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

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

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

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

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

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

Albert, Jim (2001). Using play-by-play baseball data to develop a better measure of batting performance. Retrieved from www.math.bgsu.edu/~albert/papers/rating_paper2

Jim Albert's work here is an analysis of Gary Skoog's (RIP) *Value Added* approach to measuring offensive production describe in Bill James's 1987 *Baseball* Abstract. He used Pete Palmer's run expectancy table as its basis, but the method would work just as well with an alternative. Basically, one takes the run potential at the end of a plate appearance, subtracts from it the run potential at the end of the plate appearance, and adds any runs that scored during the PA. If the result is positive, the player has contributed to run scoring, and if it is negative, the player has damaged run scoring. Each inning, the lead-off hitter is changed with .454 for before the PA, which is the mean run potential for no baserunners/no outs. The batter making the third out in an inning ends with a 0, meaning that they cannot have a positive contribution unless a run scored during the inning-ending event. It is important to remember that one cannot simply use the run potential at the beginning of a plate appearance when making these calculations, because various events can occur during a PA that change the base-out situation (SB, CS, WP, PB, Balk). Instead, one must use the run potential just before the "event" (e.g., walk, hit, out) that ends the PA. Stolen bases and caught stealing are credited to the baserunner. Getting on base on an error is not credited to the batter. The batter does get credit for baserunners getting extra bases on hits (e.g., first to third on a single), which Skoog was not comfortable with and invited discussion by interested analysts. Jim Albert (2001) recreated the Skoog method using 1987 National League data gathered by Project Scoresheet and available at Retrosheet, used it to estimate team run scoring per game, and then compared those estimates to actual team runs per game using the root mean square error (RMSE) as a goodness of fit measure. Its RMSE was .067, compared to .121 for Batting Runs, .202 for Bill James's Runs Created (described later), .212 for OPS, and .242 for OBA.

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

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

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

In another such analysis using Retrosheet data, Derek Bain (2018) presented BA, SA, and HR/AB for at bats ending on every count plus overall figures between 1998 and

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

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

Beyond base-out situation, the risk of attempting a steal (along with other speed-related moves such as taking extra bases on hits) depends on the specific abilities of the player making the attempt. Obviously, some players are better basestealers and/or baserunners than others, and the risk is lower the better the player is on the basepaths. Through a simulation based on the "team based on a given player" method for evaluating offense and using 2007-2009 Retrosheet data, Baumer, Piette and Null (2012) examined the expected outcomes of such attempts for 21 players purposely chosen for their variety of capabilities as hitters and baserunners. Their results suggest that taking the risk of the steal or extra base is more productive long-term to the extent that the player is a good baserunner and a less productive hitter. This is because the cost of an out on the attempt is unsuccessful is greater for a better hitter than a poorer one. Although they interpret this in the context of the chosen individual players, the real implication is that attempting the steal or extra base makes more sense when the next batter is weak, as that next batter could use the help of the extra base for driving the baserunner in.

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

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

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

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

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

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

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

There is not much evidence that a bad first inning is indicative of an off-day for a pitcher, such that the manager should pull him quickly and tax his bullpen for the rest of the game. Phil Birnbaum (200c), using Retrosheet data from 1979 to 1990, examined the subsequent performance of starters giving up three, four, and five first-inning runs. Overall, starters averaged an RC/27 (see the Batting Evaluation chapter for that) of 4.30. Starters who gave up three first-inning runs averaged an RC/27 of 4.51 for the rest of the game; but their overall RC/27 for the season was almost the same, 4.54. In other words, they were not having a particularly bad game for them as overall they were somewhat worse pitchers than average. The same for four runs in the first; 4.56 the rest of the game, 4.57 overall. In contrast, five runs might be an indication; 5.58 the rest of the game versus 4.67 overall. However, Phil warns us of some potential problem with this data. First, the multiple-run innings are included in the seasonal figure but not the after-the-first innings. If the multiple run innings were subtracted from the overall, as they really should be in this study. it might be noticeably lower than this study's findings and from the after-the-first performance. Second, some pitchers are removed after the first and so are not represented in the after-the-first data, and these might just be the pitchers who really are having an off-day which is recognized as such by the manager or pitching coach.

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

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

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

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

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

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

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

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

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

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

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

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

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

Bradbury, John Charles and Douglas Drinen (2006). The designated hitter, moral hazard, and hit batters. *Journal of Sports Economics,* Vol. 7 No. 3, pages 319-329.
Bradbury, John Charles and Douglas J. Drinen (2007). Crime and punishment in major league baseball: The case of the designated hitter and hit batters. *Economic Inquiry*, Vol. 45 No. 1, pages 131-144.
Baldini, Kevin, Mark T. Gillis and Mark E. Ryan (2011). Do relief pitching and remaining gams create moral hazard problems in major league baseball? *Journal of Sports Economics,* Vol. 12 No. 6, pages 647-659.

There is a surprisingly large literature on whether hit-by-pitches are the result of strategic choice on the part of the pitcher and manager of the opposing team. The impetus of this work was the substantial increase in HBP in the American League after the appearance of the designated hitter, implying that pitchers may be more willing to hit someone when retaliation against them personally will not occur. An alternative hypothesis has been that when retaliating, pitchers are more likely to throw at good batters than poor because the former are more likely to get on base anyway, so pitchers, as generally the poorest hitters on a team, are the least likely targets. Bradbury and Drinen performed two studies that provided far better examinations of the retaliation hypothesis than those previous through use of Retrosheet 1973-2003 data. Based on game-by-game information, they first (2006) noted evidence for both hypotheses in predictive model allowing for determination of the order of importance of associated variables. The variable most strongly associated with hit-by-pitches was whether the game had designated hitters, with this effect occurred in interleague games including NL teams, evidence against the idea that HBPs are just idiosyncratic to the AL but perhaps due to pitchers not batting. However, the difference between leagues largely disappeared in the 1990s. On the other side of the dispute, the second most associated variable was total runs scored, evidence that when teams are hitting well the other side finds less reason not to hit batters. Further, home runs by the other team were also associated, more evidence that a HBP against a powerful batter would be considered less harmful. Finally, and not surprisingly general pitcher wildness was also correlated. In their second (2007) paper, Bradbury and Drinen determined whether a hit-by-pitch in one half inning increases the odds of retaliation in the next. According to two analyses, one for 1969 combined with 1972 through 1974, the other for 1989 through 1992, it does, as does a home run by the previous batter in the more recent data set; both of these findings support the retaliation hypothesis. Consistently with the second argument, higher OPS was positively associated with HBP whereas pitchers were less likely to be plunked than everyone else; both of these results suggest the "less harm" hypothesis. In addition, large score differentials increase HBP, likely because there is less harm when such a differential leaves less doubt concerning which team will probably win the game. Again, wilder pitchers are, not surprisingly, more likely to hit batters.

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

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

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

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

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

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

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

John Charles Bradbury (2019) used 2000 to 2009 Retrosheet data to examine the impact of QuesTec on ball/strike calls. In short, 11 ballparks were equipped with QuesTec systems between 2001 and 2008 that allowed for the evaluation of home plate umpire calls. In short, the ballparks with QuesTec had a smaller proportion of called strikes than the ballparks without it, to the tune of .016 per PA or .81 per game on average. This impact was overwhelmed by other factors, most notably a directive to umpires to be more accurate, leading to the called strike rate to increase by two percent between 2000 and 2001 (the year of the directive) and another ½ percent in subsequent seasons. As for the effect of control variables: Consistent with past research, there were fewer called strikes for home team batters, which is part of one of the research-supported explanations for home team advantage, crowd noise; yet more called strikes due to the attendance/home team batter interaction, which is inconsistent with that explanation. In addition, there was deference for experienced batters and pitchers (consistent with past work) and more called strikes for catchers (inconsistent with the literature).

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

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

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

Bruschke (2012) offered a fielding metric based on a completely different logic than zone approaches. In his own words, “In a nutshell, zone approaches carefully measure individual performance, but estimate productivity [by that, he means total team success at saving runs via fielding). My approach measures productivity directly but estimates individual performance” (page 14). He called it Fielding Shares, and that is an apt title, as, analogously with Bill James’s Win Shares, it begins with team performance and divides it among the players responsible for it.

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

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

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

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

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

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

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

The odds of a called strike if no swing

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

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

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

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

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

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

As he has explicitly stated both online and in print (see the comment below about his 2017 book), Russell Carleton is indebted to Retrosheet for much of the data used in his many studies. My guess is that he has used it in most of them, although he has not been explicit in stating as such. Rather than including all of them, which would make this document much longer, I will describe those that he could not have performed with it.

Carleton, Russell (2007). Is walk the opposite of strikeout? *Baseball by the Numbers,* Vol. 17 No. 1, pages 3-9.

Carleton (2007) performed a very interesting (if through no fault of the author) flawed study concerning the concept of plate discipline, which I can only describe in brief. We often measure discipline through looking at the ratio of walks to strikeouts, but this ratio conflates two different capabilities: the ability to recognize which pitches to swing at and which to take, and the ability to put a ball in play (or homer, which to simplify Carleton's argument I will include in that category) given the decision to swing. Carleton attempted to get at these abilities using what data was available: Retrosheet data from 1993 through 1998 for every player season with more than 100 plate appearances (2426 in all), allowing him to distinguish balls, called and swinging strikes, foul balls, and balls hit in play. Following from signal detection theory Carleton computed a measure of “sensitivity” operationally defined as the proportion of strikes swung at that were put into play minus the proportion of pitches that should not have been swung at (those swing at and missed plus pitches that were called balls) that were swung at and missed. The idea was that the former represented pitches that should have swung at and the latter those that should have been taken, so the larger the number the more sensitive the batter for when swinging was a good idea. In short, this measures knowing when to swing and when not to. The second, “response bias,” consisted of the proportion of balls that should have been swung at that were hit (versus swung at and missed) paired with the proportion of balls that should have been taken and were (versus called strikes). The notion here is to measure how often batters swing in the first place. Players could be very high in this measure (swing too often) or very low (not swing enough). See the article for details, including how Carleton handled foul balls.

These two measures had a very small statistic relationship in the data and so measured different things. Both were also consistent over time for players (intraclass correlations of .72 for sensitivity and .81 for response), implying they are real skills. Both correlated about .5 with strikeout/walk ratio, again implying two differing but significant skills, and sensitivity correlated .22 with age, meaning that players improvement their judgment with experience. Carleton listed some players that were very high and very low in both. Vladimir Guerrero was an interesting case, as he was the most sensitive (as he made contact when he swung more than others) but had the worst response bias in the direction of swinging too often. Scott Hatteberg had the worst response bias in terms of not swinging enough.

Finally, Carleton examined how his measures predicted strikeout and walk rates in stepwise multiple regression equations. Strikeout rate was decreased by contact rate, “good decision rate” (the ratio of pitches that were either taken or into play), and surprisingly swing percentage, and again surprisingly increased by two-strike fouls (apparently giving the pitcher another chance to strike the batter out). Walk rate was decreased by the first three and decreased by the latter.

I said above that there is a flaw here that was not the author’s fault. The real measure we would want of sensitivity would be to compare pitches in the strike zone that were swung at versus taken for strikes with pitches outside of the strike zone that were taken for balls versus swung at. Retrosheet does not have data on where pitches were that were swung at, limiting Carleton’s options in this regard.

Carleton, Russell (2007). Do you have any idea how fast you were going? *By the Numbers,* Vol. 17 No. 2, pages 8-11.

Bill James’s Speed Score included six variables: stolen base attempts from first and success rate, triples per opportunity, runs scored per opportunity, grounded into double plays per opportunity, and a defensive indicator combining position and range factor. Retrosheet's availability allowed Russell Carleton (2007a) to use the following alternative indicators for speed:

1 – infield hits per ground ball

2 – times on first in which the pitcher threw there to hold the runner

3, 4, and 5 – extra bases on hits to the outfield; Russell distinguished among the three major possibilities (as in the research just described) rather than grouping them together

6 – triples divided by (triples + doubles); in other words, the ability to stretch extra base hits

7 – beating out attempts at ground ball double plays after force outs at second

Russell then converted the data for each of these seven indicators plus two used by Bill (stolen base attempts from first and success rate) to make them more amenable for analysis (for the stat savvy; took the natural log to approximately normalize the distribution and then turned the result into z scores). He then performed a form of factor analysis (principal components with varimax rotation).

> Those familiar with factor analysis can skip this paragraph. Factor analysis groups together variables that correlate with one another (and shows how well they intercorrelate with indices called "factor loadings) and differentiates groups that do not correlate with one another. The example I used when I taught was something like this: imagine the answers to a survey asking people how much they like various types of junk food. Pretzels, popcorn, and chips might form one factor; candy, cake, and pie a second factor. The first factor indicates salty options and the second sweet options. Variables can "cross load" and appear in both factors: chocolate covered pretzels perhaps.

Two factors emerged. The first included six of the nine, all except the three extra-bases-on-hits variables, which Russell took as indicating speed. The second included those three plus times on first drawing throws and attempting steals: Russell interpreted as representing motivation to get extra bases beyond hits and walks and called it "green light." Bill's original index correlated .807 with the speed factor and .718 with the green light factor, which implies in particular that two methods for measuring speed are fairly close to interchangeable. Finally, Russell surmised that players' speed score minus their green light score demonstrates their baserunning riskiness. If the former is much higher than the latter, yielding a positive number after subtraction, then the player might not be taking advantage of speed as much as they could. If the former is much lower than the latter, generating a negative number, then the player is taking more chances than they ought.

Carleton, Russell A. (2008). The foul ball, part one: What does it tell us about a batter? https://statspeakmvn.wordpress.com/2008/04/15/the-foul-ball-part-one-what-does-it-tell-us-about-a-batter/

Is hitting foul balls a skill? Russell Carleton attempted to find out. Based on Retrosheet data from 2004 to 2007 including seasons in which players had 250 or more PA, Russell distinguished between fouls per plate appearance, percentage of pitches fouled off (which differs because different batters will face a differing number of average pitches per PA), and percentage of batted balls that went foul (which differs again because different batters have differing contact rates). The intraclass correlation for foul balls per PA was .574, and those for percentage of pitches fouled off and percentage of fouls per batted balls were both over .6. So it appears from this that foul ball hitting is a skill. But this appearance is deceiving, as it does not distinguish between foul balls hit with zero and one strike, which add a strike, from those with two strikes, which do not. Expanding this and subsequent analyses to 2000-2007 Retrosheet data for seasons in which batters had at least 250 PA, the two are only correlated at .106.

And the two appear to function differently. Two strike fouls correlated .150 with the overall fouls/pitch measure and .524 with contact rate. So the two strike foul hitter seems to be trying not to strike out. And he was less likely to strike out (correlation = –.482) but also to walk (correlation = –.345). So he is trying to put the ball in play, and he is successful (correlation with singles = .347) while sacrificing power (correlation with homers = –.215 and with homers per fly ball –.300). In short, he is a contact hitter (contact rate correlated .549 with singles and –.521 with homers). In contrast, one and two strike fouls correlated .487 with the overall fouls/pitch measure and with overall batter contact rate with –.366. So the zero/one strike foul ball hitter is low on contact and has problems keeping balls fair. They struck out (correlation = .669) and homered (correlation = .410) more and singled (correlation = –454) less. Further, the overall measure, which represents the zero/one strike foul hitter a lot more closely than the two strike foul hitter, correlated .297 with fly ball rate and –.318 with ground ball rate, additional if indirect evidence of selling out for power.

Carleton, Russell A. (2008). The foul ball, part three: What does it tell us about an at-bat? https://statspeakmvn.wordpress.com/2008/04/30/the-foul-ball-part-three-what-does-it-tell-us-about-an-at-bat/

Based on 2000-2007 Retrosheet data for plate appearances between batters having a pitchers facing at least 250 PA that season, Russell Carleton examined the final OBA for each count if the next pitch were each of the three ways in which a strike can occur in PAs with no or one strike:

Count	Swinging	Called	Foul Ball	Count	Swinging	Called	Foul Ball
0-0	.263	.287	.295	0-1	.199	.219	.233
1-0	.308	.321	.329	1-1	.227	.248	.256
2-0	.397	.404	.407	2-1	.287	.315	.322
3-0	.585	.596	.597	3-1	.442	.458	.486

So in general, foul balls are signaling the best and swinging strikes the worst eventual outcome. Now, the same sort of comparison would not make sense for two strikes, as anything but a foul results in OBA = .000, so here Russell looked at the outcome from the four two-strike counts for different numbers of subsequent fouls during the rest of the PA (not distinguishing between fouls if a ball was called between fouls in the PA, although Russell claimed that the findings were about the same with that distinction made):

Count	No fouls	One foul	Two fouls	Three or more fouls
0-2	.209	.264	.231	.253
1-2	.235	.266	.279	.282
2-2	.307	.313	.314	.312
3-2	.468	.467	.451	.482

So when behind in the count, hitting at least one foul is a good sign for the batter, but when even or ahead it doesn't seem to matter, not does the number of fouls hit (which Russell points out contradicts the myth that a lot of fouls constitutes a "good at bat" at least in terms of the relevant batter).

Carleton, Russell A. (2008). On the reliability of defensive abilities, part 1. https://statspeakmvn.wordpress.com/2008/07/21/on-the-reliability-of-defensive-abilities-part-1/

Russell A. Carleton (2008) used 2004 to 2007 data (most certainly from Retrosheet) to calculate a slew of fielding reliability figures, using the intraclass correlation (which is a combination of the correlations for each fielder in the data set year by year). Here are some of them:

	Grounders				Liners	Popups
	Range	Throw	DPs	Total	Total	Total
First Base	.509	.163	X	.423	.117	.037
Second base	.224	.405	.385	.543	.036	.123
Shortstop	.507	.277	.151	.418	.239	.182
Third Base	.298	.322	x	.510	.193	.050

	Fly Balls		Grounders	Line Drives		Throwing	
	Total	Cut	Cut	Range	Cut	OPA	XBP
Left Field	.438	.149	.319	.162	.187	.226	.028
Center Field	.272	.130	.303	.179	.158	.036	.314
Right Field	.219	.133	.284	.046	.005	.257	.291

I will provide as much explanation as I can, given that Russell (at the beginning of his career as an online contributor) was not at all clear about what some of these indicate. I am guessing that "cut" means cutting off hits that are flies, grounders, or liners. OPA is Russell's fielding metric (Outs Per Attempt Above Average), about which I have not been able to find much, is I believe concerned with throwing out baserunners whereas XBP (which stands for extra base prevented) is about limiting extra bases by baserunners on hits. Anyway, note that overall reliability was pretty good for infield grounders but not for much of anything else. How much of this is inconsistency in fielder performance from year to year or in the codes assigned by (basically untrained) the many Project Scoresheet volunteer scorers (of which I was a proud participant).

Carleton, Russell A. (2008). On the reliability of defensive abilities, part 2. https://statspeakmvn.wordpress.com/2008/07/21/on-the-reliability-of-defensive-abilities-part-2/

Russell Carleton (2008) uncovered the following correlations in his OPA (Outs Per Attempt Above Average) fielding metric across infield positions:

	Second Base	Shortstop	Third Base
First Base	.278	.101	.347
Second Base		.528	.504
Shortstop			.434

The implication is that (not including first base) infielders can trade positions to some extent, which as Russell pointed out is a necessary skill for the utilityman.

Carleton, Russell A. (2008) Is Brian Bannister on to something? https://statspeakmvn.wordpress.com/2008/01/31/is-brian-bannister-on-to-something/

Russell Carleton (2008), based on a suggestion by Brian Bannister (one of the first major leaguers to take sabermetrics seriously) that batters are more likely to make bad decisions and take weak swings in pitchers' counts, used Retrosheet data to examine batting average on balls in play between 2003 and 2006 at the count at which it occurred, and came up with the following:

Count	BABIP	Count	BABIP	Count	BABIP	Count	BABIP
0-0	.2965	1-0	.3027	2-0	.3045	3-0	.3112
0-1	.2908	1-1	.2978	2-1	.3053	3-1	.3119
0-2	.2856	1-2	.2908	2-2	.2932	3-2	.3066

Consistently with Bannister's conjecture, it is pretty obvious that batters do better when the count is in their favor. The consistency for these from year to year was generally in the mid .30s, which isn't great but is high enough to be taken seriously, and Russell noted that those for the two most favorable pitcher's counts were higher (0-2, .51; 1-2, .41).

Carleton, Russell A. (2010). Why are games so long? https://www.baseballprospectus.com/news/article/10753/baseball-therapy-why-are-games-so-long/

Based on 2009 games, Russell Carleton (2010) used a method called stepwise regression, which discerns the order of importance for variables associated with the measure of interest, and determined that the number of pitches thrown was easily the most important of these factors, making up 82.3 percent of the accounted-for-variance (Russell did not tell us what proportion of total variance was accounted for). Next in line, adding 4.8 percent of variance accounted for, were mid-inning pitching changes (with an average of 2.06 per game each adding about three minutes on average) and throws to first (7.28 per game each responsible for about 40 seconds). Other significant predictors worth another 2.1 percent of accounted-for variance, were intentional walks (no longer a factor), plate appearances over and above pitches thrown, stolen base attempts, breaks between innings given that rain-shortened games have fewer and extra-inning games more. Number of walks and strikeouts were <u>not</u> predictors, and an increase in balls in play and home runs <u>de</u>creased game time; putting these four together, the impact of the first two were probably included in the all-important number of pitches, and as Russell noted the last two likely shortened the typically plate appearance and so cut down on that number.

Carleton, Russell A. (2010). Credit where it's due, part 1 https://www.baseballprospectus.com/news/article/10387/baseball-therapy-credit-where-its-due-part-1/
Carleton, Russell A. (2010). Credit where it's due, part 2. https://www.baseballprospectus.com/news/article/10533/baseball-therapy-credit-where-its-due-part-2/

One of the problems with Voros McCracken's claim that pitchers have little control over whether batted balls become hits is that his method presumes that there is no difference across pitchers in the strikeout/walk/home run tendencies of the batters they happen to face, which is akin to say that all batters have the same strikeout/walk/home run rate. Russell Carleton (2010c), based on all PA from 1993-2009 data excluding those by pitchers and those ending with intentional walks, Russell used a regression technique (logistic) designed for examining binary variables, those with only two values, in this case strikeout versus no strikeout, with each pitcher's, batter's, and league overall strikeout rate as predictors. The equation only accounted for a paltry 6 percent of the variance in strikeouts, imply that those three factors are superseded by situational influences in importance. Nevertheless, the fact that 56 percent of that 6 percent was batter effect and only 43.3 percent pitcher effect (the league received the remaining 0.7%) means that individual batter strikeout tendencies are actually more important than pitcher's. Analogously (Carleton, 2010d), batters got 63.3 percent of the accountable credit for walks and 62.2 percent for hit by pitches, with pitchers receiving 35.8 percent and 36.6 percent respectively (Russell did not include how much of the total BB and HBP variance accounted for by these last two True Outcomes).

What happens with a batted ball in play is more complicated, because now you have the fielder's ability to contend with. Russell used a couple of examples to describe his method of analysis, which considers the impact of different results in different base-out-inning-score differential situations on win probability for each team. For instance, for a ground ball toward second in a tie game with one out in the sixth inning and a runner on first, the most likely results are double play, put out at first, fielder's choice at second, single with runner going to second, and single with runner going to third. Taking everything in, the batter receives 52.6 percent of the accountable variance for the outcome, the pitcher 43 percent, and the second baseman 3.8 percent; in other words, whether the pitcher is more than ten times as responsible as the fielder concerning whether the grounder becomes a hit or an out. Further, complicating the picture even more, if the ball gets through for a single toward the right fielder, whether the baserunner on first makes it third is 39.4 percent pitcher, 26.2 percent baserunner, 14 percent right fielders, and 9.2 percent batter. In short, the pitcher has a lot of responsibility for the outcome of batted balls.

Carleton, Russell A. (2012). It's a small sample size after all. https://www.baseballprospectus.com/news/article/17659/baseball-therapy-its-a-small-sample-size-after-all/

Based on Retrosheet data for the 311 batters with at least 2000 PA from 2003 to 2011, the following table Russell Carleton (2012) computed indicates when the sample size of data for a particular index reaches an estimated .70 reliability figure (where as he out it the signal-to-noise ratio reaches 50/50; see the original for his method).

Statistic	Definition	Stabilized at	Notes
Strikeout rate	K / PA	60 PA	
Walk rate	BB / PA	120 PA	IBB's not included
HBP rate	HBP / PA	240 PA	
Single rate	1B / PA	290 PA	
XBH rate	(2B + 3B) / PA	1610 PA	
HR rate	HR / PA	170 PA	
AVG	H / AB	910 AB	Min 2000 ABs
OBP	(H + HBP + BB) / PA	460 PA	

SLG	(1B + 2 * 2B + 3 * 3B + 4 * HR) / AB	320 AB	Min 2000 ABs,
ISO	(2B + 2 * 3B + 3 * HR) / AB	160 AB	Min 2000 ABs
GB rate	GB / balls in play	80 BIP	Min 1000 BIP, Retrosheet classifications used
FB rate	(FB + PU) / balls in play	80 BIP	Min 1000 BIP including HR
LD rate	LD / balls in play	600 BIP	Min 1000 BIP including HR
HR per FB	HR / FB	50 FBs	Min 500 FB
BABIP	Hits / BIP	820 BIP	Min 1000 BIP, HR not included

Carleton, Russell A. (2012). One-run winners: Good or lucky? https://www.baseballprospectus.com/news/article/18151/baseball-therapy-one-run-winners-good-or-lucky/

Russell Carleton (2012) broke down the ways in which one-run games can occur, with about half entering the ninth inning with the eventual winner ahead by one-run (most of which had scoreless ninths but a few of which featured each team scoring the same number of runs that inning, so that for example a 4-3 game ended up 6-5), about a quarter tied after the eighth and someone scoring a run in the ninth, 14 percent with the eventual winner ahead by more than one run but the loser making it closer in the ninth, and 11 percent in which the eventual winner was behind after eight but pulled off a successful come-from-behind ninth inning rally. Anyway, the winning average of home teams in games decided by one run between 1993 and 2011 was 61 percent, which is considerably better than the 53-54 percent norm. The main reason for this appears to be the following bias: If in a tied game, the visiting team scores a run, it will play the full inning and could add several more runs. If in a tied game, the home team score a run, the game is over and they don't have the need to score more. For this reason, the home team has a greater "opportunity" to win by one run. Looking specifically at games tied going into the ninth between 1993 and 2011 organized in 40-game blocks for each team (the typical number of one-run games a team plays in a season), the reliability coefficient for team winning average in those games (measured as consistency among the blocks) was .17. In other words, there is little evidence that winning by one run is a repeatable team skill.

Carleton, Russell A. (2012). Are Three-True-Outcomes players better in the playoffs? https://www.baseballprospectus.com/news/article/18722/baseball-therapy-are-three-true-outcomes-players-better-in-the-playoffs/

Based on 1993 through 2011 (almost certainly Retrosheet) data, Russell estimated the performance of hitters in the playoffs given how they did at the level of individual plate appearances during the regular season and categorized them by the proportion of their plate appearances that ended in one of the Three True Outcomes. The following is the predicted playoff figures for the overall average hitter versus the overall average pitcher given three different TTO proportions:

TTO percentage	K	BB	HBP	1B	2B/3B	HR	OIP
20%	.182	.073	.011	.159	.047	.025	.504
30%	.184	.077	.011	.142	.045	.029	.481
40%	.186	.081	.011	.126	.043	.035	.457

TTO percentage	Fly Ball	Line Drive	Grounder
20%	.349	.174	.468
30%	.360	.182	.452
40%	.371	.190	.437

The implication of all this is that high TTO players are relatively more likely than low TTO players generating the same overall production to hit flies and liners at the expense of grounders, resulting in more homers and fewer singles, and walk more in the playoffs than during the regular season, all else being equal.

Carleton, Russell A. (2013), What is a good pitching coach worth?
https://www.baseballprospectus.com/news/article/20317/baseball-therapy-what-is-a-good-pitching-coach-worth/
Carleton, Russell A. (2013). What is a good hitting coach worth?
https://www.baseballprospectus.com/news/article/20474/baseball-therapy-what-is-a-good-hitting-coach-worth/

Based on 1993 to 2012 Retrosheet data for pitcher-seasons with at least 250 batters faced and coaches with at least 10 of those pitcher-seasons under his belt (sample size of 80), and with the proper controls for player quality, home field and league in place, Russell A. Carleton (2013) estimated that a good pitching could maintain his team's pitcher's strikeout rate by as much as 2½ percent, walk rate by up to 1 percent, and home run rate maybe one-half of a percent over the average pitching coach, and a poor one about the same worse than average. This translates to the best saving their staff and the worst costing their staff about two-fifths of a run in FIP. As Russell admits, these conclusions are confounded by potential impacts of the team's manager on the staff and the pitchers on one another. For batting coaches using an analogous sample, the difference plus or minus was about 2 percent for strikeout rate and 1 percent for walk and homer rates. Interestingly, the impact of batting coach impact on singles hitting correlated at –.409 with strikeouts and –.441 with walks; those for outs on balls in play with strikeouts at

–.730, walks at –.535, and homers at –.426. These associations imply that some batting coaches preach a risk-free contact-heavy approach and others a more aggressive stance. Despite this, there was no evidence of pure Three True Outcomes philosophies as the relevant correlations were .290 (walks and strikeouts), .137 (homers and strikeouts), and .101 (homers and walks). Overall, batting coaches could be worth a couple of wins a year either won or lost.

Carleton, Russell A. (2013). You gotta keep 'em separated. https://www.baseballprospectus.com/news/article/19907/baseball-therapy-you-gotta-keep-em-separated/

Based on 2003 to 2012 Retrosheet data including all batters with and all starting pitchers facing 250 PAs in a season, Russell Carleton (2013) uncovered no evidence that batters facing consecutive night starters who were similar in regard to handedness and tendencies for power versus finesse and groundball versus flyball performed any better than when facing dissimilar pitchers, even when these three factors were combined (e.g., two straight days facing lefty finesse groundballers). So there is no evidence supporting the myth that you need to keep similar starters separated.

Carleton, Russell A. (2013). Can't buy me chemistry? https://www.baseballprospectus.com/news/article/19704/baseball-therapy-cant-buy-me-chemistry/

Using as a defining characteristic the percentage of players on a team one season who played at least 20 games for the same team the previous year, Russell Carleton (2013f) examined whether outcomes for hitters with at least 250 PA differed in two consecutive seasons when the hitter in question either stayed with their previous team or moved to a new team, when the team(s) in question either had a lot or a little turnover. There was some impact. For one example, I cut and pasted Russell's chart for home run rate:

	High Turnover	Low Turnover
Player was here last year	2.46%	2.59%
Player was not here last year	2.72%	2.03%

For another, here would be the impact for a hitter who ended 50 percent of his plate plate appearances with outs in play:

	High Turnover	Low Turnover
Player was here last year	50.53%	49.85%
Player was not here last year	50.09%	50.89%

In contrast, there was no analogous effects either for pitchers facing 250 batters two consecutive seasons or for teams as a whole when compared with PECOTA preseason projections, either overall or for close games.

Carleton, Russell A. (2013). On the evolution of the patient hitter. https://www.baseballprospectus.com/news/article/20399/baseball-therapy-on-the-evolution-of-the-patient-hitter/
Carleton, Russell A. (2014). There's gotta be a reason for the strikeout epidemic … right? https://www.foxsports.com/mlb/story/there-s-gotta-be-a-reason-for-the-strikeout-epidemic-right-052714

These two online articles used data from Retrosheet to evaluate a number of proposed explanations for the increase in strikeouts between 1993 and 2012 or 2013 (depending on the article). Most of the numbers cited below are my guestimates from diagrams Russell used to display his findings.
Possible reason #1: Pitchers are getting better. Based on the idea that better pitchers were replacing worse ones as the years passed, Russell compared strikeout rates for pitchers in their last season one year with pitchers in their first season the next year, starting with pitchers whose careers ended in 1993 with pitchers whose careers began in 1994 (and every subsequent two-year stretch ending with 2012 and 2013). Overall there was very little difference between the two groups, maybe .02 per season on average. So there is no good evidence from here that pitchers were getting better during the relevant two decades. However, this certainly cannot be taken as a definitive test, as the underlying assumption that the first and last years of their careers adequately represent pitchers' talent is questionable.
Possible reason #2: Batters are more prone to strikeouts. Russell tested this analogously, comparing strikeout rate for batters in their last season one year and in their first season next year. The effect was actually negative .04 for 1993-1994, but topped 0 in 2000-2001 and continued rising to about a .07 increase 2012-2013. So there was some evidence in favor of this proposal, although again the underlying assumption can be questioned.
Possible reason #3: Batters are selling out for power more often. Slugging average on pitches in which the batter made contact did rise from about .465 in 1993 to about .53 in 2001 but not afterward, fluctuating around about .Slugging average on 3-0 counts, when batters can be presumed to be swinging hard, fluctuated quite a bit from year to year around .75 but revealed no up or down trend. So no evidence for this proposal.
Possible reason #4: Batters became more patient, swinging less, which leads them more susceptible to. Russell included a lot of relevant evidence across the two articles. First, pitches per at bat went up from about 3.64 in 1993 to 3.79 in 2012, which (relevant more to changes in starter usage patterns) translated to a drop from 18.33 batters faced per start in 1993 to 17.67 in 2012. In particular, batters swung less often on the first pitch over time, decreasing from about 30.3 percent of the time in 1993 to about 25.5

percent in 2010, although the figure actually rose about a percent over the next three years.   Pitchers seem to have noticed, because first pitch strikes went up during the interim from 49-50 percent during 1993 through 1999 to almost 55 percent in 2013. So combining these two factors, batters ended up facing 0-1 counts more often. A parallel if smaller increase occurred at 1-0 counts; swing rates down over time, called strikes went up, so 1-1 counts rose relative to 2-0 counts. There was also less swinging on 3-0 counts (13 percent in 1996, 7.1 in 2009), and although 1993 to 1995 were lower and 2010 to 2012 higher, a downward trend was obvious. So increased batter patience probably contributed.
Possible reason #5: Batter contact rate decreased. In 1993 it was about 81½ percent per swing, then decreased to about 80.2 percent in 1998, increased to almost 82 percent in 2005, but went down to maybe 79.6 percent in 2012. So evidence points to lower contact rate as a possible contributor.
Possible reason #6: When batters do make contact, there are more foul balls. With fewer than to strikes, a foul adds a strike to the count which would not have occurred with a fair ball. This means another opportunity for the pitcher to eventually strike the batter out. These did increase a bit, from at least one occurring in 42 percent of plate appearances in 1993 to 44 percent in 2012, so this could be a small contributor. With two strikes, the pitcher again has another opportunity for a whiff, although no additional strikes are added. Two strike fouls “boomeranged” in Russell’s terms; at about 40 percent of plate appearances reaching two strikes between 1993 and 1995, the rate soured to 46 percent from 2000 to 2002 but then collapsed back to about 40 percent in 2012.
Possible reason #7: Pitchers became better at putting batters away with two strikes. This is a bit complicated. Batters did swing more often on 0-2 counts; from about 43.6 percent of the time in 1995 up to maybe 48½ percent in 2013, and Russell claimed similar findings for 1-2 and 2-2 counts. But counteracting this tendency, there were fewer taken third strikes, about 10.4 percent in 1998 to fluctuating around about 7½ percent 2008-2013. Now, if I interpret the relevant diagram correctly – my interpretation is the opposite of Russell’s, so maybe I am in error here – contact rate with two-strike counts fell from about almost 77 percent in 2005 to maybe 73½ percent in 2012 and 2013. So, if I interpret that diagram correctly, this could be a contributing factor.

Also in the 2013 article, Russell noted that between 1993 and 2012, the correlation between team pitches per plate appearance and winning average was .14, which is not a lot of evidence that batters being patient leads to their team winning more games.
An additional tidbit: From 1993 to 2012, the correlation between team pitches per plate appearance and winning average was .14, so not huge evidence that being patient leads to more winning.

Carleton, Russell A. (2013). The high-pitch-count hangover.
https://www.baseballprospectus.com/news/article/21272/baseball-therapy-the-high-pitch-count-hangover/

Using 2003-2012 Retrosheet data for all starters pitching on the normal four day rest and controlling for batter and pitcher quality, and as usual in his work using only pitchers facing and batters having at least 250 plate appearances, there was a slight increase in singles and homers and decrease in outs in play due to previous pitch count, as follows:

Outcome	Expected — 100 Pitches Last Game	Expected — 110 Pitches Last Game	Expected — 140 Pitches Last Game
Single	15.32%	15.34%	15.51%
Home run	2.74%	2.77%	2.87%
Out in Play	45.86%	45.75%	45.40%

This only amounts to about three or five runs for a team-season for a 10 pitch per game increase. So the impact is on injury rate more than on performance.

Carleton, Russell A. (2013q). I thought he was gonna get it. https://www.baseballprospectus.com/news/article/21215/baseball-therapy-i-thought-he-was-gonna-get-it/

To what extent is there interdependence between pairs of adjacent infielders, such that fielding indices for one are associated with fielding indices for the other? This is a tricky question to answer, because one could argue that bias could occur in both directions. On the one hand, having an outstanding defender next to you could allow you could mean that you don't have to worry about the hole between you and him and can position yourself toward the other direction. On the other hand, having an outstanding defender next to you could cause you to be lazy and ignore balls hit in that hole ("it's his responsibility"). Bob Davis in *By The Nmbers* Vol. 3 No. 1 (1991; in References as Robert B. Davis) correlated 1988 Defensive Averages across the four infield positions at the level of team rather than individual players, and noted the following:

	First Base	Second Base	Shortstop
Second Base	.48		
Shortstop	.16	.34	
Third Base	.09	.29	.45

The correlations are all positive, and suggest that adjacent positions have a positive fielding influence on one another. But the fact that second and third are correlated almost as highly as second and short, and that first correlates a bit with short and third, suggests that there is bias in this data. And it turns out that there was. Russell Carleton (2013q), using 1993-1999 Project Scoresheet data on Retrosheet, assigned responsibility to infielders for balls hit to them and the adjacent holes (for example, shortstop territory was 6, 46, and 56) while controlled for BAB for pitcher, batter, and balls hit in each zone. Those controls were probably crucial, as, in general, there was a negative impact of good infielders on those adjacent. The strength of the effect was a drop of 1 percent for one, equivalent to 10 BABIP points, in what I presume could be taken as Defensive Average for every 5 percent improvement on the part of the other.

Carleton, Russell A. (2013). The truth about closers and extra-inning games. https://www.baseballprospectus.com/news/article/20927/baseball-therapy-the-truth-about-closers-and-extra-inning-games/

For pitchers with at least 20 saves in a season (unusually for him no description of how many or which seasons, my guess is Retrosheet data from 1993 to 2012), Russell Carleton (2013j) compared apparent pitch strategy between "official" save situations and games in which the pitcher came in a tied game in the ninth or later inning, and discovered a less risky approach in the latter; fewer home runs allowed, but also fewer strikeouts and more outs on balls in play.

Carleton, Russell A. (2013). How reliable are our fielding metrics? https://www.baseballprospectus.com/news/article/20571/baseball-therapy-how-reliable-are-our-fielding-metrics/

Russell Carleton (2013) estimated that the reliability for Project Scoresheet fielding data for ground balls (1993-1999) reached a reliability figure of .7 at 290 grounders in the first basemen's territory, 500 grounders for second base territory, 420 grounders for shortstop, and 400 for third base. For infield pops, the figures were 48,000 (!), 400, 320, and 3,240 respectively. For outfield flies, they were 370 for left fielders, 280 for center, and 210 for right.

Carleton, Russell A. (2013). What a difference a day off makes. https://www.baseballprospectus.com/news/article/21090/baseball-therapy-what-a-difference-a-day-off-makes/

Based on 2003-2013 (likely Retrosheet data), Russell Carleton (2013) determined that the number of games that a player played in the last week, two weeks, and three weeks was associated negatively with singles, doubles/triples, and homers, and positively with outs on balls in play; age did not impact on these relationships. The impact was about 1½ OBA points for a one game in a week difference, which could add up to a run or so per player so a win or so per team each season.

Carleton, Russell A. (2013). Does firing the hitting coach mid-season work? https://www.baseballprospectus.com/news/article/20885/baseball-therapy-does-firing-the-hitting-coach-mid-season-work/

For instances between 1998 through 2012 in which hitting coaches were changed during the season for batters with at least 100 PA under each (data from Retrosheet), batters overall improved to the equivalent of 10 points in OPA and 15 points in SA, summing to 25 points in OPS. However, as Russell admits, there is no way of knowing whether this improvement is due to the changing of the guard or of players randomly underperforming then returning to their normal production.

Carleton, Russell A. (2013). Should I worry about my favorite pitcher? https://www.baseballprospectus.com/news/article/20516/baseball-therapy-should-i-worry-about-my-favorite-pitcher/
Carleton, Russell A. (2008c). On the reliability of pitching stats https://statspeakmvn.wordpress.com/2008/01/06/on-the-reliability-of-pitching-stats/

Here are two studies relevant to Voros McCracken's claim that the most reliable indicators of pitching skill are strikeouts, walks, and home runs allowed, and that batting average of balls in play is mostly a matter of luck and fielding prowess. Based on Retrosheet data for pitchers facing at least 2000 batters from 2003 to 2012, the following copy-and-pasted table Russell Carleton (2013p) computed indicates when the sample size of data for a particular index reaches an estimated .70 reliability figure, usually through comparing two identically-sized stretches of plate appearances of ever-greater size until that magic reliability number was reached (see the original for method details).

Statistic	Definition	Stabilized at	Notes
Strikeout rate	K / PA	70 BF	
Walk rate	BB / PA	170 BF	IBB's not included
HBP rate	HBP / PA	640 BF	
Single rate	1B / PA	670 BF	
XBH rate	(2B + 3B) / PA	1450 BF	Estimate*
HR rate	HR / PA	1320 BF	Estimate*
AVG	H / AB	630 BF	Min 2000 AB's
OBP	(H + HBP + BB) / PA	540 BF	
SLG	(1B + 2 * 2B + 3 * 3B + 4 * HR) / AB	550 AB	Min 2000 AB's, Cronbach's alpha used
ISO	(2B + 2 * 3B + 3 * HR) / AB	630 AB	Min 2000 AB's, Cronbach's alpha used
GB rate	GB / balls in play	70 BIP	Min 1000 BIP, Retrosheet

			classifications used
FB rate	(FB + PU) / balls in play	70 BIP	Min 1000 BIP including HR
LD rate	LD / balls in play	650 BIP	Min 1000 BIP including HR, Estimate*
HR per FB	HR / FB	400 FB	Min 500 FB, Estimate*
BABIP	Hits / BIP	2000 BIP	Min 1000 BIP, HR not included, Estimate*

Note in particular how long it takes for home runs and for batting average on balls in play to stabilize, particularly in relation to strikeouts and walks. For the former, this is evidence that the former is not as good an indicator of true ability as McCracken believed, whereas for the latter, it suggests that, although there is some skill behind it, the impact batted balls in play for a given season is a relatively poor indicator of pitcher performance.

The 2008 data here are based on correlating pitcher measures for 2001 with 2002, 2003 with 2004, and 2005 with 2006, for 750 batters faced each of those seasons:

Rate stats:
1. K/PA – .873
2. K/BB – .806
3. BB/PA – .789
4. 1B/PA – .525
5. HR/PA – .323
6. 2B+3B/PA – .237

One-number stats:
1. AVG – .527
2. OPS – .459
3. SLG – .455
4. BABIP – .188

Batted ball stats:
1. Line drives – .936
2. Ground balls – .905
3. Fly balls – .862
4. GB/FB – .852
5. Pop ups – .764
6. HR/FB – .207

Again, homers per plate appearance are less, and batting average on balls in play are more reliable than McCracken would have claimed.

Carleton, Russell A. (2013). Will you be my mentor? https://www.baseballprospectus.com/news/article/9866/baseball-therapy-will-you-be-my-mentor/

Russell Carleton (2013) examined the myth that veteran catchers can serve as mentors for young pitchers. His sample size was every team from 1989 through 2008 with (1) a catcher at least 32 years old on opening day who caught at least 360 inning during the season (if two such catchers on a team, he used the older), (2) pitchers 27 or younger who faced at least 250 batters during the season and did not switch teams, and combining the two (3) the catchers needed to have at least 12 relevant pitcher-seasons for their sample size. The study revealed some evidence that such catcher-mentors might exist (Jason Kendall improved both strikeout and walk rate for young pitchers), but the impact was tiny, the sample size was too small, and overall there really isn't any reason to think that this wasn't a random finding.

Carleton, Russell A. (2013).The high-pitch-count hangover. https://www.baseballprospectus.com/news/article/21272/baseball-therapy-the-high-pitch-count-hangover/
Carleton, Russell A. (2013). Prioritizing the pitcher's health. https://www.baseballprospectus.com/news/article/21450/baseball-therapy-prioritizing-the-pitchers-health/
Carleton, Russell A. (2013) Leave me in, coach! https://www.baseballprospectus.com/news/article/21369/baseball-therapy-leave-me-in-coach/

This is a series of studies in which Russell Carleton examined the impact of a high pitch count game on subsequent pitcher performance. All included data from 2003-2012, undoubtedly from Retrosheet. The first examined the impact of such a game on the next start. Looking at all plate appearances in the data set with both pitchers and batters with at least 250 PA and controlled for handedness advantage, here are some significant (cut-and-pasted) findings:

Outcome	**Expected — 100 Pitches Last Game**	**Expected — 110 Pitches Last Game**	**Expected — 140 Pitches Last Game**
Single	15.32%	15.34%	15.51%
Home run	2.74%	2.77%	2.87%
Out in Play	45.86%	45.75%	45.40%

Russel figured that the difference between 100 and 110 pitches for an entire staff over a season would be only three or four runs, and 15 to 20 runs for a jump to 140. The second displays the impact of pitch count on chances of going on to the disabled list between that start and the end of the season:

Threshold crossed	**Predicted contribution to injury risk**	**Delta from above**
75 pitches	6.59%	—
80 pitches	6.91%	0.32%
85 pitches	6.59% [sic]	(0.32%)
90 pitches	6.89%	0.30%
95 pitches	6.75%	(0.14%)
100 pitches	6.59% [sic]	(0.16%)
105 pitches	6.43%	(0.14%)
110 pitches	6.32%	(0.11%)
115 pitches	6.71%	0.39%
120 pitches	5.62%	(1.09%)
125 pitches	5.85%	0.23%
130 pitches	10.19%	4.34%

This needs some interpretation. Each of these.59 percentages take into consideration not only the present game but any hangover effect from previous games, so the only meaningful figures are the percentage increases. Also keep in mind that the pitchers with 110-120 pitch counts are probably those that team management thinks can handle that many, so is not a representative sample of major league starters. The big jump at 130 is significant and consistent with other work indicating that limit. The third examined whether there was a psychological impact for pulling a pitcher throwing a good game. Although one can never be sure, there was no evidence of one. Including starts featuring seven and eight shutout innings and a pitch count over 95, Russell noted no uncovered no impact on the next start for whether or not the manager pulled him before the eighth or ninth respectively.

Carleton, Russell A. (2013). Using the closer to keep a deficit small.
https://www.baseballprospectus.com/news/article/21499/baseball-therapy-using-the-closer-to-keep-a-deficit-small/

Along with basically replicating Dave Smith and Pete Palmer's work on the myth of the proven closer, Russell Carleton (2013), using Retrosheet 1993-2012 data, compared the proportion of runs given up with a one-run deficit in the top of the ninth for a home team and bottom of the eighth for a visiting team between a team's third best reliever, who is usually on the mound then, and the team's closer:

Runs Scored	Percentage with Third-Best Reliever	Percentage with Closer	Win Probability After if this had been a "one run

					down" situation	
	Ninth	Eighth	Ninth	Eighth	Ninth	Eighth
0	74.1%	69.4%	77.2%	75.0%	18.1%	14.9%
1	14.2%	17.0%	13.2%	13.9%	7.8%	6.5%
2	6.3%	7.4%	5.9%	6.5%	3.4%	2.8%
3	3.1%	3.0%	2.3%	2.6%	1.5%	1.2%
4	1.2%	1.8%	0.9%	1.6%	0.6%	0.5%
5+	1.1%	0.8%	0.4%	0.3%	0.2%	0.2%

Taking into consideration the proportion of times each of these run-scored events occurred, the overall win probability when one run down in the top of the ninth situation was 11.7 percent for the third-best reliever and 12.3 percent for the closer; the corresponding figures for bottom of the eighth were 14.8 and 15.2. In summary, the difference between the first- and third-best reliever is about 0.5 percent a game.

Carleton, Russell A. (2013). Saving the save. https://www.baseballprospectus.com/news/article/21557/baseball-therapy-saving-the-save/

Using Retrosheet 1993-2012 data, Russell Carleton computed the following leverage scores for beginning of "the most important innings" (not defined clearly) for home (top of the inning) and visiting (bottom of the inning) teams when in the field, with the ninth including extra innings:

Inning	Score Differential	Home Team		Visiting Team	
		Win Probability	Leverage	Win Probability	Leverage
9th	Up 1	86%	2.35	82%	2.93
9th	Tied	52%	2.05	33%	1.86
8th	Up 1	76%	1.93	70%	2.29
8th	Tied	53%	1.66	36%	1.52
9th	Up 2	94%	1.60	92%	2.06
7th	Up 1	72%	1.55	62%	1.83
9th	Up 2	89%	1.44		
7th	Up 2	84%	1.39	78%	1.54
7th	Tied	53%	1.36		
6th	Up 1	69%	1.34	57%	1.52
8th	Up 2			84%	1.82
8th	Up 3			92%	1.41

Carleton, Russell A. (2013). What my four-year-old taught me about bunting.

https://www.baseballprospectus.com/news/article/21998/baseball-therapy-what-my-four-year-old-taught-me-about-bunting/

Based on an analysis by Russell Carleton (2013) of non-pitcher bunts with a runner on first and no outs as indicated in 1993-2012 Retrosheet data, its prevalence had its ups and downs; about 9.4 percent in 1993, down to about 6.1 percent in 2000 and 2001 (reaction to the steroids era?), back to 8 percent in 2003 (end of that era and subsequent decrease in power hitting), down again to 6.4 in 2010 and 6.3 in 2012, but with an 8 percent between those two. Rates of conventional success (runner on second, one out) decreased from about 70 percent to the mid 60's, with "extra value" outcomes (I assume mostly both runners safe) and "problematic outcomes" (I assume mostly a force a second) in the mid or high teens.

There has been a problem with the analyses indicating the decrease in run expectancies resulting from the conventional use of the sacrifice bunt. It is not that the conclusion is wrong as such. Russell Carleton (2013), when using Retrosheet data from 1993 to 2002 to examine the issue, noted that although they jumped around a bit during those 20 seasons, bunting in the no out/runner on first situation resulted in around .10 runs fewer than swinging away overall. The problem is that the situations in which sacrifice bunts occurred tended to be chosen strategically. First, those asked to sacrifice were weaker hitters (in 2012, averaging a wOBA of .300 whereas those not were at .319), which if taken into account decreased the deficit by about .04 runs on average. In addition, those who sacrificed were asked to do so before stronger hitters; the on-deck hitters when sacrifices occurred had an average wOBA of .322 versus .314 otherwise. The narrowed the overall gap another .03 runs. This leaves a deficit of only around .03 runs, which results in the play having far less negative impact than often supposed.

Carleton, Russell A. (2013). The effects of the shutdown (inning). https://www.baseballprospectus.com/news/article/22086/baseball-therapy-the-effects-of-the-shutdown-inning/

Using most certainly Retrosheet data from 1993 to 2012, Russell Carleton (2013) located all instances in which a team had tied or taken the lead in the previous inning when their starter was still pitching to distinguish circumstances in which that starter completed the next inning without giving up a run, i.e. pitched a "shutdown" inning. After controlling for pitcher quality, Russel learned that these occurred more often than would be expected by chance, and that they did increase the odds of winning by a tiny amount, in his words "a couple of tenths of a percentage point." Also, there was an effect such that a pitcher's rate of pitching shutdown innings correlated with performance for the rest of the game at .62. This implies the possibility of a pitcher skill difference here, but Russell determined to be potentially noticeable with a sample size of 260 or 270 shutdown innings, which is a greater number than even pitchers with very long careers would experience.

Carleton, Russell A. (2013). The corner-outfield inefficiency. https://www.baseballprospectus.com/news/article/22295/baseball-therapy-the-corner-outfield-inefficiency/

Russell Carleton (2013) addressed an interesting question; would there be strategic value in constantly switching the two corner outfielders during games so that the stronger fielder of the two were always playing the pull side, under the assumption that here is where the specific batter was more likely to do damage. In so doing, Russell reported some interesting findings from 2003 to 2012 Retrosheet data. It is true that more damage occurred on pulled balls, because (1) they were more likely line drives as compared to flies when pulled (54%) than not (32%), (2) they were more likely to become hits when pulled for both types of batted balls (19.1% of flies and 86.1% of liners) than not (14.6% of flies and 78.4% of liners), and (3) more if hits more likely to be for extra bases rather than singles when pulled (40%) than not (33%). All in all, the average pulled ball to the outfield had a run value of .206 versus .022 when to the opposite field. But counteracting these tendencies was the fact that more balls to outfielders were actually hit to the opposite field, for both righthanded (54.5%) and lefthanded (55.4%) hitters. Even so, the strategy might be worth a couple of runs a season, at the expense of lengthening game time and perhaps tiring out the outfielders having to run back and forth between left and right field.

Carleton, Russell A. (2013). Is there a pinch-running penalty? https://www.baseballprospectus.com/news/article/22155/baseball-therapy-is-there-a-pinch-running-penalty/

Russell Carleton (2013) used what was certainly Retrosheet data to conclude the following: Between 2003 and 2012, 94.6 percent of steal attempts b pinch-runners occurred in the seventh inning or late (as would be expected given when they would be used), and 81 percent with the game within two runs with no strong tendency toward being in the lead (35.8%), tied (25.0%), or behind 39.2%. As compared with former batters on base with equivalent speed, pinch-runners were about 5 percent more likely to try to steal, 4 percent more likely to draw an attempted pickoff throw and (on the bad side) about a half a percent more likely to be picked off, and slightly more successful at stealing a base and advancing an extra base on a single or double.

Carleton, Russell A. (2013). Is there a pinch-fielding penalty? https://www.baseballprospectus.com/news/article/20960/baseball-therapy-is-there-a-pinch-fielding-penalty/

Based on 1993-1999 Project Scoresheet data located at Retrosheet, Russell Carleton (2013) uncovered no evidence that substitute fielders perform any differently than starters, with the exception of the former doing a bit worse at third base.

Carleton, Russell A. (2017). *The Shift.* Chicago, IL: Triumph Books.

I begin this with a quote from the Acknowledgement section (pages 321-322). After describing what Retrosheet is, he wrote that Dave Smith "should be inducted into the Baseball Hall of Fame. I am not exaggerating. At one point, I met Mr. Smith a conference of the Society for American Baseball Research and thanked him for the fact that I was able to feed my daughter." A large proportion of the analyses in this book are based on Retrosheet data. I will list those that I found useful in my work; I'm sure there are others.

In Chapter 3, Russell extended work by Dan Levitt mentioned below on runner advancement on hits, also most certainly based on Retrosheet data:

Attempt Type	Average Attempts Per Team	Percentage of Attempts	Success Rate
First to third on single	177.5	31.2	96.6
Second to home on single	142.4	70.0	95.4
First to home on double	77.1	47.1	93.3

The implication, which we will see again below with more of Russell's work, is that teams are too conservative in trying for extra bases on hits. Here is his 2015-2016 data on sacrifice flies and distances, adding evidence to previous Pete Palmer estimates showing conservative in sending runners from third on outfield flies:

:

Fly Ball Distance	Percentage of Attempts	Success Rate Percentage
225 or less	18.4	88.0
226 to 250	17.3	100.0
251 to 275	57.7	94.7
276 to 300	91.8	99.8
301 or more	99.5	99.9

The 100 percent must be a small sample size fluke. In any case, third base coaches are being way too conservative. But if Russell's estimate that the average team losses only about three runs a year due to this conservativism, there are bigger strategic sins.

Chapter 7 includes work paralleling Dave Smith's on increases over time in the length of games, centering on years ending in 6 from 1976 to 2016. Russell's analysis shows that some of this growth is connected with more strikeouts (4.83 to 8.03) and, interestingly, more batters hit by pitches (0.18 to 0.34, with most of the jump occurring between 1986 and 1996 (why?) per game per team. But there was no analogous rise in walks per game per team; although those rose from 3.20 (1976) to 3.55 (1996), they then fell to 3.11 (2016; Dan Levitt, as announced on the SABR Statistical Committee blog on July 14, 2018, uncovered the same increase in strikeouts and inverted-U function for walks in WS games during about the same time period). Russell (page 218) also reported that the average length of time between pitches over the course of a game in his data started at 19⅓ during first three innings, jumped to about 20¼ in the fourth and fifth, to 21 in the sixth, and continued rising to 22 by the ninth. Finally, strikeouts per game per team was already up to 6.52 in 2006, which means that the increase in the visibility of baseball analytics has probably not been primarily chiefly responsible for the additional whiffs.

Also in Chapter 7, replicating a study published by the STATS folks back in 1990 (John Dewan, Don Zminda, and STATS. Inc. *The STATS Baseball Scoreboard*. New York: Ballantine Books), Russell displayed figures relevant to the final outcomes of plate appearances and number of pitches fouled off for 2016 (no fouls, .170/.232/.263; one foul, .194/.282/.310; more than one foul, .205/.308/.339). He also paralleled work by Dave Smith on the slight outcome difference that depended on the type of strike one; swinging strike (.206/.255/.328), called strike (.229/.273/.359), and foul ball (.229/.272/.367).

Even more in Chapter 7; there were 16510 throws to first base in 2016, in which 1.7 percent resulted in pickoffs and 0.7 percent were thrown away, allowing runner advancement. This means that on 97.6 percent of throws, nothing happened. However, when there were throws, the rate of successful steals went down 5 percent, likely due to shorter leads. Not surprisingly, faster runners attracted more tosses (extreme examples; Dee Gordon 66% of times and Kendrys Morales 0% of times on first).

Chapter 10 – Does how a team gets into the playoffs matter in regard to playoff series wins? Is it better if a team has to claw its way in through winning crucial September games, or eases in given a big lead in the standings? Russell defined a “meaningful game” as one in September in which a team is within three games (ahead or behind) of a playoff spot that is not yet clinched. The following shows the relationship between such games and the percentage of playoff series subsequently won between 1996 and 2016:

0	49.2%
1-5	47.7%
6-10	52.6%
11-15	49.4%

In short, it doesn’t matter. Another playoff-relevant question Russell answered was whether a team either tied or behind in 9th which ended up winning the playoff game was inspired to win the next game. At first it looked that way, as it occurred in 58 of 98 (59%) relevant instances during those seasons. Looks here are probably deceiving, as the team winning one game won 54% of the next games overall. I say probably deceiving, because the truly correct analysis is to subtract the tied-or-behind games to see the percentage for teams winning games in which they were already ahead.

Someplace in the book, during a discussion of starting pitching, he noted that whereas there were 383 starts lasting more than 120 pitches in 1987, that number had dropped to 45 in 2016. Also somewhere, he has evidence that reliever usage may not be affected by whether there is or is not a game the next day:

	Batters Faced By Starters	Pitches Thrown By Starters	Number of Relievers Used	Relievers Facing 1 or 2 Batters
Before Day Off	24.68	92.18	3.11	0.54
No Day Off	24.78	92.61	3.02	0.49

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

This is a mostly trite article on the impact of the decision whether or not to swing on the first pitch of a plate appearance on PA outcomes, based on Retrosheet data from every 2010 plate appearance. The most interesting finding was that batters tend to change their decision starting at the third PA in a game from the previous two PAs; e.g., 3rd PA from 1st and 2nd and 4th PA from 2nd and 3rd. My guess is that this is likely a response to pitching changes.

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

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

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

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

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.

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.

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).

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.

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	Number	Number	Opposition	Pitcher	Number	Number	Opposition

Season Score	of Pitchers	of Starts	Pitcher Season Score	Season Score	of Pitchers	of Starts	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.

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.baseballprospectus.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.

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 (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.

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.

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.

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).

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.

Strategy Pitcher Usage patterns (PU)

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 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.

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*).

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 <u>With Or Without You</u> 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., 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.

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/

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. I find it difficult to substantively reconcile Tom and Keith's very distinct conclusions.

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 (2009b). 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.

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.