the drop is more leniency on the part of official scorers. Of batters reaching base, between 6.0 and 6.6 percent were the result of errors every year through 1970, between 5.1 and 5.8 percent every year from 1974 through 1991, and less than 5 percent every year 1994 through 2007; that last year (4.0%) was the lowest of all. BABIP increased 29 points between 1963 and 2007; Brian thought that more forgiving scoring was responsible for six of those points.

Cartwright, Brian (2008). Different factors for different folks, part 1.
https://statspeakmvn.wordpress.com/2008/12/page/2/
Cartwright, Brian (2009). Different factors for different folks, part 2.
https://statspeakmvn.wordpress.com/2009/02/

The first part of this two-part study of relative performance examined 108 players from the "mid-1990s" through 2008 with experience playing both in Japan and elsewhere, with U.S. data from Retrosheet. Brian took their Major League Equivalent figures (which would include minor league play) outside of Japan and

compared them with Nippon Professional League performance. They were divided into the following five categories based on MLE HR percentage: A, greater than 0.65; B, 0.50-0.65, C, 0.30-0.50, D, 0.16-0.30, and E, less than 0.16.

U.S. totals.

Grade	BHFw	SDTf	SIf	DOf	TRf	HRf	SHf
A	5536	0.98	1.08	0.83	0.46	1.14	0.25
B	16069	1.03	1.05	0.92	0.31	1.39	0.23
C	22237	1.05	1.02	0.97	0.43	1.66	0.37
D	18813	1.06	1.01	1.01	0.58	1.82	0.69
E	6920	1.02	0.98	1.19	0.56	2.27	1.13
ALL	69578	1.03	1.02	0.98	0.50	1.55	0.68

Note that the lower the HR% outside of Japan, the greater the improvement there (see HRf column). Double percentage (DOf) also increased more for the lowest HR% batters, single percentage (SIf) went down a bit, and not surprisingly sacrifice bunts (SHf) went up as HR% went down.

The second part worked with a larger data set, 1953-2008 (U.S. data again Retrosheet), and only examined relative home run percentage, but this time further divided by U.S. ballpark home run factors. The categories were redefined: AA, .080+; A, .060 – .080; B, .045 – .060; C, .035 – .045; D, .020 – .035; E, .010 – .020; and F, .000 – .010

HR Factors by overall factor of ballpark vs career HR% of batter

Factor	AA	A	B	C	D	E	F
0.30	0.52	0.58	0.40	0.31	0.37	0.18	0.36
0.40	0.60	0.56	0.50	0.47	0.47	0.45	0.34
0.50	0.69	0.59	0.58	0.59	0.53	0.52	0.34
0.60	1.20	0.69	0.69	0.54	0.66	0.55	0.51
0.65	0.79	0.77	0.67	0.66	0.64	0.72	0.75
0.70	0.92	0.79	0.75	0.69	0.68	0.68	0.71
0.75	0.75	0.83	0.77	0.75	0.76	0.72	0.76
0.80	0.80	0.86	0.83	0.85	0.80	0.77	0.75
0.85	0.96	0.93	0.89	0.86	0.83	0.93	0.79
0.90	0.98	0.91	0.92	0.96	0.92	0.92	0.81
0.95	1.00	1.00	0.98	0.95	0.96	0.96	1.00
1.00	0.97	0.97	1.03	1.04	1.07	1.04	0.95
1.05	1.05	1.12	1.05	1.05	1.10	1.10	1.07
1.10	1.01	1.07	1.11	1.14	1.15	1.18	1.36
1.15	1.11	1.11	1.20	1.16	1.20	1.23	1.46
1.20	1.12	1.16	1.12	1.33	1.29	1.29	1.61
1.25	1.23	1.08	1.19	1.32	1.34	1.44	1.63
1.30	1.17	1.35	1.27	1.34	1.35	1.46	2.21
1.40	1.15	1.23	1.43	1.36	1.59	1.86	1.21
1.50	1.32	1.12	1.43	1.51	1.80	2.14	2.27
1.60	1.56	1.45	1.25	1.83	1.85	1.45	4.05
1.70	1.38	1.63	1.71	1.60	1.75	1.89	3.33

1.90 1.29 1.59 1.93 1.58 2.68 2.90 3.08

It looks like there is an interaction effect here. On top of the overall impact of Japan increasing homer production more for those who were lower in the U.S., it seems that for those with the highest HR%, playing in the U.S. ballparks with the lowest U.S. home run factors were helped more than those playing in the highest home run factor ballparks; and those with the lowest HR% were the exact opposite.

Cartwright, Brian (2009). So how long does it take for BABIP to become reliable?https://statspeakmvn.wordpress.com/2009/01/

Carleton, Russell A. aka Pizza Cutter (2007). DIPS and handedness. https://statspeakmvn.wordpress.com/2007/07/

Here are two studies of the reliability of BABIP. To the extent that not giving up hits on balls in play reflect a pitching skill, measures of it should be adequately reliable given a relatively small sample size. Brian Cartwright (2009) examined this issue with 1979-2008 Retrosheet data. It turns out that it takes a big sample size for BABIP to become reliable. Split half reliabilities for pitchers at least 500 balls in play, a split-halves correlation was 0.174. For 1000 BIP, it was 0.253. At 7500 BIP, it finally reached an almost acceptable 0.696 (sample size of only 48 pitchers). Brian concluded that 7600 BIP was needed for 0.70 reliability, the threshold for acceptance. That would takes seven years at 180 IP a year, assuming three BIP per inning. Russell Carleton, using 2000-2006 Retrosheet data for pitchers with at least 50 balls in play against both lefty and righty batters, the intraclass correlations across seasons were:
Right-handed pitcher and right-handed batter 0.181
Right-handed pitcher and left-handed batter 0.105
Left-handed pitcher and right-handed batter 0.190
Left-handed pitcher and left-handed batter –0.025 {nothing}
The take-home message of both of these efforts is that not giving up hits on balls in play is not a readily observable pitching skill.

Chiu, Yung-Chin and Chen-Kang Chang (2022). Major league baseball during the COVID-19 pandemic: Does a lack of spectators affect home advantage? *Humanities and Social Sciences Communications,* Vol. 9, Article 178.

When compared with 2015 through 2019, neither home field advantage nor fWAR were affected by having no fans in the seats in 2020.

Choe, Justin & Jun Sung Kim (2019). Minimax after money-max: why major league baseball players do not follow optimal strategies. *Applied Economics,* Vol. 51 No. 24, pages 2591-2605.

This is a mostly trite article on the impact of the decision whether or not to swing on the first pitch of a plate appearance on PA outcomes, based on Retrosheet data from every

2010 plate appearance. The most interesting finding was that batters tend to change their decision starting at the third PA in a game from the previous two PAs; e.g., 3rd PA from 1st and 2nd and 4th PA from 2nd and 3rd. My guess is that this is likely a response to pitching changes.

Comly, Clem (2000). ARM – Average Run Expectancy Method. *By The Numbers*, Col. 10 No. 3, pages 11-14.

A number of people have examined outfield throwing by using play-by-play data to compute the proportion of baserunners who advanced an extra base on a hit to a given outfielder along with the proportion of baserunners who were thrown out. Calculating the proportions for each outfielder allows the analyst to compare outfielder arms to one another. In addition, comparing run expectancies for before and after the play, these percentages can be turned into runs saved when a baserunner is thrown out or runs given up when baserunners take the extra base. Most likely the first such method was Clem Comly's Average Run Equivalent Method (ARM), based on Retrosheet 1959 to 1987 data. Clem limited his analysis to singles with runners on first and/or second. The best annual figures in Clem's data were about 10 runs saved and the worst about 7 runs lost.

Cramer, Dick and Pete Palmer (2008). Clutch hitting revisited. *Baseball Research Journal,* No. 37, pages 85-88.

This is a second response to Bill James's 2004 article critiquing the method Cramer used in his pioneering research questioning the existence of clutch hitting as a skill (see Phil Birnbaum, 200i8, above). Using the same method as before but here with a Retrosheet-based sample of 857 players with at least 3000 plate appearances between 1957 and 2007. The difference between clutch situations (defined according to the top 10 percent as defined by the Mills brothers' method) and non-clutch situations in consecutive 250+ PA seasons correlated something in the order of a nonexistent .05.

Cserepy, Nico, Robbie Ostrow, and Ben Weems (2015). Predicting the final score of major league baseball games. *CS229 Final Project*, Stanford University. https://cs229.stanford.edu/proj2015/113_report.pdf

Cui, Andrew Y. (2020). Forecasting outcomes of major league baseball games using machine learning. EAS 499 Senior Capstone Thesis, University of Pennsylvania. https://fisher.wharton.upenn.edu/wp-content/uploads/2020/09/Thesis_Andrew-Cui.pdf

Two models for predicting specific game outcomes using Retrosheet data.