Downey, Jim, and Joseph P. McGarrity (2015). Pick off throws, stolen bases, and southpaws: A comparative static analysis of a mixed strategy game. *Atlantic Economics Journal,* Vol. 43 No. 3, pages 319-335.

Downey and McGarrirty (2015) looked at the issue at hand as a cat and mouse game between baserunners on first thinking about stealing and pitchers trying to keep them from doing so. Their Retrosheet data set consisted of all pitches (and attempted pickoffs) between June 9th and 13th, 2010 with a runner on first base during games in American League parks, i.e. with the DH, purposely chosen to sidestep the complexities involved with pitcher at bats, in the middle of the season, and 5 games to include each member of the standard starting rotation once. Several models imply that there are more pickoff attempts with righty pitchers than lefties, which the authors attribute to the idea that lefties have more success when they do try a pickoff, resulting in baserunners taking shorter leads and attempting fewer steals. There were also more throws to first base with lower OPS batters (allowing pitchers to concentrate more on the baserunner), a catcher less successful at throwing out runners (giving the pitcher a greater incentive to throw over), a closer game score (increasing the baserunner's incentive to steal), better base stealers on first, and fewer balls and more strikes to the batter. As for steal attempts, they increased with better base stealers, higher pitcher ERAs (more baserunners), a closer game score (as before), and right-handed pitching (again, less success keeping batters from stealing).

Downey**,** Jim and Joseph McGarrity (2019). Pressure and the ability to randomize decision-making: The case of the pickoff play in major league baseball. *Atlantic Economic Journal*, Vol. 47 No. 3, pages 261-274.

The authors build on their previous work, in which they described when pickoff attempts were more versus less likely, with a study of the sequence of pickoff throws as an alternative to pitches. The data set (from Retrosheet) was the same as the previous study. In summary, pitchers were pretty good at randomizing their alternation between throws to plate and to first, with the exception of righty pitchers against good base stealers (those in the upper third of a measure of proficiency; stolen bases divided by times on first) in relatively close games (2 runs or less score difference). In this case they tend to alternate between pitches to plate and throws to first in a predictable pattern. In addition, the authors hypothesize that when it is more likely for batters to be successful, there is less an incentive for a baserunner to try to steal and so less reason to throw to first. As a consequence, there were more throws to first with an increased number of strikes and fewer throws to first with a three-ball count compared to fewer.

Fox, Dan aka Dan Agonistes (2004). Triples galore.
http://danagonistes.blogspot.com/search?updated-max=2004-10-18T21:42:00-06:00&max-results=20&reverse-paginate=true

1992 Retrosheet data on the relationship between triples and run scoring.

Fox, Dan aka Dan Agonistes (2004). Defensive indifference.
http://danagonistes.blogspot.com/search?updated-max=2004-10-18T21:42:00-06:00&max-results=20&reverse-paginate=true

Posting data complied by Dave Smith and posted on SABR-L on defensive indifference between 1990 and 2004.

Defensive Indifference by year:
2004 247
2003 219
2002 201
2001 213
2000 199
1999 166
1998 54
1997 122
1996 124
1995 88
1994 82
1993 85
1992 85
1991 78
1990 42

Defensive Indifference by base:
2nd base 1940
3rd base 65

Defensive Indifference by inning:
1st 1
2nd 1
3rd 1
4th 3
5th 12
6th 36
7th 69
8th 212

9th 1498
extra 172

What is interesting here is the increase over time.

Fox, Dan aka Dan Agonistes (2004). Measuring baserunning: Setting a baseline. http://danagonistes.blogspot.com/search?updated-max=2004-11-11T12:29:00-07:00&max-results=20&reverse-paginate=true

Based on play-by-play data for 2003, almost certainly from Retrosheet, Dan Fox posted the probabilities for various baserunner results from singlesand doubles: going one base on a single or two on a double (labeled +1), going an extra base on each (+2), thrown out in attempted advancement (OA), and how often the base directly in front was already occupied. Here it is:

Outs	To	Typ	+1	+2	OA	Next Base Occ
All	All	70.5%	27.2%	0.9%	1.4%	29.2%
0	All	73.4%	25.0%	0.5%	1.2%	21.1%
	7	84.5%	14.1%	0.6%	0.7%	21.6%
	8	68.6%	30.1%	0.3%	1.1%	25.0%
	9	59.7%	38.3%	0.6%	1.4%	16.5%
1	All	72.4%	25.5%	0.7%	1.3%	30.6%
	7	84.7%	13.4%	1.0%	0.9%	31.5%
	8	70.3%	28.6%	0.4%	0.7%	34.0%
	9	58.1%	39.0%	0.9%	2.0%	27.9%
2	All	66.3%	30.7%	1.4%	1.6%	33.9%
	7	81.1%	15.8%	1.4%	1.6%	33.8%
	8	60.0%	35.9%	1.8%	2.3%	33.1%
	9	48.3%	49.7%	1.0%	1.1%	31.1%
All	7	83.4%	14.4%	1.0%	1.1%	29.6%
All	8	65.8%	31.9%	0.9%	1.4%	31.5%
All	9	55.3%	42.3%	0.8%	1.5%	25.8%

7, 8, and 9 stand for balls fielded by the left, center, and right fielder respectively. I do not find anything in the fielder/outs breakdown of note.

Fox, Dan (2005). Tony LaRussa and the search for significance. https://tht.fangraphs.com/tony-larussa-and-the-search-for-significance/
Fox, Dan (2005). A short digression into log5. https://tht.fangraphs.com/a-short-digression-into-log5/

In the first of these, Dan used the Dallas Adams batter/pitcher matchup version of log5 to compute the number of batter/pitcher matchups with outcomes significantly different from chance given their and league average BA. The data was 2003-2005 play-by-play (I'm guessing from Retrosheet) for batters with at least 50 PA and at least 5 matchups; N = 30,481. Of these, only 956 (3.1%) led to more hits than chance would allow for

given batter's overall BA. The test is problematic because five matchups are too small a sample size, but further tests validated the reasonableness of this analysis. Given league average, the actual number of 5 PA matchups with either 4 or 5 hits (150) was close to the chance expectation (144). This work provides evidence that (1) the outcome of batter/pitcher matchups is random over a large sample size and (2) log5 works well in this context.
The second of these webposts provides further evidence that log5 works well using the same data.

Fox, Dan (2006). The irreducible essence of platoon splits.
https://www.baseballprospectus.com/news/article/4970/schrodingers-bat-the-irreducible-essence-of-platoon-splits/

Using Retrosheet data from 1970 through 1992, Dan Fox (2006) discovered the usual batter advantages when facing opposite side hitters, again as usual more extreme for lefty hitters than righty. More interestingly, based on 505 batters with at least 2000 plate appearances during that time, the platoon differentials for batting, on-base, and slugging averages and for walk and strikeout rates were approximately normally distributed, and the correlations between odd and even years for the first three of these were all less than +0.2, although somewhat higher for the last two. These figures imply that individual differences may be random fluctuation such that batters are not consistently more or less susceptible than one another. This in no way disconfirms the existence of the general tendency.

Fox, Dan (2007). Dropping one down.
https://www.baseballprospectus.com/news/article/6446/schrodingers-bat-dropping-one-down/
Fox, Dan (2007). Dropping one down, part 2.
https://www.baseballprospectus.com/news/article/6475/schrodingers-bat-dropping-one-down-part-two/
NOT IN BIBLIOGRAPHY

Dan Fox (2007) supplied a detailed examination of attempted bunt hits between 1970 and 2006, defined as events in which batters bunt and not charged with a sacrifice. Unfortunately this includes sacrifice attempts with a lead runner forced out, and on the other side sacrifice attempts in which the batter beats it out. Anyway, being closer to first it makes sense that lefty batters were more successful (43.8 percent of the time) than righties (37.4%). These attempts occurred most often with no outs (59.9%) and were least likely with two outs (12.2%) with one out intermediate (27.9%). This makes sense, as a runner on first is more valuable the fewer outs there are. And probably for this reason fielders not ready for one with two outs, so success rate (48.8%) was higher than with one (39.8%) or no (39.1%) outs. Attempts were most frequent with bases empty (49.8%), a runner on first (26.7%), or runners on first and second (10.9%). Among these three, success rate was higher with bases empty (45.0%) than first

(32.2%) or first and second (32.3%), as latter two allow for forces on base runners. Attempts occurred much more often on the first pitch of a plate appearance (69.4%, with a success rate 42.2%), with no other count as high as 10%. Success rate was much lower with two strikes (between 13.6% and 9% depending in the number of balls, with 58% of these ending in strikeouts), and over 50 percent for 2-0, 3-0, and 3-1 counts. Dan's follow-up (2007) includes break-even figures based on run expectancies at different base-out situations, which could be discouragingly low (.021 with runners on first and second and no outs) and encouragingly high (.690 with runner on second and two outs).

Fox, Dan (2008). Clearing the decks. https://www.baseballprospectus.com/news/article/7252/schrodingers-bat-clearing-the-decks/

Between 1959 and 2007, although the number of attempted bunt hits declined from age 21 through 37, success rate increased to its peak at age 31 before declining quickly thereafter.

Fritz, Kevin and Bruce Bukiet (2010). Objective method for determining the Most Valuable Player in major league baseball. *International Journal of Performance Analysis in Sport,* Vol. 10, pages 152-169.

Fritz and Bukiet (2010) developed a Markovian method for determining the “best” candidate for MVP awards. The authors applied Retrosheet data to determine actual advancement probabilities, in so doing halving the error in runs prediction from 4 percent in previous work but Bukiet to 2 percent here. They then used a standard lineup (e.g., shortstop leads off, outfield second and third, first base cleanup, etc.) and average offensive performance for a given position (e.g., mean shortstop in the first position etc.) to provide a baseline run distribution, substituted a given MVP candidate’s performance for the average in their position, and compared the two to provide a runs-greater-than-average figure for that candidate. Excluding MVP winners who were pitchers and so irrelevant to the model, the sportswriters’ choice and their “best player” were the same 45 percent of the time. and the winner was among their three “best players” 65 percent of the time, between 1988 and 2007.

Fuld, Elan (n.d.). Clutch and choke hitters in major league baseball: Romantic myth or empirical fact? Unpublished paper.

A well-publicized paper by a University of Pennsylavnia student named Elan Fuld that unpublished but easy to access online (search for “Elan Fuld clutch”) claims that clutch hitters really do exist. Fuld defined the importance of clutch situations according to his computation of their leverage, and then compared through regression analysis the batter’s performance in terms of bases gained per plate appearance (0 to 4) on the plate appearance's specific leverage. If a player did substantially better (worse) in high leverage situations than in low during a given season, then Fuld labeled the player as clutch (choke) in that season. The real issue was whether a player was consistently clutch or choke across their entire career. He used Retrosheet data for 1974 through 1992 for 1075 player with at least two seasons with 100 PAs, including each season reaching that threshold of play (6784 player-seasons in all). He then computed a measure of clutch tendencies across seasons with a threshold defined such that only 1 percent (11 of 1075) of players would be considered clutch and another 1 percent (another 11) choke by chance. When Fuld treated sacrifice flies under the very strange assumption that they are analogous in value to walks, as many as 24 players met the criteria of consistent clutchness across seasons, although never more than 7 reached that for chokeness. As Phil Birnbaum noted (2005c), this assumption inflates the value of a fly ball with a runner on third over fly balls in other situations, as SFs are more likely to occur in clutch situations than the average base/out configuration, while at the same time treating them as walks credits the batter an extra base they did not really earn, artificially inflating their bases gained in clutch situations. When Fuld excluded SFs from the data set, no more than 8 hitters met his criteria for clutchness. Therefore, despite a U. Penn press release claiming that the existence of clutch hitters had been proven along with the media sources that accepted that claim, Fuld's study failed to find the existence of clutch hitters.

Gantner, Ryan (2016). Never make the first or third out at third base…perhaps. *Baseball Research Journal,* Vol. 45 No. 1, pages 17-24.

Ryan Gartner has contributed a computation of breakeven points for advancing while on base. The basis of Gantner's work was an examination of the wisdom of, in his words, "the familiar adage Never make the first or last out at third base" (page 17). Beginning with the relevant baserunner on second and assuming no one else on base (Gantner also looked at an additional runner on first, with similar findings) and using 2013 Baseball Prospectus run expectancy tables, the breakeven points are success rates of 76.4% for no out, 67.1% for one out, and 87.6% for two outs; a replication for 2014 provided almost the same figures. This data appears to corroborate the adage; higher break-evens for no and two outs than for one. However, now including the impact of subsequent possible batters, in 2014 the expected number of runs forfeited by unsuccessful attempts was highest for no outs (.7999), intermediate for one out (.5373), and lowest for two outs (.2901), which stand to reason given the impact of number of outs on scoring. This implied that making the second out is worse than making the third. Further, using Retrosheet play-by-play data, Gantner noted that break-evens are way lower (.651 for no outs, .540 for one out, .806 for two outs) when only one run is needed than for higher numbers of needed runs, implying that when the score is tied in the ninth the runner should more often go for it. Gantner went on to study the impact of baserunning outs at second (overall break-evens about .70 no matter the outs, but about .60 if only one run needed) and home plate (very dependent on number of outs and again lower if only one run needed). He concluded with the following revised adage:

> Never make the last out at third base. Never make the first out at home plate. And never make any out at home plate if more than one run is needed in the inning.

Goldschmied, Nadav, Michael Harris, Damien Vira, and Jason Kowalczyk (2014). Drive theory and home run milestones in baseball: An historical analysis. *Perceptual and Motor Skills: Exercise and Sport,* Vol. 118 No. 1, pages 1-11.

In an attempt to relate drive theory to baseball, these authors examined the 24 players who had reached 505 home runs before the publication date (Albert Pujols got there too late to be included), comparing how many at bats it took for them to hit the last five home runs before their last milestone (either 500, 600, 700, 715 in the case of Henry Aaron and 756 in the case of Barry Bonds) with the first five homers after it. On average, the five leading up took 117.7 at bats and the five afterward 77.5 at bats, consistent with the authors' hypothesis that stress before the milestone restricted performance. Data came from baseball-reference.com and Retrosheet.

Green, Bret and Jeffrey Zwiebel (n.d.). The hot hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from baseball. Downloaded from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2358747

Lichtman, Mitchel (2016). Revisiting the hot hand. In Paul Swydan (Prod.), *The Hardball Times Baseball Annual 2016,* pages 213-227. FanGraphs. Green, Bret and Jeffrey Zwiebel (2018). The hot hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from baseball. *Management Science,* Vol. 64 No. 11, pages 5315-5348.

Another reported demonstration that received a good bit of publicity was an unpublished study by Green and Zwiebel, based on Retrosheet data from 2000 through 2011. In essence using the second, conditional probability method, Green and Zwiebel wanted to see if the outcome of a particular plate appearance for both batters and pitchers could be predicted more accurately using the outcomes of the previous 25 at bats than overall performance for the given season, minus a 50 at bat window around the plate appearance under question. They provided various operational definitions for hot and cold streaks. Some of these definitions seem to bias the study in favor of finding streakiness; these established criteria based on the assumption that the average player is hot five percent and cold five percent of the time, which strikes me as out of bounds given that it presumes streakiness exists. A more defensible definition required the batter to be hot or cold if in the upper or lower five percent of a distribution based on his own performance. Their equations also controlled for handedness and strength of opposing pitchers and ballpark effects, but not, as Mitchel Lichtman (2016) pointed out, for umpire and weather. Unfortunately, ballpark effect was poorly conceived, as it was based solely on raw performance figures and did not control for relative strength of the home team (i.e., a really good/bad hitting home team would lead to the measure indicating a better/worse hitting environment than the ballpark is in truth). The authors' results indicated the existence of hot/cold streaks for all examined measures: hits, walks, home runs, strikeouts, and times on base for both batters and pitchers. Interestingly, after noting improved performance after the plate appearance under question than before, the authors attributed half of the reported increase in that PA to a "learning effect," in essence true improvement in hitting. As Mitchel Lichtman (2016) pointed out, if so, then it should not be considered evidence for the existence of streakiness.

Green and Zwiebel's work elicited a lot of critical comment. Along with the ballpark problem, which Zwiebel acknowledged in email correspondence with Mitchel Lichtman, one comment was that subtracting the 50 at bat window biased the study in favor of finding streaks. Here's an example showing why: let us assume that a player is a .270 hitter. If a player happens to be hitting .300 or .240 during that window, then the rest of the season he must be hitting say .260 or .280 to end up at that .270. In this case, the .300 and .240 are being compared to averages unusually low and high rather than the player's norm. But it strikes me this would only be a problem if hot and cold streaks actually existed – if not, it would be .270 all the way. It is the case that subtracting the 50 at bat window lowers the sample size of comparison at bats, increasing random fluctuation and again adding a bias in favor of finding streakiness. Whether this loss of 50 at bats is catastrophic during a 500 at bat season for a regular player is a matter for debate. In any case, Lichtman (2016) performed his own study using 2000-2014 Retrosheet data, but in this case used the sixth PA after the 25 window, in order to insure it occurred in a different game in most cases. He also used a normal projection method (i.e. three years of past performance with more recent weighted over less) rather than a within-season window. The results were a small hot and slightly larger cold hand effects for BB/PA, OBA, wOBA, and HR/PA, and almost none for BA. Mitchel speculated that changes in both batting (such as swinging for homers after hitting a few) and pitching (such as pitching more carefully to the hit batter and less so to the cold) strategies might be at least partly responsible, along with cold batters playing with an injury.

Green and Zwiebel were finally able to publish their work in 2018, basically unchanged with an additional section in which they claimed to show that the opposition responds to hot streaks by walking the batter in question more often than the batter is normally. They also included a criticism of the Tango, Lichtman and Dolphin analysis of streaky batting described below, based on perceived problems with TMA's use of a batter's average performance as a baseline for identifying streaks. As before, I believe this criticism is flawed by the continued implicit presumption that streaks and slumps exist inherent in Green and Zwiebel's work.

Gross, Alexander and Charles Link (2017). Does option theory hold for major league baseball contracts. *Economic Inquiry,* Vol. 55 No. 1, pages 425-433.

Gross and Link (2017) likely began a new area of study in examining the factors that motivate teams to seek team options for seasons included in free agent contracts. They restricted their sample to 109 circumstances in which position players eligible for free agency signed new contracts between 2003 and 2011, with those contracts either including team options or performance standards that needed to be reached for additional years to vest. Using performance data from Retrosheet, the authors discerned that team options/performance standards were more likely to be included to the extent that player OPS had been variant over the past three seasons, which makes sense as such players were could be thought more likely to perform poorly than more consistent players.

Haechrel, Matt (2014). Matchup probabilities in major league baseball. *Baseball Research Journal,* Vol. 43 No. 2, pages 118-123.

Back in the 1983 *Baseball Analyst*, Bill James presented a formula for the prediction of batting averages in specific batter/pitcher matchups proposed by Dallas Adams which was a spin-off on James's log5 method for predicting two-team matchups. This formula only works for two-event situations; hits versus outs. Matt Haechrel (2014) proposed and mathematically justified a generalization allowing for probability predictions for multiple events (outs, singles, doubles, triples, homeruns, walks, hit by pitches), and using Retrosheet event data showed that the generalization does a good job of predicting the actual proportion of these events for the 2012 season.

Hamrick, Jeff and John Rasp (2015). The connection between race and called strikes and balls. *Journal of Sports Economics,* Vol. 16 No. 7, pages 714-734.

Hamrick and Rasp (2015) took on the issue of racial bias in umpiring, using 1989-2010 data from Retrosheet. They discovered slight increases (.004) in the probability of a strike if the umpire and batter were of different races, which accounts for perhaps a pitch every two games. That increase was greater (.005) with three-ball counts and smaller (.003) with two-ball counts. They also noted some slight differences among races in for umpires (relatively speaking, Hispanic umps favored hitters and Black umps favored pitchers in three-ball counts), pitchers (with three balls, Latin pitchers got more strikes and Black pitchers got fewer; with two strikes, this tendency was reversed), and batters (with two strikes, Black hitters got more strikes and Latin hitters fewer; both were disadvantaged relative to White hitters with three balls). However, there were no significant three-way interactions between the races of umps, pitchers, and batters; in other words, no evidence for discrimination based on similarity of race. They also noted significant but tiny increases in the probability that a pitch would be called a strike if the hitter were on the visiting team or on the team with the worse record, if the pitcher were on the better team, if either the hitter or pitcher was relatively inexperienced, if the score difference was greater, if there were more balls or less strikes, and if QuesTec or PITCHf/x were in use.

Harrison, Willie K. and John L. Salmon (2019). Leveraging pitcher/batter matchups for optimal game strategy (2019). *2019 MIT Sloan Sports Analytics Conference.*

Using 2000 to 2018 Retrosheet data, Harrison and Salmon (2019) uncovered 5170 pitcher/batter matchups with at least 35 PA (they say AB, but they include walks) and used data from that as the basis for simulating 500,000 innings in which they randomized the matchups in order to find the best sequence of pitchers for facing each simulated "lineup" of players. This provided them with 15 clusters of matchup types, with each cluster maximizing certain outcomes and minimizing others. For example, Cluster 12 (the numbers serve only as labels) maximized strikeouts and homers but minimized doubles/triples whereas Cluster 8 maximized flyouts and groundouts. They used those clusters to compare what actually occurred in two innings during the 2018 playoffs with what their simulations would predict were the best matchups from the pitcher's team's point of view.

Healey, Glenn (2015). Moedling the probability of a strikeout for a batter/pitcher matchup. *IEEE Transactions on Knowledge and Data Engineering,* Vol. 27 No. 9, pages 2415-2423.

Healey (2015) proposed models based on Dallas Adams's and Bill James's log 5 method for predicting the general outcome (strikeout versus ground ball) in specific batter/pitcher matchups. Basically, his models establish overall parameters for four categories (lefty and righty pitchers paired with lefty and righty batters), which can then be used for predicting the strikeout and ground ball tendencies for specific batter/pitcher matchups. Healey used Retrosheet plate appearance data for 2003 through 2013, and included every player with at least 150 PAs against both righty and lefty opponents. One interesting overall finding emerged; the closer the ground ball rate of the batter and pitcher in a matchup, the greater the odds of a strikeout. His explanation rings true; ground ball pitchers tend to pitch under bats and ground ball hitters tend to swing over pitches, leading to more strikes. Analogously, fly ball pitchers tend to pitch over bats and fly ball hitters tend to miss under pitches, leading to more strikes.

Healey, Glenn (2017). Matchup models for the probability of a ground ball and a ground ball hit. *Journal of Sports Analytics,* Vol. 3 No. 1, pages 23-35.

Healey's study, based on Retrosheet data from 2003 to 2014, was intended to examine a model for predicting groundball rates and batting averages on ground balls in specific matchups. It included as predictors fairly obvious individual indices; individual pitcher and batter strikeout rate, pitcher groundball rate (although not batter, but instead overall league BA on grounders), batter speed, and pitcher's team's fielding range. Healey claimed that his model allowed for smaller sample sizes than an alternative based log5 for the same accuracy rate. However, extreme cases were poorly predicted. As a byproduct of this work, Healey also gained some insight into the standard platoon advantage/handedness issue. The data revealed that same handed matchups have tended to result in more strikeouts and groundballs than have opposite handed matchups. This leads in turn has led to a lower batting average on grounders but a higher batting average on flies, perhaps due to the tendency for same handed hitters to hit pitches higher in the strike zone than opposite handed. BA on grounders was higher form righthanded hitters than lefties overall, probably due to the preponderance of balls hit to the left side of the infield and thus the longer throw needed to erase the hitter.

Healey, Glenn and Shiyuan Zhao (2020). Learning and applying a function over distributions. *IEEE Access,* Vol. 8, pages 172196-172203.

Using PITCHf/x and Retrosheet data for all 149 pitchers threwing at leasat 1500 pitches in 2016, Healey and Zhao (2010) proposed a method for modeling the odds of strikeouts based on the variation in pitch location and speed.

Hersch, Philip L. and Jodi E. Pelkowski (2014). Does general manager networking affect choice of trade partners in major league baseball? *Journal of Sports Economics,* Vol. 15 No. 6, pages 601-616.

Hersch and Pelkowski (2014), examining data from 1985 through 2011 mostly gathered from Retrosheet, were on the lookout for tendencies for general managers with connections of one type of another to another team to carry out more transactions with that other team than with others. They uncovered a small tendency for general managers who had previously worked together on the same team, and a stronger tendency for two general managers who were related to either one another or to someone else in the other’s organization, to trade more often than the average two-team pairing. General managers who had previously worked for another team were otherwise not more likely to do business with the other team. Other tendencies Hersch and Pelkowski discovered were teams being relatively unlikely to transact with teams in their division but more likely to work with teams in other divisions in their league.

Howard, Jeffrey N. (2018). Hit probability as a function of foul-ball accumulation. *Baseball Research Journal,* Vol. 47 No. 1, pages 60-64.

Based on Retrosheet data 1945-2015, Jeffrey Howard (2018) noted a big difference associated with batters hitting foul balls between when two of them both count as strikes one and two and when they don’t (fouls after strike two, which means swing and misses for strikes). With two strikes on them, batters have hit much better in the former circumstance than in the latter; .335 versus .124 with three non-strike fouls and .413 versus .079 with four non-strike fouls (keep in mind that this means five and six foul balls total respectively for the former situation).

Huckabay, Gary and Nate Silver (2003). Looking for advantages on the ground. https://www.baseballprospectus.com/news/article/1928/6-4-3-looking-for-advantages-on-the-ground/

This is a study of batter/pitcher matchups based on fly ball/ground ball tendencies, based on 1978-2000 Retrosheet data for, batters and pitchers with at least 300 PA for or against. Players have been divided into quartiles, with 1 standing for the quartile most biased toward hitting/giving up fly balls and 4 meaning most likely to hit/give up grounders. The three data lines are for BA/OBA/SLG:

Pitchers	Hitter Quartile 1	Hitter Quartile 2	Hitter Quartile 3	Hitter Quartile 4	Overall Pitchers
Pitcher Quartile 1	0.259	0.267	0.271	0.269	0.266
	0.338	0.335	0.335	0.331	0.335
	0.462	0.442	0.426	0.384	0.430
Pitcher Quartile 2	0.267	0.270	0.272	0.271	0.270
	0.343	0.337	0.335	0.330	0.336
	0.459	0.429	0.415	0.382	0.426
Pitcher Quartile 3	0.272	0.274	0.271	0.275	0.273
	0.346	0.340	0.335	0.333	0.339
	0.454	0.427	0.401	0.378	0.415
Pitche Quartile 4	0.279	0.276	0.273	0.268	0.274
	0.351	0.340	0.336	0.326	0.338
	0.447	0.416	0.391	0.358	0.403
Aggregate: Hitters	0.269	0.272	0.272	0.271	
	0.344	0.338	0.335	0.330	
	0.456	0.429	0.408	0.376	

First, the overall tendencies, which are no surprise. Fly ball hitters had about the same BA but higher OBA (and so more walks) and a greater proportion of extra base hits than ground ball hitters, and fly ball pitchers were responsible for about the same BA and OBA (so the same walks) but a greater proportion of extra base hits than ground ball pitchers. In addition, the most extreme fly ball hitters had higher BA against ground ball pitchers (unlike the other three categories), and analogously the most extreme ground ball pitchers BA went up for fly ball hitters.

Humphreys, Michael A. (2011). *Wizardry.* New York: Oxford University Press.

This in my opinion is the best effort to date to evaluate defensive skill based on conventional data, i.e., not through zone-rating analysis of actual gameplay. There are actually two procedures, both titled Defensive Regression Analysis (DRA), one using Retrosheet data and the other based on conventionally available fielding indices. I will describe procedures non-technically; those interested in the details should consult the book. The goal of the effort was to rid the available data of bias in every practical case, particularly in terms of pitching staff tendencies (i.e., strikeouts versus outs on balls in player, ground ball versus fly ball, lefthanded versus righthanded innings). These tendencies are assumed independent of one another, such that for example lefties and righties on a team are presumed to have the same ground ball/fly ball tendencies. This of course is not true, and, when available, using the Retrosheet data allowed Michael to overcome these problems also. For each position, and starting with a large set of indices, Michael transformed each *relevant index (for example, strikeouts per batters faced, assists per number of balls in play) so as to make each as uncorrelated with one another as possible. The indices for different positions were of course specific to each. For the same reasons I did, and contrary to Bill James's veiled criticisms of my work, Michael only used assists for evaluating most infielders and also catchers, and made what in my probably-biased opinion provided a very persuasive argument for that decision. For analogous reasons, first basemen are only evaluated on their ground ball putouts, although this leaves one with a bias caused by the individual player's tendencies to make the play unassisted versus tossing to covering pitchers. Outfielders are of course rated by putouts.

After that, Michael associated these transformed indices with runs-allowed data, allowing the determination of the average number of runs for each event. These numbers corresponded well with past efforts (e.g., walks worth .34 runs, home runs 2.44 runs), adding a degree of credence to the calculations. Humphrey had to make some potentially controversial decisions along the way; for example, crediting responsibility for infield popups to the pitcher under the assumption that the batter was overpowered, despite his general acceptance of the DIPS principle that the result of batted balls in play are not due to the pitcher. Michael's resulting ratings correlate at about .7 with two zone-rating-type measures, Mitchell Lichtman's Ultimate Zone Rating and Tom Tippett's, and leads to analogous findings. The best fielders save about 20 runs a year, whereas the worse cost 20 runs, when compared to the average.

Hyman, Barry (2021). Overall Offensive Performance (OOP). *Baseball Research Journal,* Vol. 50 No. 2, pages 130-139.

Barry Hyman (2021) proposed what he called <u>Overall Offensive Preformance</u> (OOP), in which players receive credit for the bases gained by their own production, including getting on base due to errors, bases gained by base runners not due to "extra effort" - one base on singles, two bases on doubles, etc. - and bases gained by "extra effort" when baserunners, such as extra bases on hits, steals, and the like. They are charged for outs made, both at bat and on the basepath. This metric clearly is biased toward batters who get to the plate with a lot of baserunners aboard and away from those usually batting with bases empty. Using Retrosheet as his data source, Barry concluded that the average batter's OOP would be in the range of 1 to 1.3.

James, Bill (2006). Relative range factors. In John Dewan, *The Fielding Bible* (pages 199-209). Skokie, IL: Acta Sports.

Bill proposed a new version of range factor in order to correct for various biases in the original measure. One of these biases was the use of games played as a denominator, because it short-changed fielders who did not play full games with some regularity. Bill used Retrosheet data to compute the actual number of innings these fielders played.

James, Bill (2008). *The Bill James Gold Mine 2008.* Skokie, IL: Acta Press.

Batting performance tends to tail off between the middle and end of seasons, but Bill James (2008, pages 310-311) uncovered evidence that player size interacts with this general tendency. Among the 1000 position players with the most plate appearances between 1957 and 2006, Bill compared the size, as measured by an undescribed combination of height and weight, the fifty largest lost 32 OPS points between June and September (from .834 to .802) whereas the fifty smallest lost 11 (from .699 to .688). Although Bill does not say so, I suspect he used Retrosheet data here.

James, Bill (2010). *The Bill James Gold Mine 2010*. Skokie, IL: Acta Sports.

At least during the 2000-2009 decade, it was not true that teams made an effort to match up their number one starters against one another; if anything, it was the opposite. Using his Season Score metric, which works well enough for this sort of analysis, here are Season Score categories for starters and their average opposition:

Pitcher Season Score	Number of Pitchers	Number of Starts	Opposition Pitcher Season Score	Pitcher Season Score	Number of Pitchers	Number of Starts	Opposition Pitcher Season Score
>299	11	366	68.88	50-99	451	10151	77.89
200-299	136	4093	77.67	0-49	980	11614	79.40
150-199	152	4660	80.13	<0	963	8711	81.63
100-149	316	8987	78.01				

The lowest (highest) average opposition starter Season Score was for the starters with the highest (lowest) Season Score. As Bill mentioned, it looks like there was a slight tendency for teams facing the absolute best starting pitchers to sacrifice the game and start their weakest.

Jane, Wen-Jhan (in press). Choking or excelling under pressure: Evidence of the causal effect of audience size on performance. *Bulletin of Economic Research.*

Using 2015 to 2018 Retrosheet performance data and attendance figures from mlb.com, along with various control variables**,** Wen-Jhan Jane (in press) examined the influence of the latter on the former. Overall, using a metric that I believe is hits divided by plate appearances, the average performance for both home and away teams were an inverted U function across five attendance categories (less than 10K, 10K to 20K, 20K to 30K, 30K to 40 K, and more than 40K). Home team players peaked in the 30K to 40K range whereas away team players did so between 20K and 30K. Although present in every inning, the effect for the away team effect players was stronger yet in the 9th and later innings, with the peak now between 10K and 20K. However, there was evidence that "star" players, defined as those who had been All-Stars the previous season, actually improved as attendance rose. Jane's study also revealed more support for home field advantage by means of higher figures on the H/PA metric.

Jarvis, John F. (1999). An analysis of the intentional base on balls. Presented at the 1999 SABR convention and retrieved from http://knology.net/johnfjarvis/IBBanalysis.html

Jarvis, John F. (2002). Career summaries and projections. Presented at the 2002 SABR convention and retrieved from http://knology.net/johnfjarvis/cftn.html

John Jarvis (1999), using the data then available from Retrosheet (1980 through 1996 with the exception of 1991), performed simulations that actually found support for the defensive use of the intentional walk, suggesting that it decreased the number of one- and two-run innings and, although it increased the number of innings with three or more runs, the former impact outweighed the latter. However, by 2002 Jarvis was changing his tune, calculating with 1969 and 1972 to 2002 data that intentional walks only helped the defense when the batter's slugging percentage was greater than .600, which occurred in only four percent of the at bats over those years.

Jarvis, John F. (2000). Mark McGwire's 162 bases on balls: More than one record in 1998. *Baseball Research Journal,* No. 29, pages 107-112.

Adding a wrinkle to research regarding the value of the intentional walk as a strategic tool, we have the unofficial "intentional" walk, when an opposing team does not signal the IBB but the pitcher does not throw anywhere near the center of the plate. John Jarvis (2000) wanted to figure out the circumstances that most often distinguish official IBBs from other walks, so that we can at least speculate the situations when walks not classified as intentional to all extents and purposes are. Based on neural net training and a regression analysis for validation, and again using Retrosheet data, John determined that a walk is most likely intentional if, in order of importance, there is a runner on second, there is a runner on third, there is not a runner on first, the relative score between opposing and batting teams, the inning is later, and there are more outs in the inning (relative score was behind inning and outs in the regression). The slugging average of the batter and (negatively) the next batter also had impact but, surprisingly, far less than the previous list. I would speculate that this is because IBBs often happen at the bottom of the lineup and not only when the best opposing hitter is at the plate.

At some point, John Jarvis did an unpublished study using 17 different seasons for which there was then available Project Scoresheet or Retrosheet data and demonstrated that attempted steals result in worse performance by batters. He also learned that over an entire league the stolen base led to an average of only 2.7 wins per season (with a range of 7 to -2.5).

Jordan, Douglas and David Macias (2019). Team batting average: A comprehensive analysis. *Baseball Research Journal*, Vol 48 No. 1, pages 64-69.

Based on Retrosheet data from 2017, team batting averages pretty much stabilized at by about game 70 and remained the same until the end of the season. They also began to have predictive value in terms of teams better or worse than league average by game 16. Overall between 2003 and 2017, team BA tends to increase until that 70th or so game, I imagine that this is a consequence of warmer and more humid weather encouraging batted ball flight.

Judge, Jonathan and Sean O'Rourke (2020). Measuring defensive accuracy in baseball. https://www.baseballprospectus.com/news/article/58243/measuring-defensive-accuracy-in-baseball/

Jonathan Judge and Sean O'Rourke (2020) used Retrosheet data to compare 2019 fielding performance with evaluations for the then-current version of FRAA (as always, details unknown) with the following set of "competitors": Sports Info Solutions' then current version of Defensive Runs Saved, Mitchel Lichtman's Ultimate Zone Rating, Chris Dial's Runs Effectively Defended, and MLB's Outs Above Average. Ignoring the details, FRAA was the most accurate for outfielders and the least accurate for infielders, OAA was the opposite, RED and DRS did okay across the board, and UZR performed relatively poorly. They speculated that fielder positioning and movement might be significant for infield defense but not for outfielders. If so, then OAA's reliance on it, as described in Tom Tango's essay, could be crucial for infielder evaluation but only add random error for outfielders. It is however important to note that to even the playing field they purposely added no controls for batter, pitcher, ballpark or overall team defense. While defensible in this case, they would need to do so if comparing FRAA to its actual closest "competitor," Michael Humphreys' Defensive Regression Analysis. For what it's worth, I challenge them to do so.

Judge, Jonathan, Harry Pavlidis, and Dan Brooks (2015). Moving beyond WOWY: A mixed approach to measuring catcher framing. https://www.baseball prospectus.com/news/articles/25514/moving-beyond-wowy-a-mixed-approach-to-measuring-pitch-framing

As part of a project designed to measure catcher framing, Judge et al., used 1988 to 2007 Retrosheet ball-strike data to estimate catcher framing abilities, resulting in a model that correlated at .7 with a model based on PITCHf/x data when applied to subsequent seasons. According to their method, the best framers saved about 20 runs in a season over average, comparable to what PITCHf/x data implies. In addition, the researchers calculated the the proportion of taken pitches that were called strikes during that period and on to 2014. The figure was around 29 percent at the beginning, eased up to about 30 percent in 2000, and then jumped to 31.5 percent the next year, perhaps as a product of umpires first answering to MLB as a whole rather than the leagues separately. At about 32 percent in 2008, it went up almost full percentage point in two years when PITCHf/x replaced Questec, and had gotten over 33 percent by 2014.

Kalist, David E. and Stephen J. Spurr (2006). Baseball errors. *Journal of Quantitative Analysis in Sports,* Vol. 2 Issue 4, Article 3.

Using Retrosheet data from 1969 through 2005, Kalist and Spurr discovered that errors tend to be higher for first-year expansion teams, in April than in later months, in day games rather than night (more variable lighting conditions?), in grass rather than artificial turf (again, more variation?), and against faster opposition, as measured by steals per game. Finally, there was a consistent bias in favor of the home team, but it decreased substantially over the period, possibly due to the replacement of active sportswriters with others with perhaps less incentive to ingratiate themselves with home-team players.

Kim, Jerry W. and Brayden G. King (2014). Seeing stars: Matthew effects and status bias in major league baseball umpiring. *Management Science*, Vol. 60 No. 11, pages 2619-2644.

This is probably the best analysis of umpire bias to date. The basic argument is that umpires are predisposed toward favoring “high-status” pitchers; more likely calling “real” balls as strikes (“overrecognition” in the authors’ terminology) and less likely “real” strikes as balls (“underrecognition”) the higher the pitcher’s status, with the bias accentuated for pitchers known to have good control. To examine the argument’s validity, all 2008 and 2009 pitches without batter swings were categorized via f/x pitch data, with a long list of control measures gathered from various sources including Retrosheet. Status was based on number of All-Star appearances, which strikes me as a good index; pitcher control via walks per plate appearance. The results were as follows: In total, overrecognition occurred on 18.8% of real balls and underrecognition on 12.9% of real strikes. Both over- and underrecognition were more likely for the home team, counts favoring the batter, later innings, high leverage plate appearances, more experienced pitchers, and as hypothesized pitchers with more All-Star appearances and better control. The status effects were still apparent for pitches by high and low status pitchers matched for pitch location and type, specific umpire, and count; All-Stars received a relative 6.7% reward in overrecognition and 5.7% bonus in underrecognition. Overrecognition also occurred for lefty batters and games with higher attendance. In my view, the authors' argument seems to generalize to more experienced pitchers, who would have status for that reason alone. In addition, the results for attendance and home team are consistent with the most strongly supported explanation for the home-field advantage; crowd noise.

In addition, analogous biases were uncovered in favor of batters with high status (again All-Star appearance) and demonstrated batting eyes (walks per plate appearance). Variance depending on catcher revealed different skill levels in pitch framing ability, which was not associated with All-Star catcher appearances; skill in pitch framing does appear less appreciated than it deserves. Finally, overcoming a problem in past umpire bias research, an on-line unpublished version of the paper included individual differences among umps in both over- and underrecognition. The authors concluded that 80% of umps are guilty of the former and 64% of the latter. Interestingly, the two biases were largely independent, correlating at only -.16.

Koch, Brandon Lee D. and Anna K. Panorska (2013). The impact of temperature on major league baseball. *Weather, Climate, and Society,* Vol. 5, pages 359-366.

Retrosheet data from 2000 through 2011 combined with data from the National Climate Data Center revealed that most offensive measures (runs scored, home runs, batting, on-base, and slugging averages) increased as game weather got hotter, with the exception of walks. Koch and Panorska also noted the impact of heat on hit batsmen; see Larrick below.

Larrick, Richard P., Thomas A. Timmerman, Andrew M. Carton, and Jason Abrevaya (2011). Temper, temperature, and temptation: Heat-related retaliation in baseball. *Psychlogical Science,* Vol. 22 No. 4, pages 423-428.

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

It has become clear that as the weather gets warmer, the number of hit batsmen goes up, and this has been explained as a consequence of discomfort resulting in increased aggressiveness. Larrick, Timmerman, Carton and Abrevaya (2011), using all games with Retrosheet data from 1952 through 2009 which included game temperature and controlling for pitcher control, discerned that the odds of a hit batsman increased as an interactive function of temperature and the number of teammates hit by the opposing team, such that more hit teammates resulted in more plunking of the opposing team, with this effect accentuated by hotter weather. Krenzer and Splan, using 2000-2015 Retrosheet data, noted both temperature and, more importantly, pitcher wildness as predictors HBPs. Further addressing the question, after dividing the season into fifths based on games played, they observed this correlation only occurring during the middle three-fifths, in other words the warmer months, implying a probable threshold temperature effect before aggression steps in. In addition, HBPs were greater against division rivals than otherwise (where the best rivalries lie), in blow-out games rather than one-runners (unfortunately they did not analyze winner versus losers separately; is this frustrating only for the blown-out team?), and for some reason against visiting teams with better records (why, and why not home teams also?).

Lei, Xinrong and Brad R. Humphreys (2013). Game Importance as a dimension of uncertainty of outcome. *Journal of Quantitative Analysis in Sports,* Vol. 9 No. 1, pages 25-36.

Of the several reasons proposed for the home field advantage in baseball, which is consistently measured at 53 or 54 percent, the most strongly backed by research is the presence of fan support, as home field advantage increases with rising attendance. Indirect corroboration comes from work by Lei and Humphreys (2013). They proposed a measure of game importance (GI), based on either how far a team leading a divisional or wild-card race is ahead of the second place team or how far a team not leading is behind the team that is. Smaller differences imply higher GI scores. Unfortunately, as the authors note, their measure it not weighted by how far in the season a game occurs, so that GI will be the same for a team one game ahead or behind after the 1st as the 161st game. Anyway, in Retrosheet data from 1994 through 2010, GI was positively related with both attendance and home team winning percentage, with the latter implying that home field advantage rises as games become more important. The authors did not know to relate all three, but we can conjecture that game importance raises attendance which increases home field advantage in turn.

Levitt, Dan (1999). Hits and baserunner advancement. *By the Numbers,* Vol. 9 No. 3, pages 20-21.

Dan Levitt (1999) has provided us with estimates of the odds of baserunner advancement on hits based on four years of Retrosheet data (1980-1983). The following is what I believe to be the most interesting of Levitt's findings. The three right-most columns display hit locations when known.

Occurrence	Result	Sample Size	Total	Left Field	Center Field	Right Field
Single with runner on first	Runner to third	31132	31.3%	19.1%	34.6%	49.4%
Single with runner on second	Runner scores	18399	65.3%	68.4%	82.6%	71.7%

Double with runner on first	Runner scores	6997	53.6%	40.5%	58.6%	37.7%

Most of the results can be explained through considering the throwing distance from the outfielder to the relevant base. As home plate is generally farther from the outfield than third base, runners successfully take extra bases to score more often than to get to third. Baserunner advancement for first-to-third after a single is more likely as we move from left field to right. Runners are more likely to score from first on doubles or second on singles to center field than to the corners. It is interesting to note that scoring from first on doubles is both less likely and less influenced by hit location than scoring from second on singles.

Levitt, Dan (2000). Speed scores and reaching base on errors. Retrieved from http://www.baseballthinkfactory.org/btf/scholars/levitt/articles/speedscores.htm

Levitt (2000), this time using individual-level data from 1980 Retrosheet files, found Speed Scores to correlate only .14 with percentage of times reaching base on error per opportunity to do so. Further, this relationship appeared to be an artifact of the number of ground balls hit, given that faster runners are more likely to hit ground balls (Speed Score and percentage of batted balls that are grounders were correlated .3), such that Speed Scores only correlated .04 with times reached base on errors as a percentage of non-basehit ground balls. In other words, faster runners do not reach base on errors more often because they are fast, but rather because they hit more grounders, which lead to more errors than fly balls.

Lyle, Arlo (2007). Baseball prediction using ensemble learning.
https://arti.franklin.uga.edu/sites/default/files/inline-files/lyle_arlo.pdf

The most trustworthy attempt to compare the accuracy of offensive projection models that I have been able to find is a M.A. thesis by Lyle (2007). The author, applying Retrosheet data between 1973 and 2006, used the previous 162-game performance of batters to predict the next 162 game outcomes for six metrics. For four of the six (runs scored, doubles, homers, and RBI), PECOTA slightly outperformed his own method and significantly defeated ZiPS and MARCEL. Lyle's did the best with triples, with ZiPS second, and with hits, which PECOTA did not project.

Mains, Rob (2020). Some bunts are OK.
https://www.baseballprospectus.com/news/article/61876/veteran-presence-some-bunts-are-ok/

This was probably but not definitely from Retrosheet data: Only twelve players attempted 100 or more what were definitely attempts at bunt hits (bunts with bases empty) between 2003 and most of 2020 (this entry was dated September 15). Even with these players, the attempt was relatively rare, with the leader in percentage of plate appearances at only 8.3 (Willy Taveras). Only twelve (with ten overlapping the two lists) had forty or more successful attempts; but among these twelve, the success rate aka batting average on bunts for hits was exactly .400, ranging from Taveras (.476) to Dave Roberts (.328). This means that there are some players who have been quite good at it.

Mains, Rob (2022). Why they're going to keep swinging for the fences.
https://www.baseballprospectus.com/news/article/72363/veteran-presence-why-theyre-going-to-keep-swinging-for-the-fences/

As part of an ongoing project relating home runs with team winning average, Rob Mains (2022) ascertained from Retrosheet game logs that since 1969 home teams have consistently had winning averages of over .700 in those games in which they outhomered the away team, with (reading off charts) that figure at around .750 in the 1980s but up to about .800 in the 2010s. Away teams with more roundtrippers than home teams have had analogous success, averaging perhaps .680 in the 1980s and .730 in the 2010s.

Marchi, Max (2010). Two dimensions of catching – and dealing with interactions.
https://tht.fangraphs.com/two-dimensions-of-catching/

Using Retrosheet data, Max Marchi (2010) devised an index for pitch blocking by dividing the sum of wild pitches and passed balls by the number of plate appearances with runners on base for each catcher/pitcher dyad, combining all of the data for (I assume) a league-year, and then using multilevel analysis to distinguish the impact of individual pitchers and catchers. Finally, he assigned a run value based on .3 runs per unblocked pitch. As would be expected, Hoyt Wilhelm and Charlie Hough ranked as the most responsible pitchers and Greg Maddux the least; the sage himself (Yogi Berra) as the best pitch blocking catcher. Max did the same with base stealing, with the third available factor (baserunner) added to the mix.

Marchi, Max (2012). The art of handling the pitching staff.
https://www.baseballprospectus.com/news/article/16096/the-stats-go-marching-in-the-art-of-handling-the-pitching-staff/
Marchi, Max (2012a). The hidden helpers of the pitching staff.
https://www.baseballprospectus.com/news/article/16199/the-stats-go-marching-in-the-hidden-helpers-of-the-pitching-staff/

Max Marchi (2012) used his multi-level analytic technique and Retrosheet data to, after removing the influence of batter, pitcher, and ballpark, estimate the amount that catchers impact on the outcomes of plate appearances, in effect devising an overall catcher evaluation system. To keep things simple, Max applied the average run value of different types of batted balls. Between 2008 and 2011, the amazing Jose Molina led the way with an estimated 103 runs saved despite being involved in about half the PAs of the closest competitors. Jason Kendall came in last at minus 80 runs. Dividing the data into even versus odd-numbered days allowed a guesstimate of reliability, with a decent correlation of 0.51.

Following up with data going back to 1948 (2012a), Tony Pena was the winner at 248 with the falsely-maligned Mike Piazza tied for third at 205; Molina was easily out front on a rate basis with 38 saved per 5000 PAs. Max also estimated that rookie catchers cost their teams about four runs, and catchers new to a team three runs, per 5000 PA. Each year spent with a team increased these figures by an average of 0.70 runs per 5000 PA. He was unable to locate any noticeable aging effects. Then, adding managers to the mix yielded analogous evaluations. Bobby Cox easily the best at 82 runs, but given his long tenure this only works out to “a couple of runs” per 5000 PA. The Cox effect may be largely due to his pairing with Leo Mazzone. The two together saved 3 runs per 5000 PA whereas Cox with other pitching coaches only coaxed 0.2 runs extra per 5000 PA. Overall, former MLB pitchers who became managers saved 0.50 and former MLB catchers 0.37 runs per 5000 PA, whereas other positions and managers never playing in the majors either saved or lost 0.11 at most.

Marchi, Max (2013). Catcher framing before PITCHf/x.
https://www.baseballprospectus.com/news/article/20596/the-stats-go-marching-in-catcher-framing-before-pitchfx/

Two years before this work, Max Marchi (linked to in the present article) had developed what was then a state-of-the-art multilevel model to estimate the impact of pitchers, catchers, batters, and umpires on ump calls for borderline pitches. In this piece, Max used used 1988-2012 Retrosheet data to estimate an analogous model for pitches in the data set that were not swung at. This model when used on 2008 to 2012 data, for which there is PITCHf/x data, correlated at .72 with his earlier model, implying that it is probably of value for getting approximate figures for earlier catchers. However, it had a far smaller standard deviation, about 7.5 versus 13 for the PITCHf/x model. The latter means that less extreme, more conservative figures are produced, which is probably good given the very provisional status of specific catcher ratings.

Marchi, Max (2013). Who's ahead of whom?
https://www.baseballprospectus.com/news/article/19716/the-stats-go-marching-in-whos-ahead-of-whom/

A different kind of matchup question – at the beginning of the season, are hitters or pitchers ahead of the other? To answer it using 1991-2012 Retrosheet data, Max Marchi (2013) first partialled out the impact of temperature on run scoring, which increases by about 0.2 runs per 10 degree difference, so as to equalize its impact across the season. After doing that, Max calculated that run scoring went down about 0.60 runs between the first and sixtieth games of the season, implying that offense were indeed ahead of defense. Just to make sure it was pitchers who were behind, Max examined Defensive Efficiency Record (the percentage of balls in play on which a team successfully makes a play; see the Fielding Evaluation chapter) and noted no large difference among months.

Maynard, M. Travis, Christian J. Resick, Quinn W. Cunningham, and Marco S. Di Renzo (2017). Ch-ch-ch changes: How action phase functional leadership, team human capital, and interim vs. permanent leader status impact post-transition team performance. *Journal of Business and Psychology,* Vol. 32, pages 575-593.

Maynard, Resick, Cunningham, and Di Renzo (2017) examined 129 in-season managerial changes between 1974 and 2008, and noted that team performance improved after the change; which of course just means that mid-season managerial changes usually occur when a team is going through a particularly bad stretch, and the new manager benefits from regression to the mean. The authors seemed to realize this to an extent, noting that the relevant teams were bad to begin with and continued to display losing records after the change. The authors also noted that player performance improvement was (of course) responsible for the improvement, and particularly when the newly-installed managers made more pitching changes during the games. These impacts were a bit stronger when the new manager was designated as permanent rather than interim. Retrosheet data was apparently used in compiling team winning percentage before and after the managerial change.

McCotter, Trent (2008). Hitting streaks don't obey your rules. *Baseball Research Journal,* Vol. 37, pages 62-70.
Pavitt, Charlie (2009). Hitting streaks and psychology. *Baseball Research Journal,* Vol. 38 No. 1, pages 6-7.

McCotter, Trent (2009). Reply. *Baseball Research Journal,* Vol. 38 No. 1, pages 7-8.
Albert, Jim (2008). Long streaks. *Baseball by the Numbers,* Vol. 18 No. 4, pages 9-13.
Albert, Jim (2010). Great streaks. *Baseball Research Journal,* Vol. 39 No. 2, pages 58-62 and 64.
McCotter, Trent (2009). Reply. *Baseball Research Journal,* Vol. 38 No. 1, pages 7-8.
McCotter, Trent (2010). Hitting streaks don't obey your rules. *Chance,* Vol. 23 No. 4, pages 52-57.

Some work by Trent McCotter has continued the debate concerning the reality of hitting streaks. McCotter's method was as follows: Using Retrosheet data from 1957 through 2006, he recorded the number and length of all batting streaks starting with one game along with the total number of games with and without hits in them. He then compared the number of streaks of different lengths to what occurred in ten thousand random simulated permutations of the games with/without hits in them. There was a consistent and highly statistically significant pattern across all lengths starting at five for more real-life streaks than in the simulations. Trent concluded that hitting streaks are not random occurrences.

Although nobody challenged Trent's analysis as such, there has been some criticism of other aspects of his work. His first attempts at explaining these patterns (batters facing long stretches of subpar pitching or playing in a good hitting ballpark, and streaks occurring more often in the warmer months) were proposed, found no evidence for the first, and claimed the second and third to be unlikely, but never empirically evaluated (although all could be). He instead opted for untestable speculations concerning a change in batter strategy toward single hitting and just the existence of a hot hand. I called him on these, and he responded with helpful analyses inconsistent with the second and third of the testable explanations and basically punted on the untestable ones. Jim Albert (2008) lauded the method and replicated it, but this time restricting the sample to five seasons of Retrosheet data studied separately (2004 through 2008). Again, real streaks occurred more often than in the random permutations, but only three out of twenty comparisons (for 5 or more, 10 or more, 15 or more, and 20 or more, for each of the five seasons) were significant at .05 and a fourth at .10, leading Jim to question the practical significance of Trent's results. This initiated a debate in the *Baseball Research Journal* Volume 39 Number 2*,* in which Jim questioned the practical significance of Trent's findings giving the huge sample size Trent used, Trent defended the huge sample size as necessary to tease out streaks buried in noisy data, and Jim challenged and Trent upheld Trent's use of the normal distribution as the basis for comparison. A later paper (McCotter, 2010) added nothing substantive to the debate.

Mejdal, Sig (2000). The recipe for a stolen base. *By the Numbers*, Vol. 10 No. 3, pages 20-22.
Loughlin, Thomas M. and Jason L. Bargen (2008). Assessing pitcher and catcher influences on base stealing in Major League Baseball. *Journal of Sports Sciences,* Vol. 26 No. 1, pages 15-20.

Given the steal attempt, what are the factors that determine its odds of success? Sig Mejdal (2000) made a nice attempt at answering this question. Mejdal began with the reasonable premise that the possibilities include the baserunner's speed, catcher's throwing ability, speed of pitcher's delivery, umpire play-judgment tendencies, and the stadium surface (turf is easier to run on than grass). One confound is between catcher and pitcher, as a particularly good or poor throwing catcher would make it appear that the pitchers he works with are better or worse than average, whereas a staff populated by pitchers particularly quick or slow at delivering the ball to the plate would make it seem that their catcher is better or worse than average. Thus it looks as if the probability of successful stolen bases against particular catchers and the probability against certain pitchers are seriously dependent on one another. However, using three years of Retrosheet data, Mejdal found that an attempt to correct the catcher's successful steal percentage by adjusting it by the average percentage of pitchers teamed up did not lead to significantly different numbers than merely computing the catcher's percentage across those years, so he used the simpler measure. Mejdal then corrected the pitcher's percentage by computing the percentage for all the catchers they have worked with, comparing the two percentages, and then using the difference between the two to represent the pitcher. To use his example, if pitcher Joe Schmo was paired up with catchers that averaged a 60 percent steal rate and his own steal rate was 40 percent, then Mejdal credited Joe with a 20 percent "stolen base value." Mejdal's method, in essence, given precedence to the catcher by presuming that his successful steal percentage, when taken over a long enough time frame, is a valid measure of ability, and that pitcher's percentage should be determined within their catchers' context.

Mejdal then entered measures for the relevant factors into a multiple regression equation predicting successful steal rate. Unfortunately, he failed to provide data on the overall predictive power of the five factors. Of that variance in successful steal percentage that was accounted for by the equation, 36 percent was attributed to the baserunner, 34 percent to the pitcher, 19 percent to the catcher, 11 percent to the surface, and absolutely none to the umpire. It is particularly interesting that the pitcher was found to be almost twice as influential as the catcher, as the correction described above in a sense gave the catcher a greater "opportunity" to influence the results.

Using Retrosheet data from 1978 through 1990, Loughlin and Bargen (2008) demonstrated that differences in catchers' ability to control the "running game," as measured by success steals divided by attempts, and of pitchers' ability to hold runners, as measured by attempted steals divided by opportunities, are statistically significant; which they claim nobody had done previously. .The variation among pitchers was greater than that for catchers, which is consistent with Mejdal's division of responsibility just mentioned.

Menéndez, Héctor D., Miguel Vázquez and David Camacho (2015). Mixed clustering methods to forecast baseball trends. In In David Camacho, Lars Braubach,

Salvatore Venticinque and Costin Badica (Eds.), *Intelligent Distributed Computing VIII* (pages 175-184). Heidelberg, Germany: Springer.

Soto Valero, C. (2016). Predicting win-loss outcomes in MLB regular season games – A comparative study using data mining methods. *International Journal of Computer Science in Sport,* Vol. 15 No. 2, pages 91-112.

Menéndez, Vázquez and Camacho (2015) and Soto Valero (2016) used Retrosheet data in methodological studies attempting to predict the outcome of games; neither have substantive import.

Mills, Brian M. (2017). Policy changes in major league baseball: Improved agent behavior and ancillary productivity outcomes. *Economic Inquiry*, Vol. 55 No. 2, pages 1104-1118.

Using PITCHf/x data, Mills (2017) concluded that the average strike zone as called by umps had expanded on the bottom by three inches between 2008 and 2014, resulting in three times as many called strikes in the zone between 18 and 21 inches off the ground. Both pitcher and batters appear to have noticed the change, with the proportion of pitches in that zone increasing from about 22 percent to about 27½ percent, and swings on pitches in that zone from about 31 percent to about 34½ percent. This change favors the pitchers, as when a batter swings at pitches in that zone, the odds of making content are 73%, putting a ball in play 48%, and getting a hit 26% lower than for pitches above it. Using Retrosheet data, Mills noted a relationship between this change and run production per game over this interim.

Morey, Leslie C. and Mark A. Cohen (2015). Bias in the log5 estimation of outcome of batter/pitcher matchups, and an alternative. *Journal of Sports Analytics,* Vol. 1 No. 1, pages 65-76.

Morey and Cohen (2015) argued that applying the log5 method to batter/pitcher matchups may result in biased findings because the method presumes a mean probability of .500, which will occur across teams but not for batting indices. Simulations for the 1996 through 2013 seasons based on data downloaded from Retrosheet and Lahman's database resulted in BA (around .300) and HR (almost 8 per 100 ABs) consistently too high, with the bias more pronounced as true performance becomes more extreme. The first author's alternative method is better, although in this case producing underestimates.

Palmer, Pete (2014). Stolen bases and caught stealing by catchers: Updating Total Player Rating. *Baseball Research Journal,* Vol. 43 No. 1, pages 23-25.

Thanks to historical information that became available thanks to Retrosheet, Pete has been able to add stolen base/caught stealing data to TPR for catchers; incidentally, his list of the top 20 all-time in controlling the running game is consistent with catchers' reputations, with Ivan Rodriguez leading the pack.

Palmer, Pete (2017). Intentional walks revisited. *By the Numbers,* Vol. 27 No. 1, pages 16-25.

Pete Palmer's well known run-value figures, popularized in *The Hidden Game of Baseball*, were, due to absence of sufficiently-detailed, estimated with the presumption that the likelihood of all relevant events is independent of base-out situation. In 2017, Pete used 1946-2015 Retrosheet data to determine the actual run values of the following events:

Single	.453	Home Run	1.413	Unintentional Walk	about .31
Double	.752	Out	-.241	Intentional Walk	.157
Triple	1.038				

The figure for unintentional walks is an estimate, as Pete actually provided a combined value of .298. The reason that intentional walks are so much lower than unintentional is that the former tend to occur in circumstances in which their impact of runs is less, particularly with runners on second, third, or both those bases, occurring in more than two percent of relevant cases (the highest is 2nd and 3rd with one out; more than 12 percent). IBBs are given in fewer than one percent in all other circumstances.

Pete also examined the IBB as a strategic tool. With the exception of when designated hitters are available, the IBB has been most often used for the #8 batter due to the weakness of the upcoming #9. Even so, it usually works in favor of the team at bat. For example, with runners on second and third and two out, it has historically decreased expected runs by .033 for that inning but increased it by .113 for the following inning, given that the #1 position is then likely to lead off in that next inning. Walking a stronger batter to face a weaker one is also usually a loser for the defensive team, as the next batter must be considerably weaker (e.g., at least a bit below average when the batter that is walked is among the upper one-sixth in performance) to be worth the tradeoff. And walking a batter to get the platoon advantage is also not worth it, as the advantage normally does not offset the value of the extra baserunner.

Palmer, Pete (2018). Relief pitching strategy: Past, present, and future? *Baseball Research Journal,* Vol 47 No. 1, pages 45-52.

Pete Palmer (2018) offered a far-reaching discussion of some of the implications that the growth in relief pitcher usage has implied for the game. Beginning with that growth., Pete calculated that the percentage of time in which a team's save leader entering the game with their team ahead but with win probability percentages of less than 50 percent due to the base-out-inning situation, has plummeted from 23 percent during the 1980s to 10 percent during the 1990s to 4.7 percent during the 2000s up to 2017. That is because only about 3 percent of them occur in the ninth inning, which has more and more become the only time the save leader (a better term than closer given previous usage patterns) appears.

In evidence relevant to the myth of the proven closer, since 1961 the difference in save percentage for a team's save leader versus other pitchers has increased, but is not as large as some might think. In the 1960s, the difference was about 4 percent with a one-run lead in the ninth inning; by the 2000-2017 interim it had increased to about 9 percent. Yet, and this is critical, even now the success rate of non-closers with a one-run lead in the 9th was more than 76 percent for visiting teams and more than 83 percent for home teams. These jump to about 89 percent for visitors and 92 percent for home with a two run lead, and over 95 percent with a three run lead, with corresponding decreases in the disadvantage they have to closers in this regard.

The increase in number of pitchers per team is of course linked with the decrease in the number of position players on the 25 man roster. This has restricted the number of substitutions managers can make with the matter. In the 1960s, there were an average of 233 fielding substitutions, 211 pinch-hitters, and 40 pinch-runners per team per season; between 2011 and 2017, these figures had dropped to 197, 183, and 28 respectively. Platooning has also dropped. Defined as a circumstance in which, for a position, a team has at least 70 starts by a lefty hitter and 30 starts by a righty hitters against opposite handed starting pitchers. Using Retrosheet data, Pete noted that platooning was almost non-existent at the beginning of the 20th century, the proportion of platooned positions had increased to about 20 percent from 1958 to 1990, but was down to about 14 percent by 2017.

Palmer, Pete (2021). Career park effects for individual players. *By The Numbers,* Vol. 30 No. 1, pages 9-13.

Pete Palmer (2021), using Retrosheet data, computed individual player OPS values for home and away games, divided the former by the latter, and then multiplied by 100, in so doing producing a career park effect figure for each player. These what I will call <u>OPS park ratios</u> I(combining two labels Pete proposed) need to be distinguished from Pete's general park factors as they measure individual player/ballpark fit. Not surprisingly, Rockies players dominate the top ten., with Charlie Blackmon (134) the highest ever at the time of Pete's work. Nonetheless, while dominating at home (1.054 and 1.072), Larry Walker and Todd Helton's respective road figures (.857 and .867) show that they excelled everywhere. Gil McDougald's 80 (OPS of .680 home and .847 road) was the lowest of anyone with 3500 at bats by an astounding eight points, making him the player with the worst ever ballpark fit and demonstrating that his honors (five-time All-Star and five-time recipient of MVP votes) were deserved.

Panas, Lee (2010). *Beyond Batting Average.* self-published, available at lulu.com

This book is a summary of sabermetic research, concentrating on player evaluation measures but short on material relevant to strategy. Panas used data from several sources; from Retrosheet, he computed a run expectancy chart for 2005-2008 (Chapter 5, Linear Weights), some RBI percentage rankings (Chapter 6, Situational Hitting), an example for a measure of baserunning performance (Chapter 7, Baserunning), and figures on BA and SA on batted ball type (Chapter 9, Fielding Independent Pitching).

Phillips, David C. (2011). You're hurting my game: Lineup protection and injuries in major league baseball. *Journal of Quantitative Analysis in Sports,* Vol. 7 Issue 3, Article 7.

Phillips (2011) performed the most thoughtful study of protection to date, with results analogous with other studies. He realized that a study of protection based on player movement within a batting order (e.g., moving a cold hitter to a different spot in the lineup) leads to ambiguous findings, because any change in the performance of that hitter could be due to the change in subsequent batter or to random changes in that player's performance irrelevant to who is batting behind. In response, Phillips looked at differences in performance for a given player remaining in the same lineup position based on changes in the next batter caused by injury. Based on Retrosheet data from 2002 through 2009 and limited to protectors with an OPS of at least .700 for a minimum of 200 plate appearances (in other words, hitters good enough to count as potential protectors), Phillips noted that injuries to protectors resulted in an overall OPS decrease of 28 points at that lineup position due to a weaker replacement. With the weaker replacement, the hitter being protected tended to receive a lot more intentional walks but fewer extra base hits (but no more hits, as additional singles compensating), indicative of the expectation that a non-protected hitter will be pitched around more often. These two tendencies pretty much cancelled one another out, resulting in little overall protection effect.

Phillips, David C. (2017). Stopping on nine: Evidence of heuristic managerial decision-making in major league baseball pitcher substitutions. *Southern Economic Journal,* Vol. 84 No. 2, pages 577-599.

Phillips (2017) examined 1992-2012 Retrosheet data to see if there has been a tendency to remove starting pitchers before their pitch count crosses a number that ends in zero. Although any such tendency was weak in the first decade of the study, there was a two percent increase in the number of times relief pitchers entered when the starter reached a pitch count ending in nine. However, the bias was weaker the closer the game score, implying that managers are less concerned with pitch counts and more with immediate strategy in those games. Finally, the bulk of the bias was for pitchers in their first three seasons, showing that managers were more concerned with in protecting the arms of the relatively young. An additional tidbit was that 80 percent of starter pitch counts were between 69 and 125 in 1992 and 78 and 114 in 2012, evidence that managers were concerned with protecting both starters (decrease in the higher number) and the bullpen (increase in the lower number, meaning fewer innings for relievers) from overwork.

Pinheiro, Ryan, and Stefan Szymanski (in press). All runs are created equal: Labor market efficiency in major league baseball. *Journal of Sports Economics.*

Here is a season-by-season run expectancy matrix for 1996-2015 (apologies that it is not lined up correctly):

Season Walk Single Double Triple Home Run Out
1996 0.331 0.485 0.784 1.105 1.403 −0.302
1997 0.307 0.465 0.761 1.083 1.393 −0.284
1998 0.312 0.469 0.780 1.014 1.400 −0.285
1999 0.311 0.477 0.789 1.059 1.408 −0.302
2000 0.332 0.482 0.765 1.085 1.406 −0.307
2001 0.298 0.460 0.778 1.084 1.380 −0.283
2002 0.303 0.466 0.755 1.052 1.398 −0.279
2003 0.307 0.466 0.775 1.080 1.391 −0.284
2004 0.307 0.462 0.786 1.041 1.396 −0.287
2005 0.295 0.458 0.768 1.056 1.412 −0.277
2006 0.317 0.467 0.766 1.070 1.389 −0.290
2007 0.310 0.468 0.798 1.044 1.406 −0.289
2008 0.312 0.460 0.772 1.081 1.405 −0.281
2009 0.304 0.459 0.762 1.004 1.392 −0.278
2010 0.299 0.451 0.763 1.076 1.404 −0.266
2011 0.289 0.442 0.736 1.064 1.392 −0.255
2012 0.284 0.441 0.747 1.039 1.396 −0.257
2013 0.285 0.439 0.740 1.035 1.371 −0.250
2014 0.283 0.437 0.739 1.054 1.400 −0.245
2015 0.303 0.442 0.743 1.031 1.386 −0.257
mean 0.304 0.459 0.764 1.056 1.396 −0.277
standard deviation 0.014 0.014 0.018 0.027 0.010 0.018

Note the year-to-year stability. The authors then ran regressions showing that these consistently accounted for 93 to 94 percent of variance in each team''s runs scored for those seasons. Finally, they used the run expectancy data to calculate run value for individual non-pitchers with at least 130 AB, and then related the individual values to salaries. Salary were roughly consistent with run values, with a slight improvement in 2005-2016 over 1996-2004 due to better valuing walks following (in their opinion) the publication of *Moneyball*.

Pinto, David (2007). Analyzing the umpires. https://www.baseballprospectus.com/news/article/6533/the-big-picture-analyzing-the-umpires/

Using the Retrosheet record, David Pinto (2007) estimated the winning average of the team with the better record against the team with a worse record in two-team matchups using the following formula:

winning average of better team X (1 minus winning average of worse team)
divided by
(winning average of better team X [1 minus winning average of worse team]) plus
(winning average of worse team X [1 minus winning average of better team])

and then compared these results with the records of individual umpires in such matchups from 2000 to 2006. As one would expect, there was a distribution of umps such that the better team consistently won more often than the formula would predict for some and less often for others. However, there was nothing apparent in the data to suggest that any of this was intentional influence on game outcomes. In addition, relevant to the home field advantage, David noted that the overall estimate for the better team during these seasons was a winning average of .587, analogous to a 95-win season, but the actual home team record for the better team was .623, or 101 wins.

plen (2010). The leadoff walk. https://community.fangraphs.com/the-leadoff-walk/

Somebody calling themselves plen used 1952-2009 data to examine the odds of leadoff hitters scoring when they get on base. In order of raw number of occurrences, these were: singles 37.69 percent, walks 37.9 percent, hit by pitches 38.77 percent, errors 37.74 percent, strike three pitches getting past the catcher 37.24 percent, and catcher's interference 34.84 percent (with the latter occurring only 155 times, leaving the possibility that the lower figure is due to random variation).

Poling, Alan, Marc A. Weeden, Ryan Redner and T. Mary Foster. (2011). Switch hitting in baseball: Apparent rule following, not matching. *Journal of the Experimental Analysis of Behavior*, Vol. 96 No. 2, pages 283-289.

Poling, Weeden, Redner, and Foster (2011), looking at play-by-play data from Retrosheet via Baseball Reference, were apparently experimental psychologists of the behavioristic school. They wrote as if they were surprised to discover that switchhitters Mickey Mantle, Eddie Murray, and Pete Rose's "apparently chose handedness based on the rule 'bat opposite the pitcher,' not on differential consequences obtained in major league games." As this was inconsistent with previous data about the impact of reinforcement of past success/failure seen in basketball shot selection and American football play selection in specific and human behavior in general, they called for more research into the variables that affect behavioral choice. Methinks that they really were not surprised by their findings.

Pope, Devin and Uri Simonsohn (2011). Round numbers as goals: Evidence from baseball, SAT takers, and the lab. *Psychological Science,* Vol. 22 No. 1, pages 71-79.

Hitting .300 is a goal for many hitters, and Pope and Simonsohn (2011) believed that the desire to do so can serve as motivation for hitters very close to that mark with a game or two left in the season to perform particularly well in those last couple of games. Examining Retrosheet data from 1975 through 2008 for all hitters with 200 or more at bats in a season (comprising a sample size of 8817), the authors showed that a higher proportion of players hitting .298 or .299 got a hit on their last plate appearance (.352) than players hitting .300 or .301 (.224). They were also, however, less likely to be replaced by a pinchhitter (.041 versus .197). The latter leads to an obvious bias; that hitters just over the .300 benchmark have less of an opportunity to drop under than hitters just under to move over it. Scott and Birnbaum (2010) demonstrate that a statistical correction for this bias removes this last at bat advantage, and in fact there is "nothing unusual about the performance of players on the cusp of .300" (page 3).

Powers. Scott, Trevor Hastie, and Robert Tibshirani (2018). Nuclear penalized multinomial regression application to predicting at bat outcomes in baseball. *Statistical Modeling,* Vol. 18 Nos. 4-5, pages 388-410.

Powers, Hastie, and Tibshriani (2018) presented a model for predicting the outcomes of specific plate appearances using 2015 Retrosheet data. All batters with at least 390 PA and all pitchers with at least 360 PA against included individually; the data for the rest combined with positions into an abstract "replacement level" player. The predictors were batter and pitcher tendencies, their handedness match or mismatch, the ballpark, and the home-field advantage. The relevant categories were strikeouts, walks, hit by pitches, homers, triples, doubles, singles, groundouts and flyouts. The model was designed to take advantage of the associations existing between these categories, which were computed using principal components factor analysis. For example, above average singles hitters also tended to ground out more than average, analogously for homers and strikeouts, and those flying out a lot tended to not ground out a lot.

In addition, the principal components analysis allowed the authors to present both trilogies of factors for describing batter skills. The first factor included negative loadings for strikeouts, walks, hit by pitches, and homers, and positive loadings for fly and ground outs, singles, doubles, and triples. Most of these loadings were very small, but nonetheless Three True Outcome type hitters were clearly being distinguished from contact hitters. The second factor included positive loadings for fly outs and homers and negative loadings for ground outs and singles. Implying a distinction based on vertical angle of batted ball, The third factor features a very strong negative loading for singles and a very strong positive loading for ground outs; and as these two were positively associated in the previous two factors, this seems to differentiate non-power hitters with high and low batting averages.

Powers et al. did the same for pitchers, with the first two factors reflecting well-established distinctions. The first factor included a strong negative loading for strikeouts and positive loadings for singles, ground outs, and fly outs, clearly distinguishing strikeout from pitch-to--contact pitchers. The second factor featured a strong negative loading for fly outs and positive loadings for ground outs and singles. The third was not as clear cut, as ground outs and fly outs (and also homers) loaded positively and walks (but also singles) negatively, signaling some division between giving up walks versus batted balls.

Rockoff, David M. and Philip A. Yates (2009). Chasing DiMaggio: Streaks in simulated seasons using non-consistent at-bats. *Journal of Quantitative Analysis in Sports,* Vol. 5 Issue 2, Article 4.

Rockoff, David, and Philip Yates (2011). Joe DiMaggio done it again…and again and again and again? *Chance,* Vol. 24 No. 1, pages 14-18.

There have been numerous attempts to estimate the odds of a 56 game hitting streak, and in my opinion Rockoff and Yates (2008) is the best of all these attempts. Their idea was to simulate 1000 seasons of play using actual seasonal game-to-game performance for each of 58 years of Retrosheet data. Out of the 58,000 simulated seasons, a total of 30 (about .005%) included a hitting streak of 56 or more games. Interestingly, Ichiro's 2004 season included 5 of them. Using this data, the authors concluded that the odds of a streak of more than 56 games in any of the 58 seasons in the data set was about 2½ percent. In a follow-up (Rockoff & Yates, 2011), they performed 1000 simulated "baseball histories" under a number of different assumptions: the odds of a hit directly determined by player batting average, including the odds of a hit determined by a varying amount centered around the player batting average, and the odds of a hit partly determined by overall batting average but also by performance in 15 and 30 game stretches around each game under question. The latter two methods assume the existence of hot and cold streaks, which I think is an error. This is because, as will be described later in this chapter, the very existence of such streaks as anything other than the results of random processes is questionable. Part of the point of examining this topic in the first place should be to address whether hitting streaks are or not random, and so to presuppose that they are not leads to an invalid bias in favor of long streaks. As a consequence, the author(s) uncovered 85 56-game or greater streaks using the "batting average" approach, 88 using the "variation around batting average" approach, 561 using the "15 game" approach, and 432 using the "30 game approach." I only consider the first two to be defensible. To make this point more clearly, the simulated Joe DiMaggio equaled or bettered his real streak once using each of the two methods and twice using an "equal at bats" approach, but four and nine times respectively for the latter two methods. Anyway, Rockoff and Yates estimated that allowing streaks to carry over across two seasons would increase the overall number by about ten percent.

Roher, David (2007). Quantifying the impact of opponent quality. *By The Numbers,* Vol. 17 No 2, pages 5-7.

Does good pitching stop good hitting? Earlier work by Dan Levitt and Tom Hanrahan suggests not, but rather implies that good pitching is better than bad pitching at stopping good hitting, and good hitting is better than bad hitting at overcoming good pitching, but nothing more. However, they worked with aggregated data, which could mask relationships which only come to light when variation among player tendencies are considered. Happily, David Roher (2007) took this on. Using Retrosheet data from 2006, David calculated the relative value of each event for run production, measured pitcher quality by Fair Run Average and batter quality through Equivalent Average, and used those to measure the impact of opponent quality on both batter and pitcher performance. The result, which he called Opponent Quality Effect, had a good deal of variation across players – in other words, a big difference among players in how much their performance was affected by opponent quality – but absolutely no relationship with measures of pitching and batting performance. The conclusion is then the same as that from Dan and Tom's work.

Rosciam, Chuck (2004). Professional thieves vs. the constabulary. *Baseball Research Journal,* No. 33, pages 81-83.

Based on a larger data set than analogous efforts (1963 and 1965-1968 A. L. games and 1969 to 2004 games for both leagues), Chuck provided stolen base success rates of 73.1 percent for second base, 71.6 percent for third base, and 37.4 percent for home. He also presented detailed indices for the most prolific base stealers and the catchers most successful at thwarting them for that period of time.

Ruane, Tom (1999). Stolen base strategies revisited. *By The Numbers,* Vol. 9 No. 1, pages 24-28.

Tom Ruane (1999), using raw game data for 1980 to 1989 compiled by Project Scoresheet and Retrosheet, found specifically for runner on first stolen base breakeven points of 70.9 percent success rate with no out, 70.4 percent for one out, and 67.1 percent for two outs. Tom also computed both run potential and probability of scoring both when a steal was and was not attempted from first on the next play, with the following differences:

	Run Potential			**Odds of Scoring**		
Outs	0	1	2	0	1	2
All runners	-.005	-.014	+.031	+.053	+.031	+.043

Fast runners	-.014	-.045	+.030	+.060	+.018	+.047

For example, looking at the first row, attempted steals from first lower run potential 1.4 percent with one out but raise it 3.1 percent with two outs. Trying to stealing second does increase the odds of scoring in all situations. The overall point, however, is how small these differences are. Interestingly enough, the speed of the base stealer has little impact. Using an informal method devised by Bill James (1987) for classifying base runner speed called Speed Scores, Tom Ruane computed the analogous figures only for the fastest runners (second row) and discovered them to be almost the same.

Ruane, Tom (2005). In search of clutch hitting. *Baseball Research Journal,* No. 34, pages 29-36.

In this study, which is also posted on the Retrosheet research page, Tom examined the difference between batting performance with runners in scoring position versus not, using Retrosheet data from 1960 through 2004 for all batters with at least 3000 career at bats during that interim. Based on each player's performance with runners on second and/or third versus not, Tom noted the difference between simulated and actual outcomes and uncovered no systematic differences in the distribution of those differences across all of the players. As a methodological note, Tom thought to take all walks and sacrifice flies out of the data set, because the former is very dependent on base-out situation (much more likely with runners in scoring position but first base unoccupied) and the latter biases batting average with runners in scoring position (i.e., they do not count as at bats). Tom found that batters averaged 7 points higher in batting and 15 in slugging with no runners in scoring position, which is likely more accurate than earlier studies that failed to include these corrections.

Ruane, Tom (2005). Do some batters reach on errors more than others? *Baseball Research Journal,* No. 34, pages 113-120.

Replicating earlier work by Clifford Blau, Bill James, and Mark Pankin using Retrosheet data to analyze batters who made at least 2000 outs between 1960 and 2004, Tom noted that batters that get on base due to errors tend not surprisingly to be faster (causing the fielder to hurry and perhaps get careless), ground ball hitters (grounder result in more errors than flies) and righthanded hitters (more errors on grounders to the left side of the infield, probably due to the longer and more hurried throw). The effects are small, with the lefty/righty difference only at 3/10 or 4/10 of 1 percent and speed effect in the same range. This research is also available at the Retrosheet research page.

Ruane, Tom (n.d.). Strikeouts, grounders and fly balls. Retrieved from http://www.baseballthinkfactory.org/btf/scholars/ruane/articles/goodout.htm

Tom's analysis, based on 1982, 1983, and 1987 Retrosheet data, showed that the expected loss in runs during an inning from strikeouts was greater than that for flyouts and, in particular, groundouts, and that the difference among the three increases as the hitter becomes faster as measured by Bill James's "speed score" metric:

	Strikeouts	Fly outs	Ground outs
Slow runners	-.278	-.261	-.262
Average runners	-.276	-.257	-.244
Fast runners	-.268	-.254	-.230

These data imply that, relative to strikeouts and fly outs, the positive value of moving up baserunners when making outs through hitting the ball on the ground outweigh the negative value of hitting into double plays. These overall numbers mask huge situational differences. I illustrate with two extremes for average speed runners hitting with one out: With only a runner on first, the type of out barely mattered (strikeout, -.305; flyout -.303; groundout, -.327), although here double plays do slightly trump moving the runner up. With runners in second and third, it makes all the difference in the world (strikeout, -.825; fly out, -.438; ground out, -.302), reflecting in particular the possibility of the runner on third scoring with any batted ball and additionally the runner on second moving to third on a groundout.

Saavedra, Serguei, Scott Powers, Trent McCotter, Mason A. Porter, and Peter J. Mucha (2010). Mutually-antagonistic interactions in baseball networks. *Physica A,* Vol. 389, pages 1131-1141.

Saavedra, Powers, McCotter, Porter, and Mucha (2010) concocted a statistically-sophisticated evaluation system based on the run potential for specific batter-pitcher matchups. They presented findings using all Retrosheet data between 1954 and 2008. The results of their model correlated almost perfectly (.96) with an index based on overall run potential.

Seidman, Eric (2008). Ballad of the fatigued: The effects of long innings. https://www.baseballprospectus.com/news/article/7641/ballad-of-the-fatigued-the-effects-of-long-innings/

Seidman, Eric (2008a). Ballad of the fatigued: Controlled results, time, and release points. https://www.baseballprospectus.com/news/article/7702/ballad-of-the-fatigued-controlled-results-time-and-release-points/

Eric Seidman (2008) used PITCHf/x data to examine the pitch velocity and vertical movement consequences for 30 starters who threw at least 40 pitches in the first inning for the 2007 season plus up to May 20th in 2008 (based on a list compiled by Dave Smith which most certainly originated with Retrosheet data). For the rest of that inning, average pitch velocity remained the same although vertical pitch movement decreased starting around pitch 20 and continued to do so for the rest of the inning. What happened during the subsequent second through sixth innings depended on the pitchers' fastball dependency. Those who threw fewer than 27 fastballs in the first did not lose velocity and added some vertical movement; those who threw 27 or more fastballs in the first lost about 1½ mph in the second but no more but lost significant vertical movement. In a follow-up copied-and-pasted table, Eric compared groupings of these pitchers based on average velocity with their performance in analogous starts in which they threw 24 or fewer pitches in the first:

	Slow		Medium		Fast	
IP	40+	Control	40+	Control	40+	Control
1	86.54	86.98	90.35	89.95	92.05	92.12
2	86.27	87.25	88.87	90.14	91.16	92.34
3	86.56	86.77	89.20	89.97	90.81	92.03
4	86.54	87.05	88.80	89.72	90.79	92.27
5	84.99	86.39	89.37	89.97	90.39	92.48
6	84.26	87.32	88.92	89.76	N/A	92.22

As for horizontal and vertical pitch movement:

	Slow		Medium		Fast	
IP	40+	Control	40+	Control	40+	Control
1	8.55/ 9.06	8.33/9.45	5.58/ 9.13	6.69/8.08	6.81/9.19	6.49/9.09
2	8.24/ 9.21	7.79/8.79	5.91/ 8.49	6.62/7.93	5.80/9.21	6.61/8.99
3	9.30/ 9.13	8.31/9.32	7.03/ 7.91	5.97/8.43	6.50/8.81	6.59/8.94
4	7.91/ 9.89	8.14/8.75	5.51/ 9.57	6.56/8.09	7.88/8.59	6.61/8.96
5	8.72/10.71	8.21/8.84	5.53/ 9.11	6.56/8.86	9.17/7.85	6.59/9.14
6	8.99/ 9.14	8.01/9.14	6.08/10.08	6.29/8.27	N/A	6.66/8.90

Seidman, Eric (2009). On the swing.
https://www.baseballprospectus.com/news/article/9841/checking-the-numbers-on-the-swing/

Eric Seidman (2009) examined a total of 897 seasons between 1974 and 2009 from 598 pitchers who both started and relieved at least ten times in those seasons to compare their performance at each. Overall, as relievers, their Fair Run Average, chosen because it sidesteps the problems with assigning run responsibility between starters and relievers that ERA has, was 0.68 lower and their strikeouts per plate appearance .023 higher, with no difference in walks per PA. Dividing the population into power pitchers (K + BB per PA greater than 28 percent), finesse (the same less than 24 percent), and neutral pitchers, the finesse group was a bit more advantaged as relievers (FRA 0.76 lower) than neutral (0.53) and power (0.52).

Seidman, Eric (2010). Drilling down on volatility and consistency. https://www.baseballprospectus.com/news/article/10005/checking-the-numbers-drilling-down-on-volatility-and-consistency/

Based on 1974 to 2009 Retrosheet data, Eric Seidman (2010) noted that predicted ERAs for pitchers with at least 20 starts in four consecutive seasons who rated in the upper fourth and upper fifth in consistency in ERA across those seasons tended to be quite accurate, whereas those in lower fourth and fifth, i.e. the most volatile, tended to outperform their projections by about a tenth of a run.

Seidman, Eric and Russell A. Carleton (2010). Side effects on pitchers' hitting. https://www.baseballprospectus.com/news/article/9932/checking-the-numbers-side-effects-on-pitchers-hitting/

Eric Seidman and Russell Carleton (2010) took on the question of whether a pitcher having to bat or run the bases results in worse pitching the next inning. 2008-2009 PITCHf/x data for pitchers with at least 30 PA revealed a drop-off of 2.6 percent in fastball usage and 0.11 in fastball velocity, with curveballs, sliders, and changeups all taking up the slack. In other words, pitch variety increased. Fastball movements decreased by 0.05 inch horizontally and 0.10 inch vertically; changeups lose 0.23 bertical inches. Having to run the bases had an analogous impact; 1.7 percent of fastballs becoming others pitches and an even slighter (.05) decrease in velocity. In some contrast with batting only, fastball movement drops more horizontally (0.15 inches) than vertically (0.04 inches), with curveballs and changeups losing as much as ¼ inch of movement. Unfortunately, this comparison appears to be against both pitchers who batted and did not get on base and pitchers who did not bat; it would be more informative to have been limited to the first of these groupings. Turning to outcomes and based on PA for seasons for batters with and pitchers facing at least 250 PA (which seasons are not mentioned, but the data surely is from Retrosheet), and controlling for pitcher and batter strength and handedness and pitch count, Eric Seidman and Russell Carleton (2010) only uncovered a .004 decrease in strikeouts per PA, with most of those K's lost becoming outs-in-play, and a slight increase in hits going for extra bases rather than singles So there are batting and baserunning effects for pitchers, particularly in terms of pitch movement, but they seem to have minimal impacts on outcomes.

Sela, Rebecca J., & Jeffrey S. Simonoff (2007). Does momentum exist in a baseball game? In Jim Albert and Ruud H. Koning (Eds.), *Statistical thinking in sports* (pages 135-151). Boca Raton, FL: Chapman & Hall/CRC

Time to examine another myth; that momentum effects exist within an inning such as when a team starts getting baserunners and scoring runs, it is likely to continue. Sela and Simonoff (2007) began with a standard Markov table of transition probabilities

between different base-out situations but added sets of variables via logistic regression allowing for a series of more complicated models. The sets respectively incorporated:
1 – player quality effects; batter on-base and slugging averages and pitcher WHIP and strikeouts per nine innings, plus whether the home or away team is batting.
2 – situational effects; the number of batters faced and pitches thrown by the current pitcher in the game and the OBA and SLG for the next batter in case "protection" was real.
3 – momentum effects, the issue at hand; the result for the previous plate appearance unless the current batter began the inning, and the number of batters and runs scored since the last out.

Using Retrosheet data from 2003 and 2004 for establishing the models and 2005 for validating them, the authors noted that addition of the player quality and situational effects resulted in more accurate modeling. However, puncturing the relevant myth, the only momentum effects uncovered were negative; with two runners on base and either one or two outs, there is a slight increase in the odds that outs begat more outs. Consistent with this result, negative binomial regressions indicated that, in those situations, average runs for the remainder of the inning for the team at bat were lower than expected given base-out situation and current batter and pitcher quality. The authors did find support for one myth; double plays really were rally killers, decreasing subsequent run scoring more than expected; the authors did not consider whether this finding was responsible for the "anti-momentum" effects.

Shamsie, Jamal and Michael J. Mannor (2013). Looking inside the dream team: Probing into the contributions of tacit knowledge as an organizational resource. *Organization Science,* Vol. 24 No. 2, pages 513-529.

There has been a lot of academic studies (mostly quite poor) examining the relationship between player and team performance. Somewhat more interesting is Shamsie and Mannor's (2013) attempt to measure the impact of factors over and above those related to sheer player skill, using data from 1985 gleaned from the Lahman Archive and Retrosheet. Although they did use one factor indirectly related to skill, the number of game appearances for a team's roster, the others included managerial experience both overall and with the relevant team, past playoff experience for manager and players, and three measures of team stability: the number of players with the team for at least three years, game-to-game consistency in starting lineups, and maintaining the same manager during a season. Every included factor has a significant, although in some cases small, impact on team winning percentage.

Shu, Pei Zhe (2016). Arsenal/Zone Rating: A PITCHf/x based pitcher projection system. *MIT Sports Analytics Conference.*

Shu (2016) proposed a pitcher projection system that combines PITCHf/x data on pitch speed, movement, and location with Retrosheet play-by-play data. Based on 2008 to 2014 data, the author claimed accuracy comparable to other projection methods and more success at predicting breakout and breakdown seasons as measured by 33 percent increases and decreases in performance.

Silver, Nate (2003). Leading off.
https://www.baseballprospectus.com/news/article/2149/lies-damned-lies-leading-off/

Here are figures showing how OBA became more centralized to #3 and #4 batters over time.

Order #	1982-1989	1999-2000	2001-2002
1	.336	.349	.332
2	.333	.346	.331
3	.349	.384	.379
4	.345	.375	.368
5	.329	.356	.338
6	.322	.345	.327
7	.315	.326	.318
8	.308	.329	.312

Silver, Nate (2003). Hitting the wall.
https://www.baseballprospectus.com/news/article/2128/lies-damned-lies-hitting-the-wall/

Do batters do better or worse after July 1st (second half of season)? Does age impact on this? The following was based on 1999 through 2001:

Improvement (Decline) After July 1

Age	n	BA	OBP	SLG	OPS
21	15	+.018	+.021	+.030	+.051
22	41	-.002	+.004	-.008	-.004
23	98	+.007	+.001	+.015	+.016
24	141	-.005	-.009	-.015	-.024
25	167	-.003	+.000	-.011	-.011
26	168	-.003	-.005	-.003	-.008
27	180	+.001	+.001	-.007	-.006
28	181	-.012	-.009	-.018	-.027
29	166	-.006	-.003	-.022	-.025
30	143	-.009	-.010	-.025	-.035
31	141	-.013	-.012	-.030	-.042
32	117	+.001	-.002	-.013	-.015
33	95	-.003	+.001	-.009	-.008
34	94	-.009	-.008	-.022	-.030
35	73	-.001	+.002	-.021	-.019
36	53	+.001	-.002	+.001	-.001
37	29	-.022	-.025	-.050	-.075
38	14	+.001	-.009	-.013	-.022
Young'uns	(21-24)	+.001	-.002	-.002	-.004
Mid-Career	(25-29)	-.006	-.004	-.015	-.019

Veterans	(30-33)	-.007	-.007	-.021	-.027
Old'uns	(34-38)	-.006	-.006	-.020	-.026

How about pitchers?

	Improvement (Decline) after July 1					
Age	n	BA	OBP	SLG	OPS	K Rate
21	11	+.005	+.035	-.024	+.011	-3.3%
22	36	+.019	+.014	+.036	+.050	-1.3%
23	69	+.004	+.002	+.004	+.006	+0.8%
24	100	+.003	-.002	-.008	-.010	+0.0%
25	101	+.007	+.003	+.011	+.014	-0.1%
26	114	-.001	-.003	-.018	-.021	+0.4%
27	113	+.000	-.006	-.011	-.017	+0.7%
28	111	+.011	+.007	+.017	+.024	-0.2%
29	96	+.001	-.007	+.002	-.005	-0.2%
30	97	-.008	-.009	-.006	-.015	+0.4%
31	88	-.002	-.007	-.007	-.014	+0.1%
32	83	+.007	+.009	+.009	+.018	-0.7%
33	69	-.007	-.007	-.023	-.030	-0.6%
34	50	+.003	+.001	-.001	+.000	+0.0%
35	33	+.002	-.004	-.009	-.013	-1.3%
36	26	+.016	+.019	+.015	+.034	-0.6%
37	23	-.006	-.006	-.022	-.028	-0.3%
38	19	-.009	-.009	-.040	-.049	+0.2%
Young'uns	(21-24)	+.006	+.004	+.002	+.006	-0.1%
Mid-Career	(25-29)	+.004	-.001	-.000	-.002	+0.2%
Veterans	(30-33)	-.003	-.004	-.006	-.010	-0.1%
Old'uns	(34-38)	+.002	+.001	-.008	-.007	-0.4%

In interpreting these figures, I suggest trusting the last four rows of each as the greater sample sizes iron out random variation.

Silver, Nate (2003). Redefining replacement level.
https://www.baseballprospectus.com/news/article/2032/lies-damned-lies-redefining-replacement-level/

Nate Silver argued that the concept of replacement level as commonly understood is problematic, because the quality of the player that a team would use as a replacement is dependent on how long the replacement needs to be used for. The longer the time the replacement player is needed, the better the player required. The following figures are for Batting Runs per PA based on career PA (1973-1992). Note that they increase very quickly until 50 PA and then ever more slowly afterward.

Min PA	Max PA	n	Mean PA	BR/PA
1	5	53	3.8	-0.1268
6	10	41	8.6	-0.0812
11	20	46	16.7	-0.0774

21	30	49	26.3	-0.0553
31	50	48	41.2	-0.0587
51	70	54	60.7	-0.0343
71	110	49	91.6	-0.0342
111	150	51	130	-0.0354
151	200	39	176	-0.0291
201	300	59	253	-0.0283
301	400	44	355	-0.0277
401	600	58	506	-0.0253
601	900	49	738	-0.0196
901	1200	50	1020	-0.0175
1201	1600	52	1409	-0.0096
1601	2200	49	1902	-0.0083
2201	3000	45	2595	+0.0015
3001	4200	48	3536	+0.0061
4201	5500	49	4835	+0.0102
5501	10184	24	7271	+0.0231

The relationship is thus curvilinear and can only be represented arithmetically by a logarithm. Here is his equation:

BR/PA = 0.0154 * ln(PA) – 0.117

Here is an equation based on the one just above that defines a replacement level as a function of career PA. He called it Progressive Runs Above Replacement (PRAR).

PRAR = BR – PA * (.0154 ln(PA) – .1324)

The replacement level player defined this way produces about 76 percent as many runs as the average player.

Silver, Nate (2003). Batter vs. pitcher matchups.
https://www.baseballprospectus.com/news/article/1986/lies-damned-lies-batter-vs-pitcher-matchups/

Nate Silver used 2002 Retrosheet data to break down batter/pitcher matchups by “power” vs. “finesse,' based on the square root of (walk rate X strikeout rate). Finesse was defined as .10 or less, power as .14 or more, and neutral as between the two, resulting in about one-third of the players in each of these three categories. First, the all batters vs. the two types of pitchers,

Pitcher	Power	Finesse	All Pitchers
BA	.241	.276	.261
OBA	.339	.327	.333
SLG	.387	.434	.417
KRate	21.9%	13.3%	17.1%

which shows that batters hit better and walk less (compare BA and OBA) against finesse pitchers, and all pitchers vs. the two types of batters,

Batter	Power	Finesse	All Batters
BA	.257	.265	.261
OBP	.361	.312	.333
SLG	.455	.390	.417
KRate	22.4%	13.6%	17.1%

which shows that pitchers give up more extra base hit and walks to power hitters. No surprises here. Now for some further breakdowns, comparing the data with matchup predictions based on (I am guessing) Dallas Adams' log5 method. First, power pitchers vs. finesse batters:

Power Pitcher v Finesse Batter, 2002

	Actual	Expected
BA	.244	.244
OBP	.311	.318
SLG	.362	.361
KRate	18.0%	17.7%

Finesse batters do poorly against power pitchers, but no worse than would be expected given their overall performance. Next, finesse pitchers vs. power batters

Finesse Pitcher v Power Batter, 2002

	Actual	Expected
BA	.278	.271
OBP	.356	.355
SLG	.487	.473
KRate	17.6%	17.7%

Better performance than just above, and perhaps a bit more production that what would be expected, Next, power pitchers vs. power batters

Power Pitcher v Power Batter, 2002

	Actual	Expected
BA	.239	.236
OBP	.371	.367
SLG	.429	.425
KRate	27.6%	28.2%

Nothing noteworthy here. Finally, finesse pitchers vs. finesse batters

Finesse Pitcher v Finesse Batter, 2002

	Actual	Expected

BA	.280	.279
OBP	.313	.306
SLG	.407	.407
KRate	10.2%	10.5%

The same. Overall, there is no evidence for platooning based on this sort of matchup.

Silver, Nate (2003). Pitcher vs. hitter matchups (Holes part deux).
https://www.baseballprospectus.com/news/article/1936/lies-damned-lies-pitcher-vs-batter-matchups-holes-part-deux/

A follow-up to the above, using 2000-2002 data. The question here is whether increasing the number of batter/pitcher matchups has any effect on long-time performance. There is a selection bias precluding a simple analysis of performance with number of times faced one another, because longer careers for each mean more matchups, and better players have longer careers. As a consequence, weaker players will drop out of the sample, meaning that overall means would wrongly imply that players improve as matchup PAs increase. Nate did a more sophisticated analysis comparing actual with expected outcome specific to that matchup. When you do, you get no consistent effect either way.

Silver, Nate (2004). Groundballs in the mix.
https://www.baseballprospectus.com/news/article/2885/lies-damned-lies-groundballs-in-the-mix/

The following is copied and pasted from the webpost:

This is why I assert ... that groundball ratio is a better predictor of home run rate than is home run rate itself. I looked at league- and park-adjusted statistics for all pitchers from 1975 onward who faced at least 500 batters in two consecutive seasons (1975 is the year in which reliable groundball-flyball data begins to be available from Retrosheet):

1. The correlation between home run rate in year N and home run rate in year N-1 is .326 (note that it is a little bit higher than in the previous example since we've increased the batters faced threshold).
2. The correlation between home run rate in year N and groundball rate in year N-1 is –.345. Though the sign proceeding the correlation figure is negative (since a *higher* groundball ratio tends to predict a *lower* home run rate), the magnitude of the correlation is a bit higher.

Of course we can do better still if we account both for home run rate and for groundball rate in the previous season. A simple regression model that uses home run rate in year N as the dependent variable, and home run rate in year N-1 as the independent variable, is capable of explaining only about 11% of the variance in home run rate for the sample of pitchers we've taken above. If groundball rate in year N-1 is included as a second independent variable, the explanatory power increases sharply to 16%. We can get up closer to 20% if we include other factors like strikeout rate and walk rate (and do considerably better than that if we look at three years worth of previous seasons data, as PECOTA does)–but all the while, groundball rate maintains the largest influence on predicting home runs allowed.

Silver, Nate (2004). Making RBIs useful.
https://www.baseballprospectus.com/news/article/2818/lies-damned-lies-making-rbis-useful/

Nate proposed a useful RBI stand-in, which he called Context-Independent Run Batted In (CIRBI). It is as follows:

((Percentage of runners on third driven in multiplied by league average for that) +
((Percentage of runners on second drive in multiplied by league average for that) +
((Percentage of runners on first driven in multiplied by league average for that)
Multiplied by number of plate appearances
Plus homers

Note that it first provides a proportion of base runners driven in that is indeed independent of the presence or absence of the number of opportunities to do so, which

is beyond the batter's control, weighted for batter opportunity as measured by PA. It then adds the run that homers contribute, which is under the batter's control.
The following was the 2003 leaderboard, which is instructive:

Player	CIRBI	RBI
Delgado_Carlos	138	145
Pujols_Albert	131	124
Sheffield_Gary	131	132
Rodriguez_Alex	128	118
Helton_Todd	124	117
Thome_Jim	124	131
Sexson_Richie	122	124
Wilson_Preston	120	141
Wells_Vernon	120	117
Lee_Carlos	117	113
Anderson_Garret	117	116

Note that the CIRBI can be interpreted the same way as RBI, and that a couple of players' figures, particularly Preston Wilson, were affected by the de-contextualization.

Silver, Nate (2004). Using the Golden Run Ratio.
https://www.baseballprospectus.com/news/article/3559/lies-damned-lies-using-the-golden-run-ratio/
Silver, Nate (2005). Introducing ORVY.
https://www.baseballprospectus.com/news/article/4003/lies-damned-lies-introducing-orvy/

Here are runs scored by one club in an inning for 2003 (copied and pasted):

Runs Scored	Frequency	Percent
0	30922	71.1%
1	6845	15.7%
2	3011	6.9%
3	1507	3.5%
4	670	1.5%
5	305	0.7%
6	117	0.3%
7	62	0.1%
8	12	0.0%
9	6	0.0%
10	6	0.0%
11	0	0.0%
12	1	0.0%
13	1	0.0%
14	1	0.0%

Note that it is a very neat exponential decay function such that the ratio between 5 and 6, 4 and 5, etc. down to 1 and 2 are pretty close, with that for 0 and 1 about twice as big as the others. Nate called the relationship the Golden Run Ratio (g), and learned that it is greater for lower-scoring teams. Nate computed a couple of g's; 4.33 for 5 runs per game, 5.64 for 3 runs per game. In his 2005c, In his 2005c, Nate used these figures to compute win probabilities for given moments in the game. Here is, using Nate's example, the probabilities of the home team winning a game following the bottom of the seventh inning:

Score	Home Win %
+5 runs	98.2%
+4 runs	96.3%
+3 runs	92.6%
+2 runs	86.0%
+1 runs	74.1%
Tied	50.0%
-1 runs	25.9%
-2 runs	14.0%
-3 runs	7.3%
-4 runs	3.7%
-5 runs	1.8%

As you can see, the probabilities differ very little with large run surpluses or deficits but quite a bit with small ones, which reflects differences in leverage. One can use these figures to compute the change in win probability if a team scores a given number of runsin an inning. To continue the example, with the score tied, one extra run would increase win probability by 74.1 minus 50 or 24.1 percent, scoring a second run by 86 minus 74.1 or 11.9 percent, and so on. Nate then introduced One-Run Value Yield (ORVY). The ratio of the first increase by the second; in this case, 24.1 minus 11.9 or 2.02. The higher the ORVY, the more valuable one-run strategies (sacrifice bunts) are relative to multiple runs. In this circumstance, a one-run strategy would be a good choice. ORVY has the following implications:
1 – The later in the game, the higher the ORVY, so the more valuable one-run strategies are compared to multi-run.
2 – When a team is one run behind, its ORVY will always be 1. This is then the break-even point for scoring one run versus two.
3 – A team should never use a one-run strategy when trailing by more than one run, because the ORVY is too small. Again, using that chart, ORVY when two runs back would be 11.9 divided by 24.1, which is 0.49. Note that this is the inverse of the first example, which demonstrated that a one-run strategy would be good for a team two runs ahead.

Smith, David W. (2006). Does walking the leadoff batter lead to big innings? *Baseball Research Journal,* No. 35, pages 23-24.

Our fearless leader's papers are customarily posted on the Retrosheet research page, but this one is not. In answer to a baseball myth expounded often (and inspired by one of those expositions by Tim McCarver), Dave showed that, from 1974 to 2002, walks to leadoff batters have the same impact on scoring as any other way to get on base, in so doing adding another piece of evidence to others showing that all ways of getting on base have equivalent impacts.

Smith, Erin E. and Jon D. Groetzinger (2010). Do fans matter? The effect of attendance on the outcomes of Major League Baseball games. *Journal of Quantitative Analysis in Sports,* Vol. 6 Issue 1 Article 4.

The most strongly supported explanation for the consistent 54% home field advantage for baseball is the impact of fan support. In one piece of relevant evidence Smith and Groetzinger (2010) combined data for the years 1996 through 2005 from the Retrosheet and Baseball Archive databases with weather information from the National Climatic Data Center, along with the Questec pitch monitoring system for 2001 and 2002. Overall, increasing attendance by one standard deviation (about 25 percent) resulted in what the authors say was .64 additional runs (I wonder if they really meant run differential) and an increase of 5.4% in the probability of a home team. Hits, doubles, and home runs were all weakly by positively related with attendance, and earned runs allowed negatively associated. In addition, there was a decrease in home team strikeouts as attendance rose, which could signal home plate umpire bias in calling balls and strikes. However, contrasting ballparks with and without the QuesTec system for reviewing umpire ball-strike calls under the questionable assumption that umpires are biased by fan support but the presence of the system would decrease that bias; they could not find any differences.

Smith, Sean (2009). Total Zone data. https://www.baseball-reference.com/about/total_zone.shtml

Sean Smith's (2009) TotalZone uses Retrosheet data to evaluate fielders, with the type of available data determining the exact method. When data on specific plays is missing, Sean would do the following:
Step 1 – Compute every batter's career proportion of batted balls for which plays were made at each fielding position. If the batter is a switchhitter, then do this separately for left- and righthanded plate appearances.
Step 2 – Assume that this proportion remains the same for hits, and based on how often the batter and each fielder play against one another, estimate how many hits each fielder should be assessed based on that proportion.
Step 2 – For every fielder, sum the result of Step 2 across all batters played against.
Step 3 – Divide the results of Step 2 by every fielder's total fielding chances, computed by

(Total plays made) + (Errors) + (Result of Step 2)

Step 4 – Do park adjustments, and convert to runs responsible for (.75 per hit for middle infielders, .80 for the infield corners, and .85 for outfielders).

When batted ball type and fielder is available, one can estimate responsibility for hits somewhat more accurately, by giving third basemen 60 percent and shortstops 40 percent of the debit for singles to right, shortstops 52 percent and second basemen 48 percent of the charge for singles to center, and first basemen 55 percent and second basemen 45 percent of the deduction for singles to right. Groundball extra base hits are presumed to be down the lines and so totally given to the corner infielders. The plays that fielders make and do not make can be compared to league average for different batted ball types and pitcher/batter handedness. I assume that outfielders would be judged based on proportion of relevant plays made.

When actual hit location is available, one can use that without making any estimates. Outfielder arms, infielder double plays, and catcher performance are also evaluated; see the referenced article on these.

Smith, Sean (2010). Relievers yesterday and today. In Dave Studenmund (Producer), *The Hardball Times Baseball Annual 2010* (pages 176-182). Skokie, IL: Acta Sports.

Using Retrosheet data and limiting analysis to seasons in which they achieved a WAR of at least 3.0, Sean Smith (2010) examined changes in elite relief pitcher usage beginning about when relief specialists became prevalent (1954) and ending in 2008. First and foremost, although the average number of appearances for top relievers has stayed about constant at 65 during this period, the average number of innings pitched has substantially dipped from about 115 to 125 until 1984 down to the current 75 or so. Second, as greater workload allows for higher WAR, this decrease resulted in only one of the top twelve WAR seasons (Mariano Rivera, 1996, 5.4 WAR) occurring after 1986. Third, as fewer innings means less overuse and more staying power, the likelihood of a reliever following up a 3.0+ WAR season with one at least at 1.0 has increased from only 50 to 60 percent through 1980 and close to 70 percent since. Thus, several indicators suggest the early-mid 1980s as a breakpoint between the 2 and the 1 inning closer. Interestingly, Leverage Index only increased a bit, from for example 1.58 from 1954 to 1969 to 1.77 in 2005 to 2008.

Smith, Sean (2011). Do catchers have an ERA? In Dave Studenmund (Producer), *Hardball Times Baseball Annual 2011* (pages 143-146). Chicago, IL: Acta Sports.

In his book with Tom House, *Diamond Appraised,* Craig Wright introduced the idea of Catcher ERA, in which a catcher is evaluated according to whether the ERA of his team's pitching staff is better or worse when he is behind the plate as compared with his team's other catchers. Sean Smith (2011) examined the consistency across seasons using Retrosheet as part of the data source and analyzing these data via the "matched inning" prorating method Craig used. In order to neutralize differences in team fielding applying a DIPS-based bottom-up estimate of runs allowed that he concocted rather than the actual total to neutralize differences in fielding. He observed a .21 correlation across consecutive seasons starting with 2003 and ending with 2009 for 70 catchers with at least 2000 "matched" plate appearances (this would double count PAs for 2004 to 2008 as each of those seasons would be included twice). This implies some but not a lot of consistency across seasons in specific catcher's relative ranking. He also noted no staff ERA improvement as catchers gain experience, inconsistently with some who had found some (Tom Hanrahan, the Hirdt brothers in the 1981 *Baseball Analyst* book) and consistently with others (Keith Woolner in the 1999 *Baseball Prospectus*).

Song, Alex, Thomas Severini and Ravi Allada (2017, February 7). How jet lag impairs major league baseball performance. *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 114 No. 6, pages 1407-1412.

Carleton, Russell A. (2017). Blame it on the plane. https://www.baseballprospectus.com/news/article/31079/baseball-therapy-blame-it-on-the-plane/

Song, Severini and Allada (2017) replicated earlier claims about the impact of jet lag on home field advantage, based on home teams using 1992-2011 data (likely from Retrosheet). In fact, the home field advantage was nullified for teams returning home west-to-east through either two or three time zones when the visiting team had stayed in the same time zone; the analogous effect for home teams traveling east-to-west also occurred but more weakly. Home-team slugging average, and even more specifically number of doubles hit, were affected identically, as were slugging average by opposing team, runs allowed, and fielding-independent pitching, the latter two due to giving up more home runs. Visiting teams were also affected negatively by travel, although direction did not matter. Displeased with Song et al. averaging across seasons and players within teams in their analysis, Russell Carleton (2017) duplicated the study at the plate appearance level using 2012 through 2016, with time lags considered significant if either two or three hours. He got several significant findings across different types of game events but none were consistent throughout.

Soto Valero, César (2016). Predicting win-loss outcomes in MLB regular season games: A comparative study using data mining methods. *International Journal of Computer Science in Sport,* Vol. 15 No 2, Article 7.

Soto Valero (2016) compared the capability of data mining methods as predictors of game outcomes using Retrosheet data for 2005 through 2014.

Swartz, Matt and Eric Seidman (2010). Introducing SIERA: Part 1. https://www.baseballprospectus.com/news/article/10027/introducing-siera-part-1/

Swartz, Matt and Eric Seidman (2010). Introducing SIERA: Part 2. https://www.baseballprospectus.com/news/article/10032/introducing-siera-part-2/

Swartz, Matt and Eric Seidman (2010). Introducing SIERA: Part 3. https://www.baseballprospectus.com/news/article/10037/introducing-siera-part-3/

Swartz, Matt and Eric Seidman (2010). Introducing SIERA: Part 4. https://www.baseballprospectus.com/news/article/10042/introducing-siera-part-4/

Using Retrosheet data from 2003 through 2009, Baseball Prospectus's Matt Swartz and Eric Seidman introduced SIERA (Skill-Interactive Earned Run Average), a very complicated pitching metric that they claim to be the most accurate predictor of them all. As with any metric based on regression analyses for specific seasons, that claim will only be true for those seasons. Nonetheless, the concept has been influential, and FanGraphs has its own version.

Tango, Tom M. (2008). With or without you. In Dave Studenmund (Producer), *The Hardball Times Baseball Annual* (pages 191-198). Skokie, IL: Acta Sports.

Tom Tango (2008) proposed a creative method that he called for With Or Without You (WOWY) for evaluating catcher ability to prevent passed balls and wild pitches, thwart attempted steals, and pickoff runners. For a given catcher:
1 – Choose a pitcher he caught.
2 – Count how many WPs, PBs, and innings occurred with that pitcher/catcher combination.
3 – Count how many WPs, PBs, and innings occurred with that pitcher and other catchers, and then use the ratio of WPs and PBs per inning to estimate the number that would have occurred if the other catchers had caught that pitcher the same number of innings and the catcher under examination.
4 – Comparing the results of steps 2 and 3 reveals how much better or worse the catcher under examination was than the other catchers for the given pitcher.
5 – Repeat these steps for all other pitchers the catcher under examination caught, and sum the results for an overall index.

Tom performed this task using Retrosheet data from 1972 through 1992. According to his chart displaying data for individuals during that era, the ones everyone thought were good (e.g., Jim Sundberg, Gary Carter) are indeed toward the top and those everyone thought were bad (e.g., Charlie Moore, Ron Hassey) toward the bottom. Tom noted that this method presumes that the other catchers to whom the catcher under examination is compared are league average; he tested the assumption and found it to be reasonably defensible. Incidentally, he noted that Tom Ruane had previously suggested this method. Michael Humphreys (2011) extended this idea to the evaluation of all fielders, by comparing a specific fielder's performance with those sharing his position on the same team in the same year.

Tango, Tom M. (2008a) With or without…Derek Jeter. In Dave Studenmund (Producer), *The Hardball Times Baseball Annual* (pages 147-152). Skokie, IL: Acta Sports.

Tom Tango's With Or Without You also works for fielding in general. Tom described it in the context of Derek Jeter; Michael Humphreys (2011, pages 84-86) did

a nice job of describing it in general, and I will use Michael's description. When evaluating a particular fielder, the analyst uses relevant Retrosheet data to do the following:
1 – Choose a pitcher he fielded behind.
2 – When the fielder in question was playing, count how many batted balls in play that pitcher gave up, and how many of these batted balls were fielded by the fielder in question.
3 – When the fielder in question was not playing, count many of batted balls in play that pitcher gave up, and how many of these batter balls were fielded b others playing the same position as the fielder in question.
4 – Comparing the results of steps 2 and 3 reveals how many more or fewer balls the fielder in question would have successfully fielded than the "typical" other shortstop would have behind the same pitcher.
5 – Repeat these steps for all other pitchers the fielder in question played behind, and sum the results for an overall index.
Rather than the fielder's team's pitchers, one can do a WOWY analysis across opposition batters, different ballparks, and different baserunner situations to see if the results look any different.

Tango, Tom M. (2009). Catcher 911. In Dave Studenmund (Producer), *The Hardball Times Baseball Annual* (pages 191-198). Skokie, IL: Acta Sports.

Using Retrosheet data, Tom Tango (2009) examined every player who caught at least one game between 1956 and 2007 to compare the fielding performance of (1) those with at least half of game appearances as a catcher in a given season, (2) those who did not catch at least half of their game appearances in a given season but had in the past, (3) those who never caught at least half of their game appearances in a given season but did catcher at least ten times in their careers, and (4) those who caught fewer than ten games in their careers. Per 5000 batters (an approximate season of catching), those in the first three categories averaged 0.8, –8.5, –3.9, and a whopping –49.6 runs per season (measured as 0.5 runs gained for every caught stealing and pickoff and –0.25 runs for every stolen base, balk, wild pitch, and passed ball). In short, true emergency catchers were far worse fielders than even those who caught only occasionally. Keep in mind that the sample sizes for the last two categories were tiny.

In a second inquiry in the same book chapter, Tom compared the batting performance of catchers when playing on consecutive days versus having a day or two off between appearances, adjusted for relative playing time in each category. Contrary to expectation, there was absolutely no impact, with wOBAs of .323 for each. Finally, Tom compared the offensive performance of players before and after their 29th birthday. First basemen, other infielders, and outfielders produced about three runs per 650 plate appearances in the older category; catchers only 1.5 runs.

Tango, Tom M., Mitchel G. Lichtman and Andrew E. Dolphin. *The Book: Playing the Percentages in Baseball.* TMA Press.

I begin with an editorial comment: This book belongs on the shelf of anybody who seriously studies quantitative baseball data. The entire book is based on sophisticated analysis using Retrosheet data (different seasons for different analyses, so I will skip the details on what seasons were employed). I will only list the themes, as describing all the findings would take too long:

In Chapter 1, entitled Toolshed, the authors explain the basics of run expectancy tables and their interpretation, and compute the “run value” of 20 possible events occurring during games, lists as demonstrations the run value of home runs at each base-out situation and the odds of scoring different numbers of runs at each base-out situation given an average of 3.2 or 5 runs per game. They also include the odds of a team winning the game given every base-out situation in every half inning (top of first through bottom of ninth) for every increment from being ahead by four runs to behind by four runs and the “win value” of the 20 events, which tells you how critical the situation is in which the event occurs on average. Finally, they define Tango’s measure of offensive performance, weighted on-base average, which in a linear weights-type formula but calibrated to be interpreted as one interprets OBA.

Chapter 2 takes on the issue of batting and pitching streaks, this time using 2000-2003 Retrosheet data. They note tiny but discernible tendencies for batters who have been hot or cold for five games to stay that way for a few more games, and the same for pitchers who have been hot over their last four appearances (but not for cold). However, as they did not correct for strength of opponent or ballpark, one should not read too much into this.

Chapter 3 is on batter/pitcher matchups and notes that specific player/player matchups probably are meaningless, replicates previous findings for lefty/righty and groundball/ flyball tendency matchups, finds no interaction effects between batters/pitchers good at controlling the strike zone or at making contact, and not much evidence that good pitching stops good hitting.

Chapter 4 addresses various situational issues. Contrary to all other research, the authors do find consistent clutch hitting tendencies for batters, but they are tiny and practically meaningless. They note no analogous clutch pitching effect for relievers. Pinchhitting indeed does lead to worse performance than being in the lineup, and it is not because pinchhitters tend to face fresh relievers in the late innings. There is no performance difference between hitting with runners on versus base empty.

Chapter 5 turns to the lineup. Here they weight run value by lineup opportunity (i.e., each lineup position has about .11 more plate appearances than the next and differing proportions across the base/out situations, i.e. leadoff batter comes up with fewer base runners than any other), and conclude consistently with received wisdom that the leadoff batter should indeed be the best on-base average player and the last four slots (with an exception to be noted below) should have the team’s worst hitters in descending order of run production. In contrast, the number 3 slot should have a weaker hitter than #s 2, 4, and 5. Again consistent with tradition, good basestealers/baserunners ought to be before batters who hit singles and don’t strike out, and the “pitcher bats eighth/pre-leadoff hitter bats ninth idea does work if the pitcher is an average or better hitter for the position.

Chapter 6 considers the standard platoon differential. Most of what is here replicates the findings of several others concerning batters, but there is one useful addition: the platoon differential is not in general large enough to counteract the performance of decrement for pinchhitters, such that one should only pinchhit for platoon advantage if the pinchhitter is considerably better than the batter replaced.

Chapter 7 features the starting pitcher, mostly concerning workload issues. Pitchers do perform a bit worse as the game continues on average. Across games, they perform best with five days rest, but the argument for a six-man rotation falters considering the (absence of) quality one’s sixth starter would likely possess. Pitchers who blow through the first nine hitters tend to return to normal for the next nine, whereas pitchers who are hammered by the first nine batters still tend to struggle with the next nine and likely are having a bad day. Finally, pitchers usually perform better as relievers as starters, with the possible exception of starters pitchers with little or no experience as relievers at all.

Chapter 8 is the relief pitcher’s turn. Conceptually, they compared the generic very good relief pitcher (analogous to one who would win 68% of their games) to the generic average one (50%). The 18% difference between the two breaks down to 2% an inning. In theory one would always do better with the very good reliever, but in practice you don’t want to overwork him and so look for situations in which you don’t lose much using the average reliever. Assuming long-term equal usage, the strategic implication is that a very good relief pitcher is worth bringing in a game rather than an average one if the odds of the good reliever winning is more than 2% more than the average reliever in a given base/out/inning situation and not if the odds are less than 2%. Using Retrosheet data from 1999-2002, they determined, for example, that the very good reliever need only be used in the ninth inning/three run lead situation (the easiest possible save given today’s scoring procedures) if there is a baserunner with no outs or two baserunners with no or one out. Using historic data, they also argue that very good relievers can be trusted to not lose effectiveness up to about 25 pitches, which on average allows bringing them in during the eighth inning. Finally, they claim (and present evidence) that relievers in general do not lose effectiveness if used two or three days in a row. I am less confident in the last of these claims is defensible given that such usage is rare for the typical pitcher, and their data may not represent what would happen long-term if such usage became commonplace.

Chapter 9 is the most detailed analysis of the sacrifice bunt as a strategic tool thus far presented, taking up more than 50 pages of their book. They used Retrosheet data from 2000 through 2004 throughout, and, using Palmer’s method, showed that the runner on first/zero outs sacrifice was overall even more harmful than in Pete’s findings, likely due to the overall increase in offense. In general, however, they applied a different and very useful method. For example, rather than comparing expected runs between runner on first/no out and runner on second/one out, they compared runs scored for the rest of the inning between runner on first/no outs when sacrifices were attempted and runner on first/no outs when sacrifices were not attempted. Note the term attempted: one can attempt to sacrifice, foul the pitch off, and then hit a home run on the next pitch; and these successful at bats ought to be included as well as the failures. Anyway, their wealth of findings are too numerous and complicated to describe in detail, and interested reader should consult *The Book*. In summary, the value of the sacrifice is affected by strength of the batter and of the batter on deck (the lower the on-deck's OBA, the better the bunt is), infield alignment (better if the infield is playing back), inning (better earlier in the game as infielders are less likely to be playing in for it), run environment (better when runs are more scarce), bunter skill, and baserunner speed. In addition, one should not use the same strategy all of the time as the other teams will respond accordingly with their defensive alignment, so randomly placed variation to decrease predictability will help.

Chapter 10 considers the intentional walk. Based on 2000-2004 Retrosheet data, there were no base-out situations in which the IBB decreased expected runs for the opposition overall. This was true even when the batter in question is much better than the batter on deck, including the #8 batter with the pitcher expected to come to the plate. There are a couple (second and third / one out, third / one out) in which it increases the defensive team's odds of winning, but by less than one percent. Interestingly, these are among the situations in which managers used it the most during those years, implying some intuitive understanding of the situation. Other exceptions are tied games in the bottom on the ninth when the IBB helps if it doesn't advance the lead runner, and when you have reached a 3-0 count against a very good hitter.

Chapter 11 is the stolen bases' turn. Overall success in basestealing during the 1999 through 2002 period of time, about 68%, was in general below the breakeven rate of 72%. The latter rate was dependent on game score (75.4% when three runs ahead and 66.9% when three runs behind) and inning (as the game progresses, the breakeven worsens when the team at bat is behind but improves when the team at bat is ahead). Interestingly, the data also provided evidence consistent with the platitude that baserunners disrupt the defense and improve the fortunes of hitters. Mean wOBA, .358 overall, was .372 with runners on first and less than two outs. Again not surprisingly, that broke down to .378 for lefthanded hitters and .368 for righties.

Finishing in Chapter 12 with the pitchout, the odds of success following a pitchout dropped to 47%. The implication that pitching out is a good strategy must be tempered by the fact that it adds a ball to the count, aiding the batter. That aid is highly dependent on the count. The TMA group (they were uncharacteristically silent on which years they used; I would guess 1999 to 2002) calculated a tiny increase in wOBA from .222 to .245 (corresponding to a scant .03 rise in runs scored) with a pitchout at an 0-2 count, but a large increase of .116 (equivalent to .15 runs) pitching out at 2-1. Combining the two, they estimated the breakeven point for pitchouts when the count is 0-2 and the opposing team believes the odds of an attempted steal are a scant 18 percent (in other words, it's a good strategy at 0-2), but this changes to 54% with a 2-1 count and one out (meaning that the opposing team has to feel that an attempt is more likely than not).

Thress, Tom (2012). Beyond Player Win Average. *Baseball Research Journal,* Vol. 41 No. 2, pages 22-29.

This to all extents and purpose is an updating of Mills and Mill's Player Win Averages analysis, providing ratings for prominent players beginning with 1948 and using Retrosheet data.

Timmerman, Thomas A. (2007). "It was a thought pitch": Personal, situational, and target influences on hut-by-pitch events across time. *Journal of Applied Psychology,* Vol. 92 No. 3, pages 876-884.

Are Black players more susceptible to being hit by pitches? Earlier evidence implied that this may have been true in the 1950s but not anymore. Timmerman (2007) examined whether pitchers from the southern U.S.A. were more likely to hit Black batters than White batters immediately after a home run, after that batter had previously hit a home runs, or one of their own teammates were hit. Using Retrosheet data from 1960 to 1992 and 2000 to 2004 and controlling for batter OPS, whether a DH was used in the game, differential in team scores (assuming the losing team's pitcher would be more likely to hit a batter), and pitcher walks per plate appearance, Timmerman noted overall increases in HBP in all three circumstances. However, opposite to what he expected, White batters were most likely to be hit by southern pitchers after they had homered and after the pitcher's teammate had been hit, with Blacks second and Hispanics last. Interestingly, pitchers not born in the south were more likely to hit Blacks than Whites and Hispanics in those circumstances.

Tollison, Robert D., and Octavian Vasilescu (2011). The designated hitter rule and the distribution of pitching talent across leagues. *Journal of Sports Economics,* Vol. 12 No. 4, pages 448-463.

It stands to reason that good hitting pitchers are a less valuable commodity and poor hitting pitchers less of a problem in a league with a designated hitter than a league without. It follows that a bias toward trading good hitting pitchers from the A.L. to the N.L. and poor hitting pitchers from the N.L. to the A.L. should have occurred around the time of the DH's imposition. Tollison and Vasilescu used the Retrosheet transaction file for trades. Examining (non-Retrosheet) data from 1960 through 1985, and controlling for pitcher quality as measured by ERA, age, and usage patterns as measured by IP, there appeared to be such a bias in 1972 and 1973 but not before and after. A second type of analysis found the same for 1970 (perhaps imagining the coming of the rule change) and 1972.

Turocy, Ted (2004). A theory of theft: The strategic role of the stolen base in baseball. Unpublished manuscript.

Turocy, Ted (2014). An inspection game model of the stolen base in baseball: A theory of theft. Available at www.gambit-project.org/turocy/papers/theft-20140822.pdf

In two unpublished papers, Ted Turocy presented mathematical models on the strategic value of the stolen base attempt.

Turocy, Theodore L. (2005). Offensive performance, omitted variables, and the value of speed in baseball. *Economics Letters,* Vol. 89, pages 283-286.

Ted Turocy (2005), using Retrosheet data from 1974 to 1992, came up with an overall breakeven point for stolen base attempts of .717.

Turocy, Theodore L. (2008). In search of the "last-ups" advantage in baseball: A game-theoretic approach. *Journal of Quantitative Analysis in Sports,* Vol. 4, Issue 2, Article 5.

Is there a last up advantage? Ted Turocy (2008) used Retrosheet data from 1973 through 1992 as data for a simulation assuming two teams of equal quality, and concluded that there is a infinitesimal last-ups advantage of .001 in winning percentage, equivalent to an extra win every six years.

Uelkes, Peter (2012). Game scores. *Baseball Research Journal,* Vol. 41 No. 2, pages 30-36.

This is a detailed examination of starting pitchers using Bill James's Game Score concept, based on more than 117,000 Retrosheet games. The most important part is the discovery that home starters have had a 14.7% advantage over road starters in strikeout/walk ratio, consistent with other research revealing pitch f/x data revealing umpire bias in ball/strike counts in favor of home teams.

Walsh, John (2008). The origin of the platoon advantage. In Dave Studenmund (Producer), *The Hardball Times Baseball Annual* (pages 165-171). Skokie, IL: Acta Sports.

John Walsh (2008) used the Neyer/James Guide to Pitchers as a source for pitcher repertoire and Retrosheet performance data from 1957-2006 to compute platoon differentials for pitchers, and compared those for one group of pitchers who had a relatively large platoon differential (he did not say how much) and a second group with a reverse differential; here is what he found concerning pitch usage (table cut and pasted from Jared Cross, 2015). John gave 5 points if the pitch was listed as the most used in Neyer/James, 3 points if the second most used, and 1 point if the third:

John Walsh's Pitch Usage Points				
		Pitch Usage Points		
Extremum	*# of Pitchers*	*Slider*	*Curveball*	*Changeup*
High-Split	22	53	21	10

Low-Split	29	22	62	42

Note that sliders were associated with high splits and curves and changeups with reverse splits. It is unfortunate that he had not included fastballs in the analysis, as subsequent research has revealed that fastballs along with sliders are the pitches associated with significant pitcher platoon differentials. On his list of pitchers included in the samples, fastball was listed first for all but one of the 22 of the high-split set whereas for the reverse aka low split group 12 of the 29 did not.

Wigley, Jay (2021). Did batters of long ago learn during a game? *Baseball Research Journal*, Vol. 50 No. 1, pages 55-59.

Jay Wigley (2021), using Retrosheet data going back to 1916, uncovered the fact that the TTOP effect appeared for the first time in 1921, the beginning of the “modern” slugging era. In the five years previous to then, a second time dip was followed by a third time return to the level of the first time through.

Wolfersberger, Jesse and Matthew Yaspan (2015). Trying to quantify recency effect. In Dave Studenmund and Paul Swydan (Prods.), *The 2015 Hardball Times Baseball Annual* (pages 360--367). FanGraphs.

Among the many routes to exploring the issue of whether streakiness is a real phenomenon, one of the more useful ones is to see if more recent plate appearances (PA) are better predictors of a given PA’s outcome than more distant-in-the-past PAs. 2013 Retrosheet data implies it does not, with the exception of the result of the immediately preceding PA, which authors Wolfersberger and Yaspan attribute to the increased tendency for both current and previous PA to be against the same pitcher.

Wolfson, Julian, Joseph S. Koopmeiners and Andrew DiLernia (2018). Who’s good this year? Comparing the information content of games in the four major US sports. *Journal of Sports Analytics,* Vol. 4 No. 2, pages 153-163.

The authors used 2010-2015 data from Retrosheet to estimate how informative different proportions of games beginning at the start of the season (first 1/8 of the game, 2/8 of the games, etc.) are for predicting team matchups for the rest of the season. Even with 7/8 of the season finished (140 games), accuracy was never higher than 58 percent for the rest, which are the authors note is not a lot higher than the 54 percent home field advantage, which they used as their comparison model.

Woolner, Keith (1999). Field general or backstop?: Evaluating the catcher’s influence on pitcher performance. In Clay Davenport (Ed.), *Baseball Prospectus 1999* (pages 466-474). Washington, D.C.: Brassey’s Sports. Available at https://www.baseballprospectus.com/news/article/432/field-general-or-backstop-evaluating-the-catchers-influence-on-pitcher-performance/

Woolner, Keith (2000). Catching up with the general: A postscript: A second look at catcher defense. https://www.baseballprospectus.com/news/article/436/catching-up-with-the-general-a-postscript-a-second-look-at-catcher-defense/

Several researchers, including Craig Wright anecdotally in his book with Tom House (*The Diamond Appraised*) and Tom Hanrahan in three articles in *By The Numbers*, uncovered evidence based on ERA that catcher performance improves with experience. The only nay-sayer that I am familiar with was Keith Woolner (1999). Using data from Retrosheet and Total Sports from 1984 through 1997, Keith performed WOWY analyses with every pitcher with each catcher with whom he faced 100 or more batters (sample size = 6347 pitcher/catcher combination). He then calculated the overall run value for the results of those plate appearances for each of the combinations. The distribution of these run values approximated the normal distribution fairly closely, implying that performance differences among catchers either do not exist or do exist but occur randomly. Further, the year-to-year correlation for catchers was a non-existent .02, meaning that performance changes randomly from year to year. Keith re-analyzed these latter data in several ways to see if a subtle effect hidden in the overall trend would appear; the correlations remained very close to zero. After reader criticism stating that his WOWY analysis was invalid because the comparisons were often with different catchers from year-to-year, Keith (2000) restricted it to teams with the same two catchers working with the same pitchers in consecutive years. The resulting correlation, 0.01, supported the original conclusion. I find it difficult to substantively reconcile Tom and Keith's very distinct conclusions.

Woolner, Keith (2001). Temperature and OPS. https://www.baseballprospectus.com/news/article/1058/aim-for-the-head-temperature-and-ops/

One thing that is almost certainly not a "skill" is a tendency to hit better in warm versus cold weather. Keith Woolner (2001), using data from Retrosheet and The Baseball Workshop, developed a database of 224 players with a least 100 PA in both cold and warm weather for both 1999 and 2000, with 72½ degrees at gametime as the cut-off, computed a ratio of cold/warm OPS for each player and then correlated these ratios for the two seasons. The correlation was actually slightly negative, –0.15, implying a small tendency for batters to reverse tendencies from season-to-season. This was a quick-and-dirty study in which Keith did not control for players changing teams, and given what could be relevant characteristics specific to individual ballparks, such changes could well be responsible for the reason the figure was negative rather than close to zero.

Woolner, Keith (2001). Reaching on errors. https://www.baseballprospectus.com/news/article/1145/aim-for-the-head-reaching-on-errors/

Woolner, Keith (2001). More reaching base on errors. https://www.baseballprospectus.com/news/article/1167/aim-for-the-head-more-reaching-on-errors/

Using play-by-play data from 1978 through 2000 (I shall assume Retrosheet) for 1704 players with at least 500 PA, Keith Woolner (2001f) uncovered a year-to-year correlation for reaching base on error (ROE) of only 0.21, but a more robust 0.41 for odd versus even years across careers. In follow-up work (2001g), Keith noted right-handed hitters (1.23% of PAs) to ROE more than left-handed (0.95%), with switch-hitters intermediate (1.12%); a correlation of 0.262 between ROE and groundball/flyball ratio and ROE but only 0.04 between ROE and grounding into double plays (incidentally, the ratio and GIDP correlated at a surprisingly low 0.148).

Woolner, Keith (2001). Response rates. https://www.baseballprospectus.com/news/article/1077/aim-for-the-head-response-rates/

Teams that score a lot tend to score in a lot of innings, and teams that score in a lot of innings tend to not be victims of shutdown innings. Between 1978 and 2000, the correlation between the latter two was .89. There was only a little evidence for a discernible team skill involved in answering opposition runs with one's own; the differences between the proportion of innings scored in and the proportion of such innings following opposition scoring was +0.17.

Woolner, Keith (2001). Walk rate spikes.
https://www.baseballprospectus.com/news/article/1107/aim-for-the-head-walk-rate-spikes/

Keith Woolner (2001) examined whether sudden increases or decreases in offensive production were signals of actual skill change rather than one-year flukes. To do so, he looked all but two of players from 1954 through 2000 (I am willing to bet that he used Retrosheet data) who amassed at least 1000 PA in a three-year span, a fourth season of at least 400 PA spike, and years five through seven (again minimum 1000 PA); the two players, Ozzie Smith (spike in 1982) and Frank Tavares (spike in 1977) had zero homers their first three seasons, which would make the result moot. There were 3220 relevant player-spans His method was:
Step 1 – compute rate per PA for a given metric across the first three years.
Step 2 – compute difference in this rate between year four and result for Step 1.
Step 3 – compute rate per PA of the metric for years five through seven.
Step 4 – compute difference in this rate between years five through seven and result of
Step 5 – compute the difference between the results of Step 2 and Step 4.
It turned out that the results of Step 5 correlated at 0.42 for hits, 0.47 for homers, 0.45 for total bases, 0.51 for strikeouts (with decreases interpreted as improvements), 0.45 for on-base average, and 0.48 for walks. This means that there was some actual skill change, approximately equivalent across the six indices, that had some degree of consistency across players. Keith then replaced the original Step 5 with a different final step:
Step 5* – divide the result of Step 4 by the result of Step 2. This gives you the proportion of the year four spike that was retained in the next three years, a more accurate measure of actual skill improvement than the original subtraction because it is not affected by the size of the metric; for example, that there are so many more hits than homers. This gives you the proportion of the spike that was retained in the following three seasons; Keith called this the “retention percentage.”
He then looked at the “retention percentage” for the 300 players with the biggest gains and 300 with the worst losses, with the following outcome:

Skill	Players who increased in skill	Players who decreased in skill
H	21.9%	45.8%
HR	41.5%	47.5%
TB	29.3%	51.4%
SO	59.7%	43.1%
OBP	42.6%	37.4%
WALK	51.7%	42.1%

It appears that for the 300 biggest improvements, the retention percentage was greatest for those most closely associated with the Three True Outcomes (HR, SO, OBP, BB) and somewhat less so for hits and total bases. That for the 300 largest decreases were more equivalent across the board. Keith admitted that there is some bias in these findings as they did not include ballpark or league effects, which could be considerable in a few cases; i.e. a player moved to Colorado in the spike year and stayed there the following three.

Woolner, Keith (2002). Set lineups.
https://www.baseballprospectus.com/news/article/1339/aim-for-the-head-set-lineups/

Between 1978 and 2000, Keith Woolner (2002, probably using Retrosheet) data uncovered instances of teams using as few as 42 (Red Sox, 1984) and as many as 155 (Angels, 1985) different lineups across a season. The correlation between number of lineups and team wins was a credible –0.39, as one would expect that worse teams would try different combinations looking for a winner. That Red Sox team had one specific lineup that appeared in 66 games; there was one unnamed teams that only used the same exact players in the same batting order twice. Only one team winning less than 70 games used a specific lineup more than fifteen times; in contrast, teams winning more than 90 were more variable in this matter.   In this case, the correlation was 0.33.

Woolner, Keith (2002). Quality Starts.
https://www.baseballprospectus.com/news/article/1623/aim-for-the-head-quality-starts/

Keith Woolner (2002) presented some interesting relevant information anout Quality Starts for the 1978 to 2000 period (likely using Retrosheet as the source). From year to year, the proportion of starts that met the definition was usually in the mid to upper 40s and occasionally lower 50s during those years. When a QS occurred, team winning average was in the upper 60s and low 70s, with that for the pitcher getting credit for the win in the mid and upper 50s and occasionally lower 60s. While none of the above trended a lot during that period of time, the odds of the pitch getting saddled with the loss went up; in the low 20s during the early 1980s, it began approximating 25 percent by the late 1980s and was well over that by the mid 1990s. That decrease was compensated for by an increase in no-decisions for the starter, rising from about 21 percent to about 26 percent in the interim.

Woolner, Keith (2003). A big change for OBP.
https://www.baseballprospectus.com/news/article/1759/aim-for-the-head-a-big-change-for-obp/

Click, James (2004). Another look at OBP: Do speedy runners force more errors?
https://www.baseballprospectus.com/news/article/2981/another-look-at-obp-do-speedy-batters-force-more-errors/

Keith Woolner (2003) proposed adding reaching first on fielder's choices to OBA because, as they are the runner's fault rather than batter's, the batter should not be penalized for the out. The reason they are not included because they are only supposed to be assigned when the official scorer thinks that the batter would have been out otherwise. And they occur more often than you might think: for players with at least 300 PA between 1978 and 2000, they occurred 28.48 percent more often than HBP and 16.62 percent more often than SF. James Click (2004e) examined the issue of whether getting on base due to opposition error should also be included, under the assumption that if faster runners do so more often then it reflects an offensive skill. However, he uncovered no evidence that they do so (but see Woolner's “Reaching on errors” above) and thus no good rationale for the inclusion.

Wyer, Colin (2008). A new framework for offensive evaluation: Total Production. https://statspeakmvn.wordpress.com/2008/10/

Colin Wyers (2008) offered a bottom-up regression-based method that he called Total Production using 1994-2007 Retrosheet data. It is pretty similar to the many others then available, and, to be honest, at this point in time was superfluous.

Wyers, Colin (2008). Run expectancy by count. https://statspeakmvn.wordpress.com/2008/11/page/3/

Here is a run expectancy chart for the end of plate appearances at each count, from Retrosheet 1994-2007 data.

BALLS	STRIKES	RUNS
0	0	0.595
0	1	0.556
0	2	0.390
1	0	0.592
1	1	0.554
1	2	0.397
2	0	0.614
2	1	0.560
2	2	0.405
3	0	0.842
3	1	0.713
3	2	0.560

Wyers, Colin (2009). The best run estimator. In Dave Studenmund (Producer), *The Hardball Times Baseball Annual* (pages 209-215). Skokie, IL: Acta Sports.

It is customary to compare specific methods for evaluating offense, but most of them are of little value because they are limited to a given period of seasons and thus biased towards those methods that were designed in the context of those seasons. A better idea is to evaluate classes of methods to see which class works better. Wyers (2009) offered a thoughtful such attempt, showing not only that but why a method such as base runs will be more accurate than runs created or extrapolated runs using a data set big enough (all Retrosheet data from 1956 through 2007) to counter the problem of formulas designed for a specific sample of years.

Wyers, Colin (2009). When is a fly ball a line drive? https://tht.fangraphs.com/when-is-a-fly-ball-a-line-drive/

It has become clear that the height of press boxes has caused variation in coder judgments concerning whether batted balls to the outfield were flies or liners. Colin

Wyers (2009b), using Retrosheet data from 2005 to 2009 for visiting teams excluding pitcher at bats, noted a correlation of .16 between press box height and line drive rate (I think as a proportion of liners plus flies), and jumps to .38 with the exclusion of the five most extreme parks in either direction, in which coders acted as if they were compensating for the problem. As the difference in overall odds of making plays on each are so great, the resulting ratings for outfielders have probably been significantly affected.

Wyner, Adi (2021). Is the 3rd time through the order effect real? Correcting for lineup order and pitcher quality selection bias. SABR Analytics Conference

Using 2010-2019 Retrosheet data, Adi Wyner (2021) reported that there was no third time through the order effect as such, as after a “settling in” process over the first few PAs, performance decrements on average begin with the first time through the bottom third of the order and continue steadily thereafter.

Zardkoohi, Asghar, Michael W. Putsay, Albert Cannella, and Michael Holmes (n.d.). Cognitive biases in decision making: Risk taking in major league baseball, 1985-1992. College Station: Texas: Department of Management, Mays Business School, Texas A & M University.

Base stealing is a one-run strategy, and as such the attempted steal should be used late in games and, in general, in high leverage situations. However, Zardkoohi, Putsay, Cannella, and Holmes (n.d.) analyzed Retrosheet data from 1985 through 1992 of more than 200,000 situations with a runner on first only and concluded that steal attempts were actually more numerous earlier in games rather than later and increased as run differentials increase from three runs behind through tied scores to three runs ahead. The authors relate this to psychological tendencies to be risky about positive things and cautious about negative things (see work on prospect theory by psychologists Amos Tversky and Daniel Kahneman, the latter a Nobel prize winner in Economics as a result of this and analogous work), such that managers are more likely to feel comfortable risking a steal when ahead than behind and when there are plenty of innings left to overcome a caught stealing then when innings are running out. Zardkoohi et al. also noted more attempts against righthanded pitchers and when there had been previous success against the same pitcher or catcher, none of which are surprising.

Zhang, Xing, Tat Y. Chan, and William P. Bottom (2022). Relational aspects of vicarious retribution: Evidence from professional baseball. *Journal of Applied* Psychology, Vol. 107 No. 6, pages 917-931

There have been several psychological studies of strategic (rather than accidental) hit by pitches as indicators of aggression, and a few specifically directed toward studying pairs of initiating HBPs followed by retaliatory HBPs as a cycle of provocation and score-settling retribution. This one is a sophisticated example of the latter set. The authors' work was based on a sample of about 20,000 HBPs from 1991-2010 games posted on Retrosheet. The data generally supported this perspective, as the occurrence of an initialting HBP increased the odds of a retaliatory one, on average an inning later, after which the odds of additional ones decreased (two retaliatory HBPs in response to the initial one were relatively rare. The authors also looked for the impact of possible measures of similarity affecting the cycle. Retaliation was more likely if the hit batter and teammate-pitcher were both from outside the U.S., and less likely if the teammate-pitcher and initiating team batter were both from outside the U.S., had been teammates in the past. In addition, retaliation was less likely if both had not attended college whereas college attendance for both actually increased retaliation; these two findings make no sense to me. Some results for control variables are also of interest, as retaliation more likely the greater the score difference, the better the retaliatory team's winning average, the retaliatory team being at home, being on more diverse teams, and in the American League. Perhaps unknown to these authors, some of these latter findings replicate earlier work; in particular, the A.L. one appears to be due to the presence of the DH leading to pitchers knowing that they will not bat and so won't be direct victims of retaliation.

Zimmerman, Jeff (2009). What factors have an effect on runs scored at MLB parks? Part 2. https://www.beyondtheboxscore.com/2009/1/7/713479/what-factors-have-an-effec

2006-2008 data. In part 1, Jeff used runs scored per game to represent park factors, but after a lot of comments switched to a method previously used by Brandon Heipp aka Patriot. This was the dependent measure in a multiple regression where the independent variables are possible factors affecting it. Rather than showing the (confusing) equation, here is his interpretation of what its results mean:

Factor	Change in Park Factor	Change Runs Scored per game (9.54 runs per
10 degree F increase	0.0077	0.073
Increase in RH by 10%	0.0120	-0.115
10,000 sq ft increase in foul area	-0.0061	-0.058
Surface is Turf	0.0090	0.085
1000 ft increase in elevation	0.0206	0.196
1 Errors for Away Team	0.0160	0.150
10 ft increase in LF	-0.0100	-0.095
10 ft increase in LC	-0.0063	-0.060
10 ft increase in CF	-0.0101	-0.096
10 ft increase in RC	-0.0020	-0.019
10 ft increase in RF	0.0106	0.101