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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

BASE STATE	OUTS	PA	BUNT RATE
1__	0	5,439	81.1%
1__	1	6,571	73.9%
_2_	0	1,611	72.9%
12_	0	1,374	81.2%
12_	1	2,396	71.4%
1_3	0	474	67.1%
1_3	1	1,005	57.0%

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

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

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

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

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

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

Change in Outcome	Correlation with Change in FB Rate
Contact Rate (per swing)	-.114
Strikeout	.093
Walk	.054
Single	-.286
Double/Triple	.093
HR	.341
Out in Play	-.047
OBP	-0.01

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

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

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

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

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

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

Baxamusa, Sal (2006). The memory remains. https://tht.fangraphs.com/the-memory-remains/

Baxamusa, Sal (2007). More on pitch sequences. https://tht.fangraphs.com/more-on-pitch-sequences/

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

Situation	AVG	OBP	SLG
After 1-1 pitch (entire PA)			
First pitch strike	.257	.314	.402

First pitch ball	.243	.312	.378
Ball in play on 1-1 pitch			
First pitch strike	.336		.528
First pitch ball	.299		.472

and

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

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

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

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

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

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

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

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

```
BBBSX        .272/.272    12
```

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

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

This piggybacks on Tom Tango's pitch count estimator:

(3.3 X PA) + (1.5 X SO) + (2.2 X BB)

which implies that the average batted ball should occur after 3.3 pitches, average strikeout after 4.8 pitches, and average walk after 5.5 pitches. Sal Baxamusa (2007) used Retrosheet data to determine that in 2006, the actual figures were 3.3, 4.8, and 5.5, so the formula did well that season. Interestingly, every hit type also averaged 3.3 pitches with the exception of triples (3.5, but with a small sample size).
Looking at a graph, approximately 18 percent of plate appearances lasted either 3 or 4 pitches, with about 17 percent going 2 pitches, 16 percent 5 pitches, 12 percent 1 pitch, maybe 10½ percent 6 pitches, 4½ percent 7 pitches, 2 percent 8 pitches, 1 percent 9 pitches, and progressively less often afterward. Sal also showed a graph for the outcomes of different numbers of pitches, but these were not as interesting as strikeouts cannot appear until the third pitch and walks until the fourth. The length of a PA had no effect on the type of batted ball.

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

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

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

In the last webpost in this series, using 2006 Retrosheet data for batters with 200 or more PAs (sample size of almost 300), Sal compared the top 20 and bottom 20 to see if the length of their PAs differed systematically. Contact was measured by percentage of

PAs ending with balls in play, and passivity by percentage of strikes that were called rather than whiffs or fouls,

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

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

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

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

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

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

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

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

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

Biolsi, Goff and Wilson used Retrosheet data from 2000 to 2017 to examine another possible defensive interdependence, that between pitchers and catchers regarding getting outs of all types and strikeouts. There were a total of 5519 pitcher-catcher matches between 2000 and 2017; the authors used the 75 percent most active of those for each analysis. There was a lot of variation across seasons, but in general for individual seasons, pitchers had the most impact on both outs and strikeouts, then catchers, and finally the specific pitcher-catcher match the least. However, when combined across seasons, the match had more impact than the catcher and, in the case of outs, almost as much as the pitchers. The authors' proposed explanations for the difference between within and across season findings were that (1) the increase in sample size obtained from combining seasons reduced noise that appeared in the yearly individual pitcher and catcher coefficients, and that (2) good pitcher-catcher matches take time to develop and the development time was reflected in the single season data; this latter proposal was supported when examining factors potentially affecting the overall results. In addition, pitcher-catcher matches were slightly more influential when the two came from the same country and, more strongly, spoke the same first language, and when their MBL debuts had been closer together in time.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Phil Birnbaum (2011), in response to the claim by Moskowitz and Wertheim (hereafter MW) in their book *Scorecasting* that pitch calls favor the away team in low-leverage situations, argued that this implies that the home team scoring advantage over away teams should be highest when leverage is highest, which tends to be in the last innings. Using Retrosheet data from 1957 to 2007, here are the inning-by-inning differences in run scoring, contrary to MW.

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

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

Inning	Runs	Percent
2	2543-2139	+19

3	4583-4176	+10
4	8817-7801	+13
5	10940-10057	+9
6	14371-13279	+8
7	15698-14583	+8
8	16935-16180	+5

And just to away teams ahead by four or more

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

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

Bond, Brittany, and Ethan Poskanzer (in press). Striking out swinging: Specialist success following forced task inferiority. *Organization Science*.

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

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

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

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

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

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

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

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

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

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

Brill, Ryan S., Sameer K. Deshpande, and Abraham J. Wyner (in press). A Bayesian analysis of the time through the order penalty in baseball. *Journal of Quantitative Analysis in Sports.*

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

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. aka Pizza Cutter (2007). The triumph of Pythagoras.
https://statspeakmvn.wordpress.com/2007/10/

1980-2006 including 1981 and 1994, which was a mistake. Halfway through the season, using runs scored and given up at that point to calculate Pythagenpat correlated with end season winning average at .494 whereas using 81 game winning average to predict end of season winning average was at .464.

Carleton, Russell A. aka Pizza Cutter (2007). Still more Pythagoras musings. https://statspeakmvn.wordpress.com/2007/10/

Russell took the 100 most over- and under-achieving teams 1901 to 2007 given the difference between their Pythagenpat and actual winning average, which over(under) achieved by an average of about 8 wins, and noted that in the next season, the former over-achieved by 0.36 wins, later under-achieved by 0.37 wins. The conclusion; that there was some hold-over effect the next season. Incidentally, Year 1 winning average correlated at 0.603 and year 1 Pythagenpat at 0.626 with year 2 winning average.

Carleton, Russell A. aka Pizza Cutter (2007). Pythagoras solved? An R-squared of 97.8 percent. https://statspeakmvn.wordpress.com/2007/11/

Actual winning average, average margin of victory and average margin of loss predicted 97.8% of variance in difference between Pythagenpat and actual winning average. So as before, teams with small average margin of victory and large average margin of losses more likely to outperform Pythagoras. Incidentally but importantly margins of victory and defeat could be hypothesized to be negatively associated assuming that good(bad) teams win(lose) by big margins and lose(win) by small margins. However the two were actually positively correlated at about 0.2, implying that they are largely unrelated.

Carleton, Russell A. aka Pizza Cutter (2007). Managers and the Pythagorean Theorem. https://statspeakmvn.wordpress.com/2007/12/

All available games from 1871 to 2006, designated each as either an odd or even game depending on its sequential number in a season. Russell the Pizza Cutter took all managers with at least 500 games experience and did split halve correlations, looking for the degree of consistency in team outcomes. The correlations were all positive; 0.342 for one run games, 0.323 for two run games, and 0.756 for blowouts (defined as a 6 run differential or greater). This implies some consistent difference in managers winning vs losing close games, plus huge consistency in blowouts (which is likely a function of team quality). Managerial experience (games managed) had no impact (correlations of 0 for one run games, 0.01 for two run games, 0.04 for blowouts; and the difference between both managers in experience correlated 0.007 in one and two run games, 0.06 in blowouts.

Carleton, Russell A. aka Pizza Cutter (2007). Stats 204: The proximity matrix OR Re-visioning similarity scores. https://statspeakmvn.wordpress.com/2007/11/

Russell Carleton under his Pizza Cutter alias (2007z) invented a more sophisticated method for computing similarity score limited to batting, by in effect assigning players locations in a four-dimensional statistical space defined by K, BB, and HR rates and BABIP and computing the distances among the players.

Carleton, Russell A. aka Pizza Cutter (2007). On throwing to first, part I. https://statspeakmvn.wordpress.com/2007/03/31/on-throwing-to-first-part-i/
Carleton, Russell A. aka Pizza Cutter (2007). On throwing to first, part II. https://statspeakmvn.wordpress.com/2007/04/06/on-throwing-to-first-part-ii/
Carleton, Russell A. aka Pizza Cutter (2007). On throwing to first, part III. https://statspeakmvn.wordpress.com/2007/04/12/on-throwing-to-first-part-iii/
Carleton, Russell A. (2015). The wonderful world of throwing to first. https://www.baseballprospectus.com/news/article/26734/baseball-therapy-the-wonderful-world-of-throwing-to-first/

Russell Carleton contributed three entries early on (all 2007) that he continued later (2015) with a wealth of relevant data regarding the value of pitchers throwing to first when that base is occupied. I will report much of it next, beginning with the first three (based on 2006 Retrosheet data). All are based on runner on first/second empty situations.

Overall figures: The proportion of relevant base-out-inning events and the total number of throws was correlated at about .75 with steal attempts and Bill James's Speed Scores figure. Throws to first cut down the stolen base success rate from 76.8 percent to 65.4 percent, and was a significant predictor of success with runner Speed Scores controlled (which is important because faster runners get more throws). Russell estimated that thwarting stolen base attempts was worth about five runs saved a year. Although errant throws cost a run or so a team, successful pickoffs saved about another four, meaning that throws in and of themselves would be worth close to a win a season if there were no potential tradeoffs; as we shall see shortly, there are.

The game situation mattered. Pitchers were more likely to throw to first with no outs (27.6% of the time with runner on first and second empty, versus 26.6% with one out and 23.2% with two outs), a closer score, early innings (34.4% in the first, decreasing steadily until 16.1% in the ninth). There was quite a bit of variation in team usage of this strategy (Brewers tops at 36%, Dodgers bottom at 15%), but within teams, there no difference in usage depending on the catcher.

Because first basemen play closer to the bag with a runner on, it follows that the number of throws impact on fielding. In this case using 1997 Retrosheet data as it has hit location data based on Project Scoresheet Event Form fielding sectors. Balls hit close to the foul line (zone 3L) are easier for the first baseman with a runner on, and so became hits less often (34% with a throw, 37.6% without). Balls hit farther from the foul line are harder, and so became hits far more often (zone 3, 22% with a throw, 8.3% without; zone 34, 63% with a throw, 47.5% without). As Russell pointed out, the data is not definitive as there are not always throws when there is a runner on first, it certainly is very suggestive, and implies a significant tradeoff for throwing in and of itself as mentioned above.

Hitters did slightly better without a throw (back to 2006 data:.281/.345/.446 without one, .270/.343/.428 with one), but Russell did not have the time to see If this was a real effect or an artifact of the likelihood that throws are more likely with weaker hitters as there is more reason for the team at bat to try and steal.

All of this was based on just one year of data, and so one should beware small sample sizes. When Russell returned to the issue eight years later, he had a far bigger (2010-2014 Retrosheet) data set.

Overall figures: There was a 1.6 percent success rate on pickoff attempts, with 0.6 percent becoming errors. There was a lot of variation among pitchers in throw, from none to more than 50 percent of opportunities, but the number of pitcher attempts correlated at only .26 with attempts to steal and a nonexistent .04 with success rate. As for the runners themselves, there was a lot of consistency across seasons (joint correlations [ICC] of over .90) in drawing throws. Controlling for inning, number of outs, whether there was a runner on third, and game score, runners on first were three percent <u>more</u> likely to attempt a steal if the pitcher had thrown over and a bit more likely to try for third on a single, but were 12 percentage points more likely to get caught. An errant throw slightly lowered the chance of pitcher throwing again that game but had no effect on subsequent error throw rate, and a successful pickoff both increased the odds of throwing over for the rest of the game and lowered the probability of attempted steals, but not the success rates

Turning to batter outcomes and controlling for pitcher's overall tendency to throw to first, a throw resulted in fewer singles/doubles/triples/outs on balls in play, translating to more strikeouts and also more walks, such that OBA and BABIP were unaffected. Russell surprisingly wrote nothing about any impact on homers.

Given all possibilities, he estimated that a throw helps the pitcher .0064 runs.

The last of the 2007 entries also included data on stolen base rates. Excluding "automatic" 3-2 count 2 out take-off-with-the pitches, attempts to steal second were highest in the first inning (16.3%), dropped to about 12 percent for several innings, 9.2 percent in the eighth, but back to 13.5 percent in the ninth. Success rates were 76.1 in the first inning, down to 65.2 percent in the second, then rising steadily to 75 percent in the seventh but slightly less in the eighth and ninth. Only one year of data, so reader beware.

Carleton, Russell A. aka Pizza Cutter (2007). Runner tagging from third, here's the throw... http://baseballpsychologist.blogspot.com/2007/03/runner-tagging-from-third-heres-throw.html

Carleton, Russell A. aka Pizza Cutter (2007). There's gonna be a play at the plate. http://baseballpsychologist.blogspot.com/2007/03/theres-gonna-be-play-at-plate.html

Carleton, Russell A. aka Pizza Cutter (2008). How to make your team better by firing your third base coach. https://statspeakmvn.wordpress.com/2008/04/page/3/

Russell Carleton (2007) posted a series of studies detailing the ultraconservatism of third base coaches in sending runners from third on outfield flies. In the first, Russell used the hit location data from 1993-1998 Project Scoresheet scoresheets available at Retrosheet to estimate fly ball distances for plays with a runner on third, 0 and 1 out, and a fly ball/liner caught by outfielder (sample size = 9415). The runner tried for home 84 percent of the time with a success rate 97.1%. Fly ball distance accounted for 49.5 percent of the variance in the decision to go for home but only 17.2 percent of variance on whether they were successful given an attempt; the first clue that runners/coaches are too conservative. A follow up from the same year revealed that Bill James's version of Speed Score accounted for a paltry 1.1 percent of variance on the decision to run.

Returning to the topic the next year, Russell used 1993 run expectancy matrix to show that the break-even point for sending a runner from third with no outs on was a 75.6 percent success rate, more evidence of conservatism. He then constructed a model to predict odds of successful scoring based on fly ball distance from the data. In 1993, given fly ball distance, only 22 of 1322 fly balls gave a probability less than the break-even, and the runner stayed at third on 19 of these, a good decision. Of the remaining 1300, runners held 232 times, clearly a mistake given the known success rate. Overall, even including the 19 who should have stayed, the 251 total holds cost the team 0.365 runs each.

Carleton, Russell A. aka Pizza Cutter (2007). Third base coaches, get your windmill arm ready. https://statspeakmvn.wordpress.com/2007/05/

Carleton, Russell A. aka Pizza Cutter (2007). Is speed really that important? https://statspeakmvn.wordpress.com/2007/08/

Carleton, Russell A. aka Pizza Cutter (2007). Is speed really that important, part II. https://statspeakmvn.wordpress.com/2007/08/

This continued the demonstration of ultraconservatism on the base paths. In the first of these, Russell used 2000-2006 probably Retrosheet data to compute break-evens for advancement on hits. They were:

First to home on double, 86.7% with 0 out, 79.4% with 1 out, 43.1% with 2 outs

Second to home on single, 91.7% with 0 outs, 70.3% with 1 out, 39.8% with 2 outs.

First to third on single, 91.2% with 0 outs, 76.9% with 1 out, 91.6% with 2 outs. This one supports the myth to not make the first or third out at third base.

Success rates were in the 90's for everything, once again demonstrating harmful risk aversion in base running strategy.

In the second and third, Russell used 2003-2006 data to estimate the amount of variance in success and attempt rates accounted for by Bill James's Speed Score.

First to home on a double, Speed Score predicted only 1.0 percent of variance in success rate and 2.0 percent of attempt rate.

Second to home on a single, Speed Score predicted only 1.2% of success rate, 1.7% of attempt rate.

First to third on a single, Speed Score predicted only 0.2% of success rate, 1.4% of attempt rate.

Additional analyses: Attempts at stealing second, Speed Score predicted only 4.2% of success rate, 10.2% of attempt rate.
Runner on first, batter success at beating out throw to first on double play attempt 5.5%. Part of the reason for these small numbers was the absence of variance in some of these, but Russell noted how many other factors come into play in these events.

Carleton, Russell A. aka Pizza Cutter (2007). How much is that closer worth anyway? https://statspeakmvn.wordpress.com/2007/07/page/2/

Inspired by the save rule, Russell computed for 2000-2006 how often leads of 1 to 3 runs were lost, resulting in behind tied, or behind, in the top and bottom of every inning. Leads cannot be lost in the top of the first. Cut and pasted data:

The full chart:

Inning	Top	Bottom
1st	—	24.6%
2nd	19.5%	19.9%
3rd	18.1%	19.1%
4th	20.0%	20.6%
5th	17.0%	19.3%
6th	19.0%	19.2%
7th	16.9%	17.6%
8th	14.8%	15.8%
9th	13.0%	13.9%

Carleton, Russell A. aka Pizza Cutter (2007). I thought we were all professionals here. https://statspeakmvn.wordpress.com/2007/10/

The point here is to see if there is any consistency in making so-called "productive outs," measured by Win Probability Added (or lost) divided by leverage to make the PA context neutral. The data was 2003-2006 situations with fewer than two outs and baserunners aboard when the batter (minimum 100 PA per season) made out. Russell measured a possible skill in making so-called "productive outs" by WPA (or lost) divided by leverage to make the PA context neutral. A year-to-year intraclass correlation of 0.16 (0.14 when weighted by player PA) provided little evidence for productive out making as a skill.

Carleton, Russell A. aka Pizza Cutter (2007). A small update on "clutch relief." https://statspeakmvn.wordpress.com/2007/10/page/2/

At some point, according to Russell Carleton (2007), Tom Tango and David Appleman on FanGraphs started defining clutchiness based on the difference between player's actual Win Probability Added and what WPA would have been had all PAs occurred with leverage of 1. This makes sense for hitters, but as Matt Souders pointed out, it doesn't work for relievers that generally pitch in situations with leverage greater than 1. In a study I could not find, Russell used the original definition and noted no clutch effect for relievers. Matt recommended comparing WPA corrected for the specific leverage a given reliever faced on average with WPA corrected for league average. Russell did that for 2003-2006, and uncovered a year-to-year intraclass correlation of –.059 for pitchers with at least 100 PA. This provides no evidence that pitchers differ consistently in clutch ability. Russell did the same for batters (I assume 100 PA minimum), and this time got an even smaller ICC of –.015, with the same implications as for pitchers.

Carleton, Russell A. aka Pizza Cutter (2007). Does swinging at the pitch really protect a base-stealer? https://statspeakmvn.wordpress.com/2007/05/

Russell Carleton's (2007) work on this issue was based on 2006 Retrosheet data for attempts to steal second and third, not including pickoffs. For attempts at stealing second, success rates were 65.7 percent with a swing and 77.9 percent no swing. For attempts at stealing third, they were 56.8 percent with a swing and 82.1 percent with no swing. Discussion with readers following up on this webpost led to the hypothesis that the reason success rate were lower with swings is that a significant proportion of those are busted hit and run attempts, generally done with slower runners. As for other proposals, runners going on attempted steals with a swing averaged a tiny ½ point less on Bill James's Speed Score metric than those without a swing. Further a respondent named John Beamer (based on suggestions from “Guy” (Molyneux?), with 2000-2005 Retrosheet data basically extinguished the swing-no swing difference when controlling for specific baserunner, supporting the first hypothesis. Russell also noted that success rate was not affected by whether pitches, assuming with no swing, were called a strike (76.3%) or ball (78.5%). Finally, there was no substantial difference based on batter handedness (lefty batter, 76.1%; righty batter 75.0%).

Carleton, Russell A. aka Pizza Cutter (2007). What's the most important at bat in an inning? https://statspeakmvn.wordpress.com/2007/05/12/whats-the-most-important-at-bat-in-an-inning/#more-179

This is the average leverage for different base/out situations for 2006, from Retrosheet data.

1.99��2 outs, bases loaded
1.74��2 outs, runners on 1st & 3rd
1.68��2 outs, runner on 3rd
1.65��1 outs, bases loaded
1.62��1 outs, runners on 1st & 3rd
1.61��2 outs, runners on 2nd & 3rd
1.60��2 outs, runners on 1st & 2nd
1.56��2 outs, runner on 2nd
1.45��1 outs, runner on 3rd

1.39��1 outs, runners on 1st & 2nd
1.34��1 outs, runners on 2nd & 3rd
1.14��1 outs, runner on 2nd
1.06��0 outs, runner on 1st
1.04��1 outs, runner on 1st
1.04��0 outs, runners on 1st & 2nd
1.00��2 outs, runner on 1st
0.87��0 outs, runner on 2nd
0.82��0 outs, no runners
0.76��1 outs, no runners
0.76��0 outs, bases loaded
0.76��0 outs, runners on 1st & 3rd
0.72��2 outs, no runners
0.71��0 outs, runner on 3rd
0.65��0 outs, runners on 2nd & 3rd

Carleton, Russell A. aka Pizza Cutter (2007). Testing the Ewing Theory. https://statspeakmvn.wordpress.com/2007/07/

The Ewing Theory (named after Patrick Ewing and named by sportscaster Bill Simmons) claims that teams with a superstar play better without them. Using 1980-2006 Retrosheet data, Russell determined that teams with exactly one “superstar” (defined as top 30 OPS in majors in a given season with at least 400 AB) had a winning average of .504 when the superstar played versus .472 without him. Forget about the Ewing Theory.

Carleton, Russell A. aka Pizza Cutter (2008). The foul ball part 2: What does it tell us about a pitcher? https://statspeakmvn.wordpress.com/2008/04/

Inducing foul balls (FB) seems to be a pitching skill. Using 2004-2007 probably Retrosheet data, intraclass correlations of 0.696 for foul balls per plate appearance and 0.753 for FB per contact imply that pitchers give up roughly the same proportion from year to year. Foul ball percent for pitchers correlated positively with fly balls (0.411), strikeouts (0.440), negatively with ground balls (–0.440), and not with walks (–.020). Fouls per contact was negatively related with slash line metrics at –0.535, –0.352, and –0.387, and also (for some reason) positively with walks at 0.205.
Pitching to contact is a mistake. The intraclass correlation for contact rate of pitchers was 0.805. Correlations between contact rate and slash line metrics were 0.610, 0.381, and 0.494 (note that it was lower for OBA than BA because of walks, which correlated with contact rate at –0.245. The correlation with singles was 0.519 with singles. And with strikeouts, –0.844.

Carleton, Russell A. aka Pizza Cutter (2008). Who gets the credit/blame for that home run? https://statspeakmvn.wordpress.com/2008/02/page/3/

Based on 1993-1998 Project Scoresheet fly ball location data (so approximate only) for pitchers giving up at least 25 fly balls and popups. Mean distance consistency year-to-year correlated at 0.312, implying some consistency in fly ball distance.

Carleton, Russell A. aka Pizza Cutter (2008). Do hitters get more jumpy during a slump? https://statspeakmvn.wordpress.com/2008/02/

When batters feel that they are in a slump (defined as 0 or 1 hit in their previous 10 AB), does their strategy change? Addressed by Russell Carleton (2008) using 2006 data for all batters with at least 50 AB. Although average pitches faced per AB was not affected, those in slumps increased their response bias as defined in one of his earlier pieces on plate discipline.

Carleton, Russell A. aka Pizza Cutter (2008). On the 100 pitch limit. https://statspeakmvn.wordpress.com/2008/05/page/2/
Carleton, Russell A. aka Pizza Cutter (2008). More on pitcher fatigue. https://statspeakmvn.wordpress.com/2008/06/page/3/
Carleton, Russell A. aka Pizza Cutter (2008). Pitcher fatigue, batted balls, and DIPS. https://statspeakmvn.wordpress.com/2008/06/page/2/

The first webpost in this set revealed the results for regression equations predicting K rate for different pitch counts for starters, controlling for pitcher and hitter quality; 2000-2006 Retrosheet data.

Pitch Count	K%	BA	OBA	SLG
0	.1731	.2545	.3248	.4023
10	.1688	.2568	.3263	.4068
20	.1646	.2591	.3277	.4113
30	.1605	.2614	.3291	.4158
40	.1564	.2637	.3306	.4203
50	.1525	.2659	.3320	.4249
60	.1486	.2682	.3334	.4294
70	.1448	.2704	.3348	.4340
80	.1411	.2726	.3362	.4385
90	.1375	.2749	.3376	.4431
100	.1339	.2771	.3390	.4477

110	.1304	.2793	.3404	.4523
120	.1271	.2815	.3418	.4569

Note the continual increase in slash metrics and decrease in strikeout rate with pitch count. Singles, extra base hits, home runs, and BABIP are each analogous; the latter is more evidence that BABIP is not totally random. In contrast, walks, which become less likely as pitch count increases.
The second webpost continued the theme, including games with 10 or fewer days of rest as more could imply injury/MiLB time etc. All of these effects are over and above specific game pitch count effects just described. Adding days of rest to the equation had no impact except for fewer HRs. Adding total pitches in the season up to that point to the regression equation also resulted in fewer walks over time, as did adding pitches for the previous start. Adding total pitches also increased fly ball rate and decreased ground ball rate, but had no effect on HR rate. Interestingly, even with the overlap between this and pitch count, adding times through the order had a big impact, decreasing K, BB, and GB rates and surprisingly BABIP and line drive rate (probably because a pitcher having good luck that game with balls in play will face more batters).
The third webpost recaps the first two.

Carleton, Russell A. aka Pizza Cutter (2008). He always gets off to a hot start.
https://statspeakmvn.wordpress.com/2008/12/page/2/

Russell Carleton (2008v) examined whether there was any evidence supporting the notion that some players are consistently hot or cold in given months, i.e. get off to hit/cold starts in April or hot/cold finishes in September year after year. Using 2004-2008 Retrosheet data for hitters with at least 70 PA in the relevant month and 400 for the season, he correlated monthly OBA with seasonal OBA (and admitted that there is a confound in that each month is included in the year). The answer; the highest intraclass correlation was a very small 0.11 for May. In particular, Russell was looking for hot starts, but the April figure was a non-existent 0.01. In other words, there is no evidence here for batters being consistently good or bad in specific months.
The second webpost continued the theme, including games with 10 or fewer days of rest as more could imply injury/MiLB time etc. All of these effects are over and above specific game pitch count effects just described. Adding days of rest to the equation had no impact except for fewer HRs. Adding total pitches in the season up to that point to the regression equation also resulted in fewer walks over time, as did adding pitches for the previous start. Adding total pitches also increased fly ball rate and decreased ground ball rate, but had no effect on HR rate. Interestingly, even with the overlap between this and pitch count, adding times through the order had a big impact, decreasing K, BB, and GB rates and surprisingly BABIP and line drive rate (probably because a pitcher having good luck that game with balls in play will face more batters).
The third webpost recaps the first two.

Carleton, Russell A. aka Pizza Cutter (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. aka Pizza Cutter (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. aka Pizza Cutter (2008). Wanted: players who like to run into things? https://statspeakmvn.wordpress.com/2008/05/page/2/

This applied 2004-2007 Retrosheet data for the following situaiion;, runner on first with fewer than two outs and a grounder fielded by an infielder, aka double play situation, when the runner on first was retired at second. The issue at hand was second base and shortstop year-to-year consistency in double play completion (25 chances minimum per year). Russell uncovered a 0.44 intraclass correlation, implying that this represents an infielding skill.

Carleton, Russell A. aka Pizza Cutter (2008). Playing the blame game with ground balls. One of the posts at https://statspeakmvn.wordpress.com/2008/06/

Carleton, Russell A. aka Pizza Cutter (2008). Vindicating Derek Jeter's fielding at short (sorta). https://statspeakmvn.wordpress.com/2008/06/24/vindicating-derek-jeters-fielding-at-short-sorta/

Russell Carleton's Out Probability Added Above Average (OPA!), a method for evaluating infielder fielding performance, relied on the batted ball location data present in 1993-1998 Retrosheet play-by-play data. It was limited to grounders that made it on to the infield dirt, adjusted for pitcher and batter handedness. Russell's method was as follows:

Step 1: For each of the four batter/pitcher handedness combinations, see what proportion of ground balls that make it to an outfielder are hit in each of the ten relevant infield zones (the Project Scoresheet Hit Location diagram shows only eight, but two right down the foul lines were added by these years). Russell's example: With batter and pitcher both righties, ground singles fielded by the left fielder went through the 56 zone 84.1 percent of the time, through the 5 zone 7.0 percent of the time, through the 6 zone 6.0 percent of the time, through 5L (one of the added zones) 2.2 percent of the time, and through the 6M zone 0.5 percent of the time.

Step 2: When a ground ball does not get through, find the proportion that fielded it for each of the infielders for each of the zones. Continuing Russell's example, again assuming righty/righty plate appearances:

Fielder	5L	5	56	6	6M
Shortstop	1.1%	0.3%	41.9%	97.6%	88.1%
Third Base	98.6%	98.8%	57.4%	1.2%	0.1%

These do not sum to 100 percent due to occasional plays made by others, usually the second baseman.

Step 3: Multiply the results of the two steps. Russell's example: As 84.1 (groundball singles fielded by the left fielder that were hit into the 56 hole) multiplied by 57.4 (56 hole (ground ball outs hit to the 56 zone fielded by third base) equals 48.2, that is the proportion of responsibility for these singles that third base is assigned.

Step 4: Sum the figures; third base ended with 54.2 percent responsibility for singles to left with righty/righty matchups.

There was a link to the full breakdowns, but it no longer works.

Step 5: For a given ground ball, determine whether the infielder got to the ball (measure of range), the infielder got the ball in the glove (measure of hands), the infielder threw it accurately to the first baseman (or other infielder if relevant; measure of arm), and whether the infielder successfully caught accurate throws from one another. The following is cut-and-pasted from Russell's account:

1. Take 2007 [his example season] and isolate all ground balls
2. Figure out the rates of expected outs by play state (after it leaves the bat, fielder got there, clean pick, good throw, 1B catches) controlling for who fielded it, and batter and pitcher hand.
3. Create a separate look at double play grounders, in which we isolate the *two* plays that will hopefully happen, and account for the fact that it's harder to turn the second leg of a double play.
4. On each play, code for whether the play was completed with no problems or where the play broke down (ball went through to the OF, it broke down at the "range" stage; fielder was charged with a fielding error, "hands" stage; no error, but the batter reached base OR fielder gets a throwing error, "arm" stage; 1B is charged with an error on the catch, "catch" stage) and if it broke down, who was at fault.

5. Aggregate it all together, including a total “outs added above average” column. …To control for the number of chances each player received, I gave him credit for a ball in his area if a) he fielded it or b) if he bore more than a 20% blame on the ball getting through, using the division of responsibility chart from last week.

The analyst can combine these into an overall measure, but a strength of OPA! Is the ability to isolate each skill from the others. In 2007, Derek Jeter, consistently with his sabermetric reputation, came in last in range among the 43 shortstops receiving at least 100 relevant grounders, but he came in ninth in arm and in turning double plays and eighth in receiving throws. Russell did not say where Jeter ranked in hands, but I would not be surprised if he did well there also.

Carleton, Russell A. aka Pizza Cutter (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 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 Probability Added 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. aka Pizza Cutter (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 Probability Added 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. aka Pizza Cutter (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. aka Pizza Cutter (2008). Wanted: players who like to run into things? https://statspeakmvn.wordpress.com/2008/05/page/2/

This was a heroic but obviously flawed attempt to see if breaking up double plays is a skill. Russell used 2004-2007 Retrosheet data for situations with a runner on first, fewer than two outs, and a ball fielded by infielder, aka double play situations, when runner on second out, thus completing the DP. Intraclass correlation indicating consistency across seasons was a non-existent 0.04. There is then no evidence for this proposed skill. In contrast, the intraclass correlation for the batter not out at first, aka beating out the throw from second, was 0.47. Speed was almost certainly why.

Carleton, Russell A. aka Pizza Cutter (2008). Ah, so we meet again... https://statspeakmvn.wordpress.com/2008/11

Russell Carleton (2008q) in his Pizza Cutter identity explained his method for measuring expected OBA in batter/pitcher matchups (which he did using Retrosheet data) as follows:
Step 1 – Convert batter, pitcher, and league OBA into an odds ratio (OR) by dividing it by (1 minus OBA).
Step 2 – (Batter OR divided by league OR) X (pitcher OR divided by league OR), which gives you an expected OR for the matchup relative to the league.
Step 3 – Convert back to OBA by (expected OR divided by [expected OR + 1]).

Carleton, Russell A. aka Pizza Cutter (2008). Power scores (or at least my attempt). https://statspeakmvn.wordpress.com/2008/05/

An attempt at something analogous to Bill James's Speed Score measure for baserunning, including players with at least 100 PA in seasons from 2000-2007. Russell performed a factor analysis that produced two factors, accounting for 59.7% of the total variance. The point is to see what metrics intercorrelate, and so as a group could be viewed as indicators of some basic batting skill.
Factor 1 – Homers per fly ball, extra base hits, and isolated power positive, percentage of balls in the air that made it to the outfield negative (surprise).
Factor 2 – BABIP, ground ball base hits, and line drive rate.
The first factor had an intraclass correlation of 0.740, which is quite good and indicates year-to-year consistency in what is clearly a representation of power. The second factor has an ICC of only 0.380, which is not surprising as it is based on indicators that are not as consistent as power from year to year.

Carleton, Russell A. aka Pizza Cutter (2008). Who gets the credit/blame for that home run? https://statspeakmvn.wordpress.com/2008/02/page/3/

This was based on 1993-1998 Project Scoresheet fly ball location data (which was approximate as it was based on scorer judgments) for batters with at least 25 fly balls and popups. The year-to-year correlation for fly ball distance was 0.612, implying consistency in power. For those with at least 50 fly balls and popups, the year-to-year correlations were 0.239 to pull side, 0.359 to the opposite field 0.359, and 0.591 to center field, which seems to imply that the last of the three is a greater indicator of power.

Carleton, Russell A. aka Pizza Cutter (2008). What can my glasses teach me about home runs? https://statspeakmvn.wordpress.com/2008/09/page/2/

In this webpost, Russell looked for statistical interactions in 2003-2007 (most likely Retrosheet) data for batters with a minimum of 250 Pas. He noted the following:
Overall, extra base hit rate was usually positively related with home run rate, but that relationship is moderated by contact rate, with the relationship stronger for players with higher contact rates.
Despite the above, overall home run rate related negatively with contact rate (power hitters tend to be free swingers), but this relationship was stronger for players with low swing rates.
Overall, home run rate was positively related with strikeout rate, with this relationship stronger for players with more pitches per PA.
Overall, more fly balls hit positively related with HR rate, with the relationship stronger for players with more pitches per PA.

Carleton, Russell A. aka Pizza Cutter (2008). The developmental curveball. https://statspeakmvn.wordpress.com/2008/08/page/2/

This is a study of within-season development of younger players. The sample was 99 batters with at least 100 PA in 2006 under the age of 26 on July 1st. Russell calculated strike zone sensitivity (his method for measuring plate discipline in the sense of swinging when, but only when, the pitch is a strike) for the player's first 50 PA of the season, and then repeatedly for each PAs 2-51, 3-52 etc. until the season ran out. This provided a string of moving averages for each player. Russell then performed a regression analysis across the moving averages for each. A positive regression coefficient means an increase in plate discipline as the season progressed, a negative coefficient a decrease, and the higher the amount of variance accounted for, the steadier the progress (or backtracking as the case may be). Overall, the coefficient correlated only 0.184 with the next season's (2007) strikeout rate. But for 20 of the 99 with regression line accounting for at least 30 percent of variance in moving averages, the correlation with 2007 was 0.525; for the 8 with over 50 percent of variance (he admits too small a sample size) it was 0.667. The point is that those who more steadily changed over 2006 were more likely to retain the change the next season.

Carleton, Russell A. aka Pizza Cutter (2008). Closers and non-save situations. https://statspeakmvn.wordpress.com/2008/08/page/3/

Russell Carleton (under his former alias Pizza Cutter) looked at all relievers with at least 15 saves between saves 1980 and 2007 (sample size 220 pitchers and 696 seasons) and compared basic metrics for when they pitched in designated save and non-save situations. ERAs were respectively 2.91 vs 3.15, OPS was .629 vs .652, walks per 9 innings were 3.08 vs 3.39 and strikeouts per 9 innings were 8.12 vs 7.79. Russell speculated that this may be due to closers primarily being used in non-save situations when they had not pitched for a while, and so the deficit would be due to rust. The more obvious explanation, adrenaline differences, is not as clearly indicated, as it should be more evident in comparing low vs. high leverage situations. These do not correspond well with the save situations; there is far higher leverage in a tie game in the 7th or 8th innings then in 9th innings with three run lead.

Carleton, Russell A. aka Pizza Cutter (2009). What really happens in the clutch. https://statspeakmvn.wordpress.com/2009/05/

2005-2008 data for late-inning pressure situations for batters with at least 50 PA per season in these events. The difference between swing percentage in and out of these situations had an intraclass correlation of 0.24. This implies some consistency in how batters respond to high leverage plate appearances, but there was a lot of variation across them; some players consistently swung more in the clutch, some less, some about the same. And given that low contact hitters who swing more often strike out more and hit fewer HRs, this may signal that those sort of pure sluggers perform poorly in the clutch.

Carleton, Russell A. aka Pizza Cutter (2009). The measure of a man, part 1. https://statspeakmvn.wordpress.com/2009/02/03/the_measure_of_a_man_or_10_things_i_didnt_know_about_you/

Carleton, Russell A. (2017). The secret powers of the foul ball. https://www.baseballprospectus.com/news/article/31412/baseball-therapy-the-secret-powers-of-the-foul-ball/

In a early post (2009) using 2008 data likely from Retrosheet, Russell noted that two strike fouls loaded positively on the same factor (correlated highly and positively with) his sensitivity score and contact rate whereas zero- and one-strike fouls loaded positively with his response bias score and negatively with contact rate. Eight years later (2017), based on 2016 data for non-pitchers and excluding intentional walks, when batters got to 0-1 counts, they were less badly off at the end of the plate appearances when they got to that count through a called strike (.229/.273/.359) or foul ball (.229/.272/.367) then by a swinging strike (.206/,255/.328). With two strikes, the ability to foul pitches off was also helpful thereafter; no subsequent fouls .170/.232/.263, one subsequent foul .194/.282/.310, more than one subsequent foul

.205/.308/.339.

Carleton, Russell A. aka Pizza Cutter (2009). If you're happy and you know it, get on base. https://tht.fangraphs.com/tht-live/if-youre-happy-and-you-know-it-get-on-base/

Looking at 2008 data for all batters with and pitchers faced at least 250 PA, and after controlling for their overall performance, the difference in OBA for batters when their team was winning versus losing was on average 7 points.

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 not predictors, and an increase in balls in play and home runs decreased 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). Saving the save. https://www.baseballprospectus.com/news/article/21557/baseball-therapy-saving-the-save

For 1993-2012, the innings in which the home team was pitching (that is, the tops of innings), the combination of innings and score differentials with the highest leverage scores.

Inning	Score Differential	Home Win Probability	Leverage
Top 9th	Up 1	86%	2.35
Top 9th	Tied	52%	2.05
Top 8th	Up 1	76%	1.93
Top 8th	Tied	53%	1.66

Top 9th	Up 2	94%	1.60
Top 7th	Up 1	72%	1.55
Top 9th	Up 2	89%	1.44
Top 7th	Up 2	84%	1.39
Top 7th	Tied	53%	1.36
Top 6th	Up 1	69%	1.34

Remember that all "9th" innings include extra innings as well.

The same when the visiting team was pitching (bottoms):

Inning	Score Differential	Visitor Win Probability	Leverage
Bot 9th	Up 1	82%	2.93
Bot 8th	Up 1	70%	2.29
Bot 9th	Up 2	92%	2.06
Bot 9th	Tied	33%	1.86
Bot 7th	Up 1	62%	1.83
Bot 8th	Up 2	84%	1.82
Bot 7th	Up 2	78%	1.54
Bot 6th	Up 1	57%	1.52
Bot 8th	Tied	36%	1.52
Bot 8th	Up 3	92%	1.41

Carleton, Russell A. (2013). Who has the momentum? And does it matter? https://www.baseballprospectus.com/news/article/21920/baseball-therapy-who-has-the-momentum-and-does-it-matter/

Russell Carleton (2013) examined whether teams that had played a lot of “crucial” games toward the season and then made it into the playoffs performed better in the postseason. He considered a game “crucial” if played in September by teams that had not clinched a playoff berth, were within 3 games either way of a playoff spot, and there was a playoff spot available for that team. Russell used play-by-play data for the last 15 games between 2003 and 2012 for matchups between batters and pitchers with at least 250 PA for each. He uncovered a little bit of evidence that teams that had better pitching outcomes in those games, particularly in terms of fewer walks, hit by pithces, and extra base hits, did better in the postseason assuming that they made it, perhaps in the order of a .20 or .25 runs per game advantage. There was no evidence for momentum effects for batting.

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. (2013). 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 Numbers* 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 (2013), 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. (2008). 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 (2013) 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. (2014). Do some pitches do more damage than others? https://www.baseballprospectus.com/news/article/24034/baseball-therapy-do-some-pitches-do-more-damage-than-others/

One possible predictor of pitcher injuries is added arm stress caused by pitching in high leverage situations. Russell Carleton (2014), based on a proposal contributed by Ben Flajole to the RotoScouting website, used 2002-2012 almost certainly Retrosheet data to try and find out. As always, the biggest predictors of elbow and shoulder injuries for starters were past elbow and shoulder injuries, respectively. However, pitch counts in situations in which an opposition home run would either tie the game or break the tie and give the other team the lead was the second most powerful predictor of shoulder injuries, before overall game pitch count. As Russell noted, it is very possible that those two variables actually reinfore one another multiplicatively in the “real” baseball world. Analogously, the previous season’s total number of such pitches were a stronger predictor than total pitches. High-stress pitches defined this way was not a significant predictor for elbow injuries. A more inclusive definition for high-stress as any coming when the score was tied or within a run also failed. This is a fairly crude method of measuring leverage, and a more exact measure of high-stress situation plus a more sophisticated model might find better evidence.

Carleton, Russell A. (2014). Do innings limits work?
https://www.baseballprospectus.com/news/article/23438/baseball-therapy-do-innings-limits-work/

Another article on what predicts pitcher injuries. Russel Carleton (2014), based on 2000-2013 data (likely Retrosheet), while controlling for previous injury, examined different measures of past pitching load beginning with age 19 (innings, games, batters faced, pitches thrown) at different ages until age 23 (this would mean pitching load between 19 and 22) and then every year up to age 28 (load between 19 and 27). For every year along the way, a DL trip due to shoulder injury was associated with (along with previous injuries as always) innings pitched. As for elbow injuries worthy of disablement, again along with previous injuries, innings pitched both overall and from the previous three or four seasons were predictors. The impact is not huge; an increase from 160 to 200 IP increases the overall injury risk by one percent at the most, and overall increases in IP did not have the negative impact of single game pitch counts above 120 even once. Additionally the impact of innings did somewhat ease up the innings pitched impact on elbow injuries as pitchers became older, either because of "survivor bias" (those who survived this long were less injury-prone to begin with) or the elbow area becoming stronger. Keep in mind that all of this says nothing about future pitching effectiveness.

Carleton, Russell A. (2014). I guess you just throw the next pitch.
https://www.baseballprospectus.com/news/article/24549/baseball-therapy-i-guess-you-just-throw-the-next-pitch/

What happens right after a pitcher gives up a home run? Based on 2009 to 2013 data almost certainly from Retrosheet, Russell Carleton, comparing the outcome of plate appearances right after opposition homers with other plate appearances with bases empty with suitable statistical controls, noted an overall seven point drop in OBA due to both more outs on balls in play and fewer walks, plus more strikeouts. Just allowing someone to get on base versus not also decreased performance a couple of points. Not surprisingly, the after-homer performance varied widely across pitchers, but did not correlate across seasons; in other words, it cannot be considered a skill.

Carleton, Russell A. (2014). Sure as day follows night…
https://www.baseballprospectus.com/news/article/24810/baseball-therapy-sure-as-day-follows-night/

Given that batters could well be tired when playing a day game after a night game whereas opposition starting pitchers would be relatively fresh, it would follow that offensive performance would decrease in that circumstance compared to other games. Russell Carleton (2014) studied the issue using (almost certainly Retrosheet) data from 2003 through 2013. It turned out that this was the case, with outs on balls in play going up and extra base hits going down, to the tune of six or seven OBA points. The day game performance of players who had not appeared in the night game were unaffected.

Carleton, Russell A. (2014). Why Joe Maddon matters. https://www.baseballprospectus.com/news/article/24988/baseball-therapy-why-joe-maddon-matters/
Carleton, Russell A. (2014). Against the grind. https://www.baseballprospectus.com/news/article/25065/baseball-therapy-against-the-grind/
Carleton, Russell A. (2015). The 10th man in the lineup. https://www.baseballprospectus.com/news/article/25647/baseball-therapy-the-10th-man-in-the-lineup/
Carleton, Russell A. (2015). A veteran and his presents. https://www.baseballprospectus.com/news/article/27388/baseball-therapy-a-veteran-and-his-presents/
Carleton, Russell A. (2015). The thirty-run manager. https://www.baseballprospectus.com/news/article/27388/baseball-therapy-a-veteran-and-his-presents/

Russell Carleton used Retrosheet pitch-level data from 2009 to 2013 to address two questions about the impact of managers on their team. In the first of these, Russell noted that in games immediately after losses, batters took more swings but had a lower contact rate (he failed to describe the size of the effect; I imagine that it was small). He proposed that this could possibly be a result of the psychological impact of that loss. Further, batter contact rate after losses differed among managers to the order of five percent. But managerial impact did not correlate from year to year, and without evidence of such stability, there is no reason to believe that he was tapping into a dimension of managerial skill. In the second of these, he uncovered evidence that the grind of the season had a similar impact, to the tune of about one percent. Again, managers differed in this effect, with the extremes differing by a significant 15 or so runs a year, and this time it correlated at .73 across them. Thus, we have a bit of evidence that managers differ by a game or two in their ability to help players cope through the long season.
Russell continued to examine the grind issue the following year. The following table is based on 2010 through 2014 data with batter/pitcher matchups controlled. The main diagonal displays the overall correlation for managers across seasons for players getting better/worse across season at the relevant skill; the other cells show correlations across the managerial tendencies. (A manager must have managed four of the five relevant seasons to be included in the consistency correlations and three of the five to be included in the cross-skill correlations):

	count goes 1-0	contact/swing	contact/swing (2 strikes)	balls/taken pitch	

count goes 1-0	.57	.062	-.001	.434	
contact/swing		.69	.889	.142	
contact/swing (2 ks)			.72	.043	
balls/taken pitch				.55	
pitch not a strike					

It seems as if there are unrelated managerial skills here, one helping or not helping player make contact when he swings, and the other helping or not helping players recognize when to take a pitch. (Russell included a fifth variable, pitches that are not a strike, but as this includes both contact and pitches correct taken it conflates the two skills). The difference between the best and worst managers amounted to about two wins in a season. Incidentally, in another study using the same data and method Russell uncovered no evidence that veterans (defined as players aged 35 or greater with at least 250 PA in a given season) had an impact on teammate's plate discipline, defined here as their teammates doing a better job of keeping pitches from being strikes, as the seasonal "grind" progressed.
Finally, using the same data for examining pitchers' ability to handle the grind, there was also evidence that managers differed in their ability to aid pitchers' ability to avoid inducing swings and making, contact, with correlations across seasons for managers relative to one another in the .6 range for both skills. Combining the two, the range in managerial ability in these added up to close to 4 wins at good versus bad extremes.

Carleton, Russell A. (2014). The timeshare DH. https://www.baseballprospectus.com/news/article/25107/baseball-therapy-the-timeshare-dh/

Is there a performance advantage to rotating the DH spot among different players rather than employing a full-time one? Based on 2009-2013 (almost certainly Retrosheet) data, Russell Carleton (2014) noted that, compared with playing in the field the day before, players that had DH'd the previous day had a few more singles and extra base hits, but their impact was mostly nullified by fewer walks, leaving a mere half a run surplus for this strategy over the course of a season.

Carleton, Russell A. (2015). How much the DH rule matters. https://www.baseballprospectus.com/news/article/27818/baseball-therapy-how-much-the-dh-rule-matters/

Of all the proposed explanations for the home field advantage, the two that have received the most support have been crowd support and travel. A third that may be relevant is the difference between leagues in the use of the designated hitter. Russell Carleton (2015) described the issue as follows: "The AL team is hurt significantly in the NL park by the loss of its DH. The NL team just plays its usual lineup. The NL team is hurt by the DH rule in the sense that the AL team has a guy who is a hand-in-glove fit for the role already on their roster, while the NL team can only match it by playing a bench guy." Russell was interested in seeing how great these advantages have been. Now, any DH-caused impact would be over and above the normal home field advantage, with Russell computed (almost certainly using Retrosheet data) as 53.7 percent and 0.11 runs per game between 2010 and 2014. To distinguish the two, Russell compared games when the team at home was playing A.L. versus N.L. teams. The A.L. teams at home scored 0.33 more runs when facing an N.L. team than an A.L., whereas N.L. home teams' advantage against the A.L. was only 0.06. Nonetheless, the average of the two (0.19) is much greater than the overall effect, implying that it should increase home field advantage. And indeed it did; the advantage in interleague games was two percent higher (55.5 percent for interleague games versus 53.5 for intraleague matches.

Carleton, Russell (2015). Going streaking.
https://www.baseballprospectus.com/news/article/27524/the-bp-wayback-machine-going-streaking/

This is more piece of evidence showing that streaks and slumps are pretty much random events. Russell Carleton (2015a) used a method for evaluating the existence of streakiness that had previously been used by several analysts; seeing if previous performance over a period of time or number of plate appearances affected the odds of getting a hit. The data (almost certainly from Retrosheet) encompassed 2000 through 2009 data for plate appearances between batters with and pitchers facing at least 250 plate appearances that season (the study was actually performed in 2010 but re-posted five years later), Russell determined that the probability of an "on-base event" given the batter's OBA for his previous 10, 25, and 100 PAs when controlled for batter and pitcher OBA barely changed. The difference in odds amounted to one more on-base event for every "few thousand" (Russell was not more specific) PAs for hot versus cold hitters. In other words, the difference had no practical significance. Always keep in mind that this could mean either that (1) streaks and slumps are totally random events with no real cause or that (2) streaks and slumps have real causes that crop up randomly.

Carleton, Russell A. (2015). Why do pitchers get tired.
https://www.baseballprospectus.com/news/article/27517/baseball-therapy-why-do-pitchers-get-tired/

Based on 2010-2014 (probably Retrosheet) data, Russell Carleton (2015b) uncovered no evidence that the leverage of the situations pitchers faced had any impact on their performance over and above the pitch count.

Carleton, Russell A. (2015). Say you'll remember me.
https://www.baseballprospectus.com/news/article/27723/baseball-therapy-say-youll-remember-me/

Based on 2003-2014 (probably Retrosheet data), Russell Carleton (2015c) found no evidence that a reliever's performance in one outing carried over if he pitched again the next day.

Carleton, Russell A. (2015). Seven months has gone so fast. https://www.baseballprospectus.com/news/article/27461/baseball-therapy-seven-months-has-gone-so-fast/

Carleton, Russell A. (2016). Stop blaming the September call-ups. https://www.baseballprospectus.com/news/article/30306/baseball-therapy-stop-blaming-the-september-call-ups/

Carleton, Russell (2016). The 26th man. https://www.baseballprospectus.com/news/article/30763/baseball-therapy-the-26th-man/

Russell Carleton (2015 and two from 2016), using what was almost certainly Retrosheet data, examined the extent to which fringe players were used during different months. Not surprisingly, given all of the call-ups, the proportion was much greater in September. The following table shows the percentage of PAs by month by position players who had, or pitcher who faced fewer than 250 PAs and 100 PAs that season (2011-2015 data):

Month	**Position Players**		**Pitchers**	
	<250 PAs	<100 PAs	<250 PAs	<100 PAs
March/April	17.5%	7.2%	9.2%	6.6%
May	16.8%	6.5%	6.1%	4.4%
June	14.9%	6.1%	6.0%	3.3%
July	14.7%	5.7%	4.7%	3.0%
August	18.0%	6.0%	5.6%	3.3%
September/ October	23.6%	10.1%	12.00%	9.6%

Their performance does not change much and in fact that for pitchers improves, which might imply that called-up pitchers are better than those at the end of the bullpen during the previous months:

Month	**OBP of < 250 PA Batters**	**OBP Allowed by <250 PA Pitchers**
March/April	.276	.366
May	.278	.377
June	.270	.367
July	.280	.380
August	.287	.357

September/October	.286	.345

In 2015, the average number of pitchers used per game went up from 4.06 to 4.62 and of pinch-hitters from 0.68 to 1.17 in A.L. parks and 1.70 to 2.25 in N.L. parks when moving from August to September/October. However, teams vying for the playoffs (defined as within three games either way of a playoff spot when they were still in the hunt) were much less likely to use September callups (position players, 2.4 percent of PAs; pitchers, 4.4 percent of PAs) than those not (position players, 6.3 percent of PAs; pitchers, 8.7 percent of PAs), and teams not vying for the playoffs were a bit less likely to use September callups when playing a team in the hunt (position players, 4.7 percent of PAs; pitchers, 7.1 percent of PAs) than when not (position players, 5.5 percent of PAs; pitchers, 7.7 percent of PAs).

Carleton, Russell A. (2015). No relief for starters.
https://www.baseballprospectus.com/news/article/27131/baseball-therapy-no-relief-for-starters/

Russel Carleton (2015), using 2003-2014 (likely Retrosheet) data, isolated those relief pitchers who had at least five appearances in which they faced more than twelve batters and another five in which they faced six or fewer within a year. It turned out that in the short appearances, strikeout rate was a significant 1.5 percentage points higher than in the long outings; walks were also up but singles were down, resulting in pretty much the same OBA. In short, it appears that relievers change strategies depending on the expected length of their appearance.

Carleton, Russell A. (2015). The most important player on the field.
https://www.baseballprospectus.com/news/article/25922/baseball-therapy-the-most-important-player-on-the-field/

Carleton, Russell A. (2015). Chopping up the credit.
https://www.baseballprospectus.com/news/article/25989/baseball-therapy-chopping-up-the-credit/

In two posts, Russell Carleton used the amount of variation among batters, pitchers, and other relevant players in several performance indices to estimate the proportion of responsibility each should be given for the results of plate appearances. The following are based on 2010-2014 Retrosheet data:

Event	Batter	Pitcher	Catcher	Random Noise
Strikeout	63.5%	35.3%	0.1%	1.1%
Walk	64.2%	33.9%	1.0%	0.8%
HBP	65.7%	32.3%	1.7%	0.3%
Grounder	60.6%	39.2%	0.1%	0.04%
Line Drive	54.5%	32.2%	1.4%	11.8%
Fly Ball	42.6%	51.2%	—	6.2%

HR given FB	78.8%	19.7%	—	1.6%

The following proportions measure responsibility for the hit-versus-out outcome of batted balls in play, this time based on 1993-1999 data from Retrosheet with hit location information:

Event	Batter	Pitcher	Random Noise
Fly Balls	48.1%	32.5%	19.4%
Ground Balls	46.6%	40.6%	12.7%
Line Drives	18.7%	20.9%	60.2%

Russell emphasized the point that these are probabilities over the multi-year samples of these events, and say nothing about the responsibility for the outcome of any given PA. Those would be in the six to seven percent range for a True Outcome event and something like one percent for a batted ball in play. His conclusion: McCracken's claims that pitchers have no control over batted balls in play was half right. Neither pitchers nor batters are responsible for the outcome of a given plate appearance.

Carleton, Russell A. (2015). The credit card game.
https://www.baseballprospectus.com/news/article/26535/baseball-therapy-the-credit-card-game/

Based on the variance accounted for by each in Retrosheet data (not sure which years, probably 2010-2014), Russell Carleton (2015) estimated (1) that responsibility for the outcome of stolen base attempts should be divided up 45 percent to the runner, 49 percent to the pitcher, and six percent to the catcher; (2) that the runner should receive 34 percent, fielder 20 percent, pitcher 21 percent, batter 14 percent, and unknown factors 12 percent of the responsibility for attempts at getting extra base on hits in play or advancing on flyouts; (3) as for success at these attempts, 29 percent go should be assigned to the runner, 23 percent to the fielder, 24 percent to the pitcher, and 20 percent to the batter, leaving 4 percent unknown, and (4) that the pitcher should be assigned almost all of the responsibility (94%) for passed balls and wild pitches, leaving just six percent to the catcher. Official scorers' decision between the two are pretty close to that split; 83 percent wild pitches versus 17 percent passed balls.

Carleton, Russell A. (2015). Collage or jigsaw?
https://www.baseballprospectus.com/news/article/27253/baseball-therapy-collage-or-jigsaw/
Carleton, Russell A. (2015). Interaction effects and credit.
https://www.baseballprospectus.com/news/article/27309/baseball-therapy-interaction-effects-and-credit/

Russell Carleton (2015) did a nice job of suggesting a type of team interdependence. The difference between the best and worst infields in successfully fielding ground balls is not huge; the 2014 range was 76.1 percent (A's) to 69.8 percent (Rays). Nonetheless, that is enough to impact on groundball pitchers. Dallas Keuchel, who was the most extreme groundball pitcher among those pitching 162 innings that year, was tops at 63.5 percent) would be estimated to get 24 more outs on ground balls per 600 PAs if he were to have pitched for the 2014 A's as compared to the 2014 Rays. In contrast, the most extreme fly ball pitcher (Chris Young at 22.3 percent grounders) would only have gained 9 outs. Although Russell did not mention it, the corresponding point is that Young would have benefited more from a very good outfield threesome than would have Keuchel. The point is the interdependence between pitcher type and team fielding capability. In a follow-up, Russell uncovered a tiny (i.e. untrustworthy) bit of evidence suggesting that at least some pitchers might try to pitch to contact in front of a good infield defense, but if so the total impact might be two or three more grounders a season.

Carleton, Russell A. (2015). Hit the pitcher eighth?
https://www.baseballprospectus.com/news/article/26045/baseball-therapy-hit-the-pitcher-eighth/

Russell Carleton (2015) looked at historical trends concerning pitchers batting eighth and discovered that it was not a new strategy, having occurred 66 times in 1957, but it became rare after that. Based on 2014 (I am sure Retrosheet) data, he simulated 100,000 games in which he traded the 8th and 9th place hitters' performance. The verdict: 3.7118 with the pitcher batting eighth, 3.7079 with the pitcher batting ninth. Also, further in the 7th inning, when N.L. starting pitchers still in the game only hit for themselves 27.1 percent of the time, the 8th spot was two percent more likely to come up than the ninth position, which means that the team is more likely to have to blow a pinch hitter with the pitcher batting eighth. Batting the pitcher eighth does not look like a particularly helpful strategy.

Carleton, Russell A. (2015). Should they pitch to the eighth hitter?
https://www.baseballprospectus.com/news/article/26123/baseball-therapy-should-they-pitch-to-the-eighth-hitter/

Assume that there are runners in scoring position and two outs in the sixth inning or later, with the #8 batter up and the pitcher batting #9. 2010-2014 (almost certainly Retrosheet) data examined by Russell Carleton (2015) revealed that intentional walks occurred 36.8 percent of the time, plus the walk rate when the eighth sport was pitched to (11.8%) was far higher than the 6.9 percent walk rate with bases empty, suggesting that there were quite a few "unintentional-intentional" walks.
Russell then looked at the implications of the strategy; in short, it is complicated (I supply a summary; see the article for details). The team in the field ended up a bit ahead for that inning by walking the #8. But there's the next inning; if you pitch to #8 there is a greater chance that he won't make the third out than for a pitcher batting at #9 after an IBB. This means that with an IBB there is a greater chance that the #1 batter

will lead off the next inning rather than #9. This will cost the team in the field a bit. So combining all of these possibilities, with a runner on third only and two outs the team in the field would have lost a tiny 0.04 runs by walking an average hitting #8, making the strategy a wash in general and dependent on how good a hitter #8 is. However, with runners on second and third and two outs the team on the field gained 0.10 runs, which is enough to make the IBB a good overall strategy in that situation. And teams seemed to have realized it; the IBB rate was 56.6 percent in that base/out situation with a 15.3 percent walk rate otherwise (meaning a ton of "unintentional/intentionals").

Carleton, Russell A. (2015). Are you over 18?
https://www.baseballprospectus.com/news/article/26325/baseball-therapy-are-you-over-18/

Russell Carleton (2015) examined changes in the outcomes of plate appearances as games progress, based on the outcome of each PA in 2010-2014 games (Retrosheet data for sure) in which both batter had, and pitcher had faced 150 PAs for the season. , Relative to expectations, outcomes in which the ball is not put into play (strikeouts and walks) went down over the game, whereas outcomes in which the ball is put in play (singles, doubles, triples, homers, outs on balls in play) all went up. Despite the fifth of the latter, BABIP went up; despite the decrease in walks, OBA went up too. At the level of individual pitches with controls for batter and pitcher tendencies, as games progressed, there are more swings, a higher contact rate on swings, and a lower likely of the batter taking a called strike.
Russell's actual intent in this post was to examine the reason for these changes. A problem with the "times through the order penalty" concept is that it conflates two different explanations for pitcher ineffectiveness later in the game; batters' increased familiarity with the pitcher's repertoire and pitcher fatigue. Russell wanted to distinguish between the two as much as possible by, at the beginning of each PA, seeing whether the time through the order (ordinal number; 1, 2, or 3) or the pitch count was a better predictor of PA outcomes. It turned out that the changes in strikeouts, doubles, and triples were more statistically dependent on times through the order, whereas those for walks, singles, and homers were statistically dependent on pitch count. This sure seems random to me. Russell also discovered that, as the two were very highly correlated, once one predictor entered his regression equation there was almost no variance left for the other, which means that there is no way to discern the relative amount of variation in PA outcome. In other words, pitch count and times through the order are what statisticians call "multicollinear," which means that the overlap between the variance accounted for by each is so large, and correspondingly the variance each accounts for separately is so low, that attempts to separate the two are invalid. Russell's conclusions bear this out; he believed that the loss of pitcher effectiveness is due both to pitcher fatigue and batters attempting more seriously to put balls in play. In any case, if a reliever is in by the third time through the order, all of these effects disappeared.

Carleton, Russell A. (2015). Can a manager "win" spring training? https://www.baseballprospectus.com/news/article/25795/baseball-therapy-can-a-manager-win-spring-training/

Using 2010-2014 (probably Retrosheet) data, Russell Carleton (2015) found no evidence that managers showed consistency across seasons in their ability to "get their team ready for the season," in the sense of whether or not pitches are strikes during April.

Carleton, Russell A. (2015). The clock is ticking… https://www.baseballprospectus.com/news/article/25603/baseball-therapy-the-clock-is-ticking/

Russell Carleton (2015) tried to estimate the impact of a strategic choice inherent in the concept of a pitch clock; if a pitcher is not ready to throw and the clock is winding down, whether to throw anyway or accept the automatic ball. He started with the following 2014 (certainly Retrosheet) data on the run expectancy "penalty" for a ball call at each count (RE is short for run expectancy):

Count	Run Ex.	Count	Run Ex.	Count	Run Ex.	Count	Run Ex.
0-0	.0314	1-0	.0521	2-0	.0940	3-0	.1194
0-1	.0228	1-1	.0445	2-1	.0908	3-1	.1743
0-2	.0200	1-2	.0367	2-2	.0886	3-2	.2372

Russell presumed the following strategic choice: Let us assume that making a bad pitch means an extra base hit of some type. The pitcher needs to estimate those odds and how much they would cost him. Now, the odds of an extra base hit on a pitch in 2014 was 1.9 percent, with each extra base hit worth an average of 1.06 runs. Multiply the two together, and you get a run value of almost exactly .02 for making a bad pitch. So basically the decision is whether that cost is higher or lower than taking the automatic ball. If we could trust the data table, the only count in which accepting the automatic ball is acceptable would be 0-2. The problem is that we cannot trust the data table. Russell realized that the 1.9 is likely an underestimate of what the odds for an extra base hit are when a pitcher is not psychologically ready to pitch, but there is no data to work with on that, and I cannot imagine how there could be. As he also noted, the estimate should also include damage from singles and hit-by-pitches (he forgot about wild pitches). But the general point holds; the pitcher has to choose between the damage for accepting an automatic ball and the damage for making a pitch that the batter clobbers.

Carleton, Russell A. (2016). Should someone save Salvy? https://www.baseballprospectus.com/news/article/28606/baseball-therapy-should-someone-save-salvy/

Using (likely Retrosheet data) for 2011-2015, Russell Carleton (2016) uncovered no evidence that days off had an impact on the performance of catchers, in terms of batter outcome, baserunning steal attempts or success rate, or blocking balls. Days off might help a bit on offense, as days off seemed to help catchers maintain their plate discipline across seasons better than other players. Russell reminded us that these findings should not be interpreted as suggesting that catchers never need a day off, but rather that as a group managers do not appear to be overworking them.

Carleton, Russell A. (2016). Can teams come back from a comeback?
https://www.baseballprospectus.com/news/article/29023/baseball-therapy-can-teams-come-back-from-a-comeback/

We know that strategically the current standard reliever usage patterns (fixed roles for pitchers including "official" ninth inning closer) is substandard because a team's best reliever(s) should be used in the highest leverage situations no matter the inning. However, there is a psychological argument in its favor, that teams should save their best reliever until the end because a ninth-inning lose is the most upsetting to the players are carries through to poorer performance the next day. Russell Carleton (2016) demonstrated that this is not true Based on 2000 to 2015 data (likely from Retrosheet), team winning averages the next game after a ninth inning comeback lose were actually better (.481) than when losing a lead in the eighth (.466) or seventh (.442). If anything, this is an argument for team's using their best reliever earlier than the ninth.

Carleton, Russell A. (2016). The pink elephant effect.
https://www.baseballprospectus.com/news/article/29868/baseball-therapy-the-pink-elephant-effect/

Applying 2011 to 2015 (probably Retrosheet) data, Russell Carleton (2016) examined whether replacing one pitcher with a second who is demonstrably different impacted on batter performance in the first PA against the new pitcher. When compared with overall batter and pitcher performance, there was no such impact for pitcher handedness beyond the standard platoon differential or in pitcher strikeout tendencies, but putting in a reliever with ground ball tendencies actually decreased the odds of a first-PA ground ball.

Carleton, Russell A. (2016). What would 7-inning baseball look like?
https://www.baseballprospectus.com/news/article/30023/baseball-therapy-what-would-7-inning-baseball-look-like/

At the time that Russell Carleton (2016) speculated about the consequences of seven-inning major league games, it was a mere possibility; of course, it became fact in 2020. Much of what he wrote was imaginative speculation, but here I will stick with those ideas which have relevant data. Pitcher usage patterns would probably change.

Here are per-inning RAs for the 3990 games in 2015 in which starters went at least five innings:

Inning	1st	2nd	3rd	4th	5th	6th
ERA	3.74	2.78	3.06	3.12	3.17	3.89

Looking beyond the fact that these are well-pitched games as a group, note the big decrease in effectiveness during the sixth inning. Now here is Russell's cut-and-pasted table of RAs by relievers that season:

Inning	All Games	Game within 3 runs
6^{th}	5.68	5.90
7^{th}	5.03	4.92
8^{th}	4.09	3.92
9^{th}	3.73	3.42

Note how much it improves beginning in the eighth inning. This is of course because most managers save their best relief pitchers for the late innings. Implication: there is good reason to consider pulling a starter after five innings and, in particular, after six, but the weaker relievers can now be bypassed.

All of leads to some speculations: starters might have less concern with saving themselves and air it out earlier, innings eaters have less value and guys good for only two times through the order more value, alternate starter rotations (such as four-man with three days rest or with four day rest plus spot fifth starter) seem more viable, and smaller pitching staffs with more position players and, with them, greater platoon possibilities. But this would be counteracted by the fact that there would be more extra-inning games. In 2015, extra inning games occurred 8.7 percent of the time, but the score was tied 12.8 percent of the time after seven innings. So you would need all of those extra relievers after all.

There also could be changes in who gets to play. Here are the 2015 percentage of PAs for each lineup position for the first nine and first seven innings:

Lineup Position	1	2	3	4	5	6	7	8	9
9 innings	12.3	11.9	11.7	11.4	11.1	10.9	10.6	10.2	9.9
7 innings	12.6	12.2	11.8	11.5	11.1	10.8	10.4	10.1	9.6

Note that in the first seven innings of games, the top-of-the-order aka better hitters comes up relatively more often and the bottom-of-the-order aka weaker hitters relatively less. Russell speculated that this could lead to a bit of an incentive to play glove guys at some positions as they would bat less often at the bottom of the order and ironglove sluggers at other positions as they would bat more often at the top of the order than in a nine-inning game.

Carleton, Russell A. (2016). Bullpen contagion.
https://www.baseballprospectus.com/news/article/30464/baseball-therapy-bullpen-contagion/

Between 2003 and 2015, based on what is likely Retrosheet data, Russell Carleton (2016) searched for impacts of blown saves on subsequent games, but only uncovered an increase in walks of about one percent for a game or two afterward and nothing for any other index.

Carleton, Russell (2016). The 26th man.
https://www.baseballprospectus.com/news/article/30763/baseball-therapy-the-26th-man/

In 2015 (Carleton, 2016; Retrosheet data) gave us the following PAs in which the pitching team had the platoon advantage in 2015 during games in which that team was ahead by three runs or less, tied, or one run behind, with a reliever on the mound:

Inning	All hitters	Left-handed hitters
6	60.0%	35.6%
7	57.9%	34.9%
8	55.6%	26.8%
9	49.3%	11.8%

This seems to imply that lefty specialists are most likely used in the sixth and seventh and that traditional closer and setup pitchers are tilted toward the right.

Carleton, Russell A. (2016). The dark side of pitch framing?
https://www.baseballprospectus.com/news/article/28350/baseball-therapy-the-dark-side-of-pitch-framing/

Not surprisingly, the better the pitching framing, the more strikeouts and thr fewer walks. Russell Carleton (2016a) examined other consequences using (I assume Retrosheet) data from 2006 throught 2015 meeting the following criteria:

> · A pitcher who had faced at least 250 batters in the season in question faced off against a batter who came to bat at least 250 times that season.
> · The pitcher had not switched teams during the course of the season. This allows us to rule out cases where a pitcher was traded from a good catcher to a bad catcher (or vice versa), but where he also went to a less (or more) favorablepark or was put in front of seven other guys who were better (or worse) at fielding.

· The pitcher and catcher in question worked together for at least 100 batters that year.
· The pitcher worked with two different catchers for at least 100 PA that year. One of them was in the top fifth of the league in BP's called strikes above average (CSAA) metric. The other was in the bottom fifth.

There was no impact on any component of BA, SA, or BABIP, yet at the same time groundball rate was almpst 2 percent higher for the pitchers with good framers as opposed to poor, implying an additional four of five runs a year saved on top of that for the extra strikes. Russell realized that these two findings were incompatible with one another. He did not have a good explanation for this discrepancy, other than speculating that a pitcher knowing he had a good framer behind the plate might take more risks in their pitch selection, resulting in harder hit balls.

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 at 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

Carleton, Russell A. (2017). Taking the weirdness out of baseball. https://www.baseballprospectus.com/news/article/31175/baseball-therapy-taking-the-weirdness-out-of-baseball/

Russell Carleton (2017) examined scoring in extra innings with an eye on seeing the potential impact of beginning each with a runner on second. Between 2012 and 2016 (likely Retrosheet data) 8.8 percent of games (1064 in total) went into extra innings, with an average of 2.3 extra frames and a breakdown as follows:

Inning	10	11	12	13	14	15	16	17	18	19	20
%	43.3	24.1	13.9	8.6	5.2	1.8	1.4	0.4	0.7	0.6	0.1

The following cut-and-pasted table first shows the average number of runs scored in extra innings. If you compare this with overall figures for runs scored per inning, you will see that the chances of scoring at all are not much different, but the odds of scoring only one run have increased at the expense of higher numbers. This is almost certainly because if the home team scores one the game ends, denying them the unneeded chance of adding to it. The second and third columns display what happened in the 329 instances in which the leadoff batter in an extra inning reached second base:

Number of Runs	No runners, No outs (extra innings only)	Runner on second, No outs (extra innings only)	Runner on second, No outs (top of extra innings only)
0	73.9%	43.8%	39.7%
1	17.4%	36.5%	27.0%

2	5.3%	12.2%	19.5%
3	1.9%	2.7%	5.2%
4	0.9%	3.0%	5.2%
5	0.3%	1.5%	2.9%
6+	0.3%	0.3%	0.6%

Russell then performed a simulation to estimate the number of innings that games would require to end if they began with a runner on second; compare these to the earlier table:

Inning	10	11	12	13	14	15	16	17	18	19	20
%	70.5	20.8	6.1	1.8	0.5	0.2	<0.1	<0.1	<0.1	<0.1	<0.1

This would drop the average of extra innings from 2.3 to 1.4.

Carleton, Russell A. (2017). Is win probability broken?
https://www.baseballprospectus.com/news/article/31219/baseball-therapy-is-win-probability-broken/

Using what is almost certainly Retrosheet data from 1993 to 2016, Russell Carleton (2017) determined that win probability estimates can be improved by an average of three percentage points, and in about ten percent of the cases ten points, by including average runs scored and given up per game for both teams in the calculation. The more equal the teams were in quality and the later the inning, the less adjustment the classic win probability figures needed.

Carleton, Russell A. (2017). Baseball needs some new words.
https://www.baseballprospectus.com/news/article/31259/baseball-therapy-baseball-needs-some-new-words/

The following table cut and pasted from Russell Carleton (2016e) displays the proportion of starter and reliever pitch counts for each ten percentage point category in 2016.

Pitch Count	Starters	Relievers
0-9	0.1%	20.1%
10-19	0.1%	47.1%
20-29	0.3%	22.1%
30-39	0.6%	7.0%
40-49	1.0%	2.2%
50-59	2.0%	0.8%

60-69	3.8%	0.4%
70-79	7.3%	0.1%
80-89	17.7%	0.1%
90-99	31.5%	0.0%
100+	35.6%	0.0%

Russell wanted us to note the gap between 40 and 70 pitches, which indicates the absence of long relief outings despite the likelihood that there are some failed starters who could probably fit this role successfully.

Carleton, Russell A. (2017). Is defense slump-proof?
https://www.baseballprospectus.com/news/article/31307/baseball-therapy-is-defense-slump-proof/

Russell Carleton (2017) studied hit location data for infielders (not including first base) with at least 250 grounders hit in their territory from the 1993 to 1997 version of Project Scoresheet found in Retrosheet and concluded that there was a small but noticeable impact for the last 10 chances on the odds of fielding the eleventh chance for an infielder beyond the infielder's normal odds of success. In short, there was a bit of evidence in favor of fielding being susceptible to streakiness. It was unrelated to the infielder's recent batting success. I would caution the reader, however, that Project Scoresheet fielding data was observer-based and there could be some judgment bias at work here.

Carleton, Russell A. (2017). Bring back ball four.
https://www.baseballprospectus.com/news/article/31360/baseball-therapy-bring-back-ball-four/

Using what must be Retrosheet data from 1993-2016, Russell found that the average Leverage Index for a non-eighth-spot-in-the-lineup (to get to the pitcher) intentional walk is 1.4.

Carleton, Russell A. (2017). The disappearing left fielder.
https://www.baseballprospectus.com/news/article/31686/baseball-therapy-the-disappearing-left-fielder/

The thirty left fielders who had the most PAs had fewer of them as time went on, from 69.3 percent in 2006 to only 56.2 percent in 2016. Here are all the positions in 2016, note that LF is even less than C:

Position	Number of Players	Percentage of PA by “starters”
LF	241	56.2%
C	104	60.7%
RF	206	67.0%
CF	157	67.4%
3B	160	71.1%
1B	170	72.7%
2B	152	74.1%
SS	121	78.1%

Here are the “normal” positions of the replacement left fielders:

Primary Position	2016 (n = 181)	2006 (n = 136)
C	1.7%	0.7%
1B	9.4%	11.0%
2B	11.0%	8.8%
3B	7.2%	1.5%
SS	6.6%	2.9%
LF	26.0%	19.1%
CF	18.8%	25.0%

RF	19.3%	30.9%

Note now few of them are strictly substitute left fielders, and how many of them are now infielders, the latter implying that management now believes that almost any competent players outside of catchers can play LF. One more table: percentage of PAs without platoon advantage:

Position	Same Hand
LF	41.1%
1B	41.6%
CF	43.9%
RF	46.2%
2B	46.5%
SS	49.2%
C	55.2%
3B	57.9%

Implies that left fielders are replaced/platooned the most. The reason for this is that, relative to other positions, left (and right) fielders have had HR rates that have been declining (or in recent years increasing more slowly) than other positions, making it more attractive to use players from other positions.

Carleton, Russell A. (2017). Do strikeouts spread?
https://www.baseballprospectus.com/news/article/31783/baseball-therapy-do-strikeouts-spread/

Using 2012-2016 for batters with at least 250 PAs per season, Russell Carleton looked at the percentage of teammates at least 1 standard deviation above or below mean in strikeouts and walks. The more high strikeout teammates, the more likely a player on that team will have increased strikeouts from previous season. The same relationship existed for low numbers of walks, but not for fewer strikeouts or more walks. This could mean that bad plate discipline and contact issues are contagious. As Russell says here, it could also be a product of team philosophy.

Carleton, Russell A. (2017). How long can you keep a secret? https://www.baseballprospectus.com/news/article/31891/baseball-therapy-how-long-can-you-keep-a-secret/

The transition into what was standard closer usage during the 2010s (and perhaps beyond, we shall see), three outs in the ninth inning only, took several years. Between 40 and 50 percent of what are now defined as saves were the 2010s standard between 1950 (and perhaps earlier) and 1987, when Pete Rose used John Franco pretty much exclusively as a three-out pitcher. A decade later, the percentage had increased to about 75 percent, and after about one more decade 90 percent, where it stayed at least through 2016. By 2010, at least 75 percent of saves were ninth inning only for every team (data likely from Retrosheet and calculated by Russell Carleton, 2017).
In the same entry, Russell came up with a nice indirect way to examine the onset of team concern with catcher framing; the standard deviation across teams in framing runs as estimated from Retrosheet data for 1988 through 2007 and other sources from 2008 (when framing runs were first calculated by sabermetricians) through 2016. It was as low as 5 and rarely higher than 10 through 2006, but then increased to consistently above 15 and as high as about 21 in 2011. This is of course speculative but it is indirect evidence that some teams had become aware of the issue and so were going after catchers who were known to be good framers whereas other teams remained ignorant or uninterested. After that, the s.d.'s starting decreasing and were back down to about 14 in 2015 and 2016. The again speculative implication is that more teams began taking catcher framing seriously, lowering the variation among teams. A corollary of this is that with less variation among teams, the relative advantage of teams in the know has lessened, taking with it the relative value of framing itself.

Carleton, Russell A. (2017). The Justin Smoak problem. https://www.baseballprospectus.com/news/article/32359/baseball-therapy-the-justin-smoak-problem/

Based on a sample of both position players and from 2012 to 2016 who had at least 250 PA one season and 300 PA the next, Russell Carleton (2017) determined that there was no difference in how predictable performance was in PAs over 250 for the second season for strikeouts, walks, singles, double/triples, homes, or outs in play, whether the predictors were performance earlier during 250 PA stretches in that second season or the first season. For pitchers, this was the case even when the predictors differed more than 10 percent between seasons. In other words, there was no evidence for sudden changes in performance over two seasons.

Carleton, Russell A. (2017). The case of the missing fireman. https://www.baseballprospectus.com/news/article/32480/baseball-therapy-the-case-of-the-missing-fireman/

On average, pitchers who earned saves required five outs to get them from 1950 through the late 1980s, and then nosedived to three-and-a-half in only five years before diving further to three around 2010. Overall, relievers were in for about four outs until the mid 1980s, when it collapsed to three in fewer than 10 years and was at maybe 2 2/3 on average by 2016. There was not an associated increase in the percentage of, in Russell's term, “save-worthy” i.e. lead of three or fewer runs until about 2010, when it increased by a couple of percentage points., and the half-a-run per inning and .03 strikeouts per PA advantages for relievers over starters came about in the mid 1970s, before the strategy change. The proportion of relief pitchers who pitched at least 70 innings in relief was well over half until the 1960s, as many were also spot starters, and then drifted steadily down to about 10 percent by 2016. In the 1950s, pitcher received an average of 3.6 days of rest between appearances no matter how long their outing from 1 through 9 outs recorded. In the 1960s, the average dropped to 3.4, again no matter how many outs. However, starting in the 1970s, days between appearances began becoming dependent on number of outs achieved, such that by the 2010s three outs meant about 2.7 days off, six outs about 3.3, and nine outs about 4.4. There was a stretch in the 1970s in which pitchers who recorded 4 or more outs pitched the next day abut a quarter of the time, which was about .04 more often than earlier and far more often than later, with the figure in the mid teens by the 2010s. In contrast, a 4-out appearance followed by 2 days off actually increased from about .4 to .5 of the time during the entire duration. Incidentally, the proportion of 27-out starts was at about .35 in 1950 and went steadily down to maybe a tenth of that by 2016.

Carleton, Russell A. (2017). Whether to Waxahachie. https://www.baseballprospectus.com/news/article/32527/baseball-therapy-whether-to-waxahachie/

The “Waxahackie Swap” is a strategy credited to Paul Richards (who was from Waxahachie Texas; the name was suggested by a reader of Rob Neyer's column) in which a right-handed pitcher goes to left field so that a left-handed pitcher can face a lefty batter or two, after which, the righty returns to the mound and a substitute outfielder takes over in left. Russell Carleton (2017) thought out some of its advantages and disadvantages. On the good side, it would give you a platoon advantage worth .025 runs per plate appearance. On the bad side, if the batter hit a fly ball to left, then you would probably have a sub-par fielder trying to grab it, but from 2012-2016 only 8.7 percent of PAs ended with flys or liners to left. Russell believed that the difference between an average and bad left fielder was .02 runs per inning, so he guesstimated that the overall risk on a given play was .005 runs. In addition, a typical substitute left fielder is about .02 runs worse than a typical starter per PA, which along with the fielding issue just mentioned wipes out the advantage. So the Waxahachie swap would only be worthwhile if it were unlikely that the left fielder's spot in the batting order comes up again, or if the manager can use it for more than one opposing batter.

Carleton, Russell A. (2018). Reimagining the defensive spectrum. https://www.baseballprospectus.com/news/article/41948/baseball-therapy-reimagining-the-defensive-spectrum/

Carleton, Russell A. (2018). The “tell him, Wash” theory of WAR. https://www.baseballprospectus.com/news/article/42103/baseball-therapy-the-tell-him-wash-theory-of-war/

Russell Carleton (2018) made an interesting point about these sort of analyses as part of his work on “emergency players” discussed just below; that there is a hidden assumption here that a player can be moved from any position to any other position with a given run penalty as represented by the figures in those adjustments. But this just isn't true. As he wrote, except in rare and desperate situations, “The only guy who’s going to replace a catcher is a catcher. The only guy who’s going to replace a shortstop is a shortstop. The same seems to go for second and third basemen. “ Further, for a player with little experience at a position, here will be a period of time getting used to the special skills needed to competently play it, such as first baseman (the lowest on the totem pole) picking errant throws out of the direct. So players from higher up in the Defensive Spectrum should be expected to be worse than players already there when first moving there.

In order to evaluate this point, Russell (2018) used 1993-2017 data (I'm guessing from Retrosheet) for players with at least the equivalent of 150 games at a position trying another position. for at least the equivalent of five games. The figures in the last two column are for the first five games for those new at positions who subsequently played it either part-time or full-time – Russell surprisingly did not define the differences between the two – with the with number in parentheses signifying how long on average it took players to reach league average.

Position	Play	League Average	New; Part-time	New: Full-time
First base	Balls hit in “territory”	73.5%	72.2%	77.0% (0)
First base	Throws from other infielders	96.3%	95.9%	96.4% (0)
Second base	Grounders hit in “territory”	68.2%	74.1%	68.7% (0)
Second base	“Throw out rate” on caught grounders	97.7%	97.9%	96.9% (31)
Second base	Line drives caught	8.7%	7.4%	7.3% (28)

Second base	Double plays turned	65.3%	51.6%	67.3% (8)
Third base	Balls hit in territory	78.0%	79.2%	81.3% *
Third base	"Throw out rate" on caught grounders	94.4%	91.5%	91.8% (46)
Third base	Catching line drives	12.1%	11.7%	12.0% (42)
Left field	Flies caught	87.90%	98.50%	84.2% (38)
Left field	Liners caught	22.1%	25.5%	23.1% **
Left field	Stopping runners from advancing	64.2%	65.4%	63.9% (10)
Right field	Flies caught	88.2%	87.5%	88.5% (0)
Right field	Liner caught	23.5%	23.3%	36.7%*** (30)
Right field	Stopping runners from advancing	49.7%	51.8%	48.7% (45)

* - Strangely, new full time third basemen actually worsened over their first fifty games and as a group became below average after game 23.
** - As with third basemen, left fielders actually got worse over their first fifty games and as a group became below average after game 14.
*** - Likely a small-sample fluke, followed by a far smaller percentage; note the number of parentheses.

The reason that shortstops and center fielders are not listed is that converts were few and far between, and when they did occur tended to be those who had played there in the minors and were forced to play another position for a while because of an incumbent veteran. Russell's point was not that someone changing positions could not be as good or better than someone with experience there, but that it takes time for that to happen.

Carleton, Russell A. (2018). Bunting the the value of being honest. https://www.baseballprospectus.com/news/article/39497/baseball-therapy-bunting-value-honest/

This web spot of Russell Carleton (2018) was part of the re-assessment of the sacrifice bunt that began occurring at about that time. Between 2013 and 2017, the run expectancy for runner on first/no outs was 0.858 and runner on second/one out was 0.662, which makes bunting look bad. But if it is the ninth spot in lineup that bunts, the run expectancy for runner on second/one out for spots 1 through 3 in the order was 0.725, which is better. Further, the specific run expectancy for runner on first/no outs and the batter bunts was 0.773, better yet as it includes batters beating bunt out for hit, throwing errors etc. For the 8th and 9th spot in batting order, runner on first/no outs was 0.830 if the batter didn't bunt and 0.773 if the batter did. The point was not that the sacrifice is really a good strategy other than in one-run-really-matters situations, but that it was not as bad as earlier analysts had made it out to be.

Carleton, Russell A. (2018). The surprising evolution of the bullpen. https://www.baseballprospectus.com/news/article/40786/baseball-therapy-the-surprising-evolution-of-the-bullpen/

This Russell Carleton post included a couple of historical trends that when I did not remember seeing previously. Along with demonstrating one more time the increases in relievers per game and decrease in reliever innings per appearance between 1950 and 2017, the gap in strikeouts per plate appearance between relievers and starters has widened over the decades. The gap was rarely ore than 0.1 K/PA between 1950 and 1970, but starting in the mid-1980s up to at least 2017 relievers have had a fairly consistent 0.3 K/PA advantage over starters. In addition, the number of players per game appearing in multiple positions was around 0.2 in the early 1950s but jumped to between 0.50 and 0.60 by the late 1960s and stayed there through 2017.

Carleton, Russell A. (2018). Is it time to spit out the LOOGY? https://www.baseballprospectus.com/news/article/45661/baseball-therapy-is-it-time-to-spit-out-the-loogy/

Using 2014-2018 Retrosheet data, Russell Carleton (2018) determined that one out guys, whether left- or right-handed, give up fewer walks and get more strikeouts and outs on balls in play than those pitchers going longer, whether or not the standard platoon advantage was in play.

Carleton, Russell A. (2018). The surprising evolution of the starter. https://www.baseballprospectus.com/news/article/41513/baseball-therapy-the-surprising-evolution-of-the-starter/

The following table reveals the drastic downturn in the percentage of pitchers starting and, in particular, ending their third time through the batting order between 2012 and 2017, as compared with previous five-year intervals.

Year	Made it to 3rd Turn (batter 19)	Finished 3rd Turn (batter 27)	Median Last Batter Faced
2017	90.2%	24.6%	24th
2012	93.4%	40.2%	26th
2007	93.0%	42.5%	26th
2002	92.5%	46.1%	26th
1997	92.1%	50.7%	27th
1992	90.9%	54.0%	27th

1987	88.4%	53.2%	27th

Carleton, Russell (2018). The openers are coming, the starters are fine. https://www.baseballprospectus.com/news/article/43410/baseball-therapy-the-openers-are-coming-the-starters-are-fine/

I decided for this one to skip the detail and provide a quick summary; see the article in interested. Russell provides 1950-2017 examinations of the comparison of strikeout rate for starters versus relievers (originally about even, but relievers have been ever more ahead starting in 1970), the proportion of 70-inning relievers who started at least once (down from over 80 percent to under 20 percent), the proportion of relief appearances lasting more, less or exactly three outs (originally a majority was more, now a majority is exactly three), the average number of outs per start (from about 21½ to 17), the standard deviation of the number of outs per start (decreased from about 8 to 4, showing a lot less variation), the percentage of games in which starters got the most outs (increase from around 85% to around 96%, showing that “the bulk guy” has actually become more likely to be the starter), the percentage of games in which the starter faced 20 or fewer batters (around 21 percent in the mid 1950s, down to 12 percent or less in the 2000s up to 2014, when openers appeared and the number jumped to 18 percent in 2017).

Carleton, Russell (2018). Bullpen day in Tampa Bay. https://www.baseballprospectus.com/news/article/38633/baseball-therapy-bullpen-day-tampa-bay/

As part of an article describing the potential strategic value of pitchers entrusted with about three innings of work, Russell Carleton (2018) presented this copy-and-pasted table (probably from Retrosheet data) showing the percentage of appearances falling into various pitch count categories in 2017.

Pitch Count	Starters	Relievers
0-9	0.1%	23.7%
10-19	0.2%	46.3%
20-29	0.4%	20.0%
30-39	0.7%	6.6%
40-49	1.1%	2.0%
50-59	2.0%	0.8%
60-69	4.7%	0.4%
70-79	10.1%	0.1%
80-89	20.2%	0.1%
90-99	33.8%	0.0%
100+	26.8%	0.0%

Note how few relief appearances were for more than 40 and starts for less than 60 pitches. Russell's discussion is relevant to the advantages and disadvantages of filling that 40 to 60 hole on “bullpen days” with pitchers adept at going twice through an order and no more.

Carleton, Russell A. (2020). Are aces underrated?
https://www.baseballprospectus.com/news/article/62277/baseball-therapy-shane-bieber-ace-workload/
Carleton, Russell A. (2020). Aces high, sixes low.
https://www.baseballprospectus.com/news/article/62407/baseball-therapy-aces-high-sixes-low/

The point of this sequence of posts is to see the actual WAR value of starting pitchers given that their replacements not only are less successful but pitch fewer innings. He used 2015-2019 probably Retrosheet data, first looking at the win probability for starters getting a given number of outs, off of diagram. For 14 and fewer outs, win probability fluctuated between 20 and 40 percent, but starting with 15 it went up fairly linearly and dramatically, reaching 50 percent at about 17 outs, 60 percent at about 19 outs, 70 percent at about 21 outs, 80 percent at about 24 outs, and 90 percent at about 25 outs. Russell then divided starters getting 150 outs in each season into six bins depending on xFIP; 30 starters into each of five bins assumed to represent each team's five starters in order of outs and the remaining into a replacement starter bin. Here are relevant data:

Tier (by xFIP)	Average Outs Recorded by Starter	Percentage of Games with at least 15 outs recorded	Average win probability at starter exit	Average number of opponent runs given up at exit
Top 30	18.33	88%	59%	2.25
31-60	17.42	84%	52%	2.58
61-90	16.95	82%	50%	2.66
91-120	16.69	79%	48%	2.77
121-150	16.33	76%	46%	2.95
Replacements	13.83	56%	43%	2.79

In addition, the run expectancies for the rest of the inning when the starter departed were, respectively for the six bins, .21, .26, .30, .31, .32, and .34.
The sixth bin stood for replacement level in the WAR figures, so for example starters in the first bin had a 16 percent higher win probability, equivalent to 4.8 more wins when in there – WAR would represent that – but also 4.5 more outs.

Tier	Average Pitch County at Exit	Batters Faced at Exit
Top 30	96.57	24.74

31-60	93.61	24.28
61-90	92.04	23.94
91-120	91.92	23.85
121-150	90.85	23.58
Replacement	79.13	20.39

Note that #1 starters compared to replacements faced 4.35 more batters, and given that they got 4.5 more outs, basically they got all of those extra batters out. This in effect saved an inning and a third of pitching. An attempt to simulate that extra time through one full and one partial inning pitched by the replacement yielded an average of 1.68 runs, to add on to the 2.79. This equals 4.47 runs compared to the #1 starter's 2.46 (I added the run expectancy on to the runs given up, basically 2 more runs which is 60 more over 30 starts and so a WAR 0f 6 for the #1 starters). Russell believed that the true figure was actually more, but that at least provides a ballpark estimate.

Carleton, Russell A. (2020). Where have all the productive outs gone?
https://www.baseballprospectus.com/news/article/62811/baseball-therapy-where-have-all-the-productive-outs-gone/

This is a descriptive study of "productive outs," defined as outs in which a baserunner advances, between 1950 and 2019. I am interpreting findings from Russell's graphs. The proportion of circumstances in which productive outs occurred was about 28 percent in the early 1950s, went down to less than 25 percent in the late 1950s, up to 26 percent through about 1990 and 27 percent around 2000, but down to 25 percent in the 2010s. In these circumstances, successful baserunner advancement was around 32 percent in the 1950s, up to about 36 percent in the 1980s and 38 percent in the early 1990s, then back to 36 percent through the 2000s and early 2010s, then down to 33 percent in the late 2010s. These two figures appear to be correlated.
Non-pitcher PAs with a runner in first, none on second, and less than 2 outs – in other words classic sacrifice bunt situations – productive outs occurred at about 6 percent in the early 1950s, started falling in the late 1960s, and was by 2012 at about 2 percent and continued down to 1½ percent by 2019. This was pretty much responsible for the change in advance rate over the past few decades.
Stolen base attempt rate in the same situations at 4 percent in 1950, up to 12 percent by the early 1980s, and down to 6 percent by 2019. Success rate bounced between 50 and 55 percent in the 1950s, up to about 65 percent by the mid-1970s to the mid-2000s, and then up a couple of percentage points through 2019.

Carleton, Russell A. (2020). How could we ever replace you?
https://www.baseballprospectus.com/news/article/58746/baseball-therapy-how-could-we-ever-replace-you/

The following cut-and-pasted table lists the percentages of time (I would assume from Retrosheet data) that a team whose regular, on a day off, was replaced by someone who was a regular at another position, broken down to time periods.

Position	1950-2014 rate	2015-2017 rate	2018-2019 rate
C	3%	4%	3%
1B	35%	39.00%	44%
2B	21%	28%	40%
SS	19%	26%	37%
3B	26%	33%	41%
LF	30%	32%	48%
CF	35%	37%	49%
RF	32%	36%	60%

Russell's point was that at the time of this work, this strategy was becoming much more popular than it had been.

Carleton, Russell A. (2020). Remember some stats: Quality Start.
https://www.baseballprospectus.com/news/article/58127/remember-some-stats-making-a-quality-stat/

Russell Carleton (2020) provided what was sort of a postmortem for the Quality Start in 2020. From 1950 to about 2015, about half of all starts qualified, with a stretch from the mid-60s to mid 70s where this figure was more like 55 percent. From 1950 to 2019, a team receiving one from their starter won 67.6 percent of the time, which increased to 88.6 percent of the time when the opposing team's starter did not. The point for Russel is what happened to the percentage of starts that qualified after 2015 or so, which dropped like a rock annually down to around 37 percent in 2019. A major reason for this appears to be the drop in the proportion of games in which starters went the required six innings, which remained at or above 50 percent until that time and also plummeted to below 5⅓ by 2019. In short, the Quality Start as John Lowe defined it became obsolete beginning in the mid-2010s because teams no longer expected starters to last for six innings.

Carleton, Russell A. (2020). Why aren't you running.
https://www.baseballprospectus.com/news/article/57480/baseball-therapy-why-arent-you-running/

Carleton, Russell A. (2020). Why aren't you running part 2.
https://www.baseballprospectus.com/news/article/57527/baseball-therapy-why-arent-you-running-part-2-baserunning-aggression/

In these posts, Russell Carleton continued his work on, in his words “the rates at which players attempted and made it safely on several could-be advancements on the basepaths." This is from 2019 almost certainly Retrosheet data.

Situation	Attempt Rate	Success Rate
First to third on a single	32.1%	97.0%
First to home on a double	45.1%	93.1%
Second to home on a single	70.3%	96.2%
Tagging up from 2nd to 3rd on a fly ball	33.0%	94.3%
Tagging up from 3rd to home on a fly ball	78.9%	97.5%

Note how low attempts rates are on a couple of them. Next, using the 2019 run expectancy figures for each base-out situation, Russell computed the following break-even figures:

Situation	0 outs	1 out	2 outs
First to third on a single	81.4%	75.5%	90.0%
First to home on a double	87.8%	74.9%	45.9%
Second to home on a single	71.0%	73.9%	41.7%
Tagging up from 2nd to 3rd on a fly ball	71.4%	86.8%	***
Tagging up from 3rd to home on a fly ball	65.5%	35.0%	***

Comparing the two, it is clear that baserunners are not attempting these moves enough. Note the low breakeven for tagging up from third with one out; this is because the probability of scoring in this situation in other ways is very low. In a follow-up (2020i), Russell included the following table of averages for teams (I assume 2019 figures):

Situation	Number of Instances	Attempt Percentage	Number of Unsent Runners
Potential sacrifice fly	49.4	78.9%	10.4
Second to home on a single	134.0	70.3%	39.8
First to home on a double	81.6	45.1%	44.8
First to third on a single	170.2	32.1%	115.6

He used all of this to estimate a loss of 38 runs per season if teams would "send everybody."

Carleton, Russell A. (2020). Bullpen everything?
https://www.baseballprospectus.com/news/article/60922/baseball-therapy-bullpen-game-relievers-starters-mlb/

The point of this webpost was to determine the soonest that pitchers were reused based on the number of batters they faced in their previous outing, irrespective of their effectiveness on that outing. The following data is for the lowest 10 percent days off for each number of batters faced. For example, after facing as many as 7 batters, pitchers were used the very next day as often as 10 percent of the time. Here are the data:

Days	1	2	3	4	5 or more
Batters	1-7	8-11	12-14	15-16	17 and more

Carleton, Russell (2021). Baseball's other loose screw.
https://www.baseballprospectus.com/news/article/63809/baseball-therapy-baseballs-other-loose-screw/

The following data on the length of games and run scoring were contributed by Russell Carleton, almost certainly from Retrosheet data.
Games in 2020 were on average 19 minutes longer than in 1990 and 33 minutes longer than in 1970.
For games with relevant relevant data available, plate appearances divided by game length (not quite average PA time as it also includes time between innings, pitching changes, etc.) averaged about 1½ minutes in the 1920s and 1.6 minutes around 1940, but jumped to around 2 in the 1960s, 2¼ in the 1990s, and close to 2½ by 2020. The times between balls in play was at about 2 minutes in the 1920s, rose over 2½ around 1950 and 3 soon after 1980, and was over 3½ and still climbing in the 2010s.
The proportion of games in which the two teams are within two runs of one another at the beginning of the seventh inning, a good operational definition for a close game, is a function of the general run scoring environment; more runs, fewer close games. Therefore it is not surprising to see that that proportion, that kicked around 56 percent through the 1950s, was up over 60 percent in the run-poor late 1960s, back to about 56 percent in the 1980s, down during the 1990 steroid period to around 52 percent, back again to about 56 percent by 2010, but down to 54 or so by 2020.

Carleton, Russell (2021). The risk of the short(stop).
https://www.baseballprospectus.com/news/article/64393/pecota-2021-the-risk-of-the-shortstop/

From 200 through 2019, the proportion of shortstops who played a minimum of 729 defensive innings in two consecutive seasons was well over .8 until age 29 but dropped afterward to as low as about .25 at age 34, as they moved down to Defensive Spectrum.

Carleton, Russell A. (2021). Deconstructing the double hook.
https://www.baseballprospectus.com/news/article/66310/baseball-therapy-deconstructing-the-double-hook/

The average number of outs earned by starting pitchers has not dropped that much since 1950. Between 1950 and 1980 it annually jockeyed between 19 and 20; between 1995 and 2019 it was normally between 18 and 18½. Interestingly, Russell also showed a diagram for the average number of outs for the pitcher getting the most in games, which was generally well over 20 until the late 1970s as long relievers sometimes pitched for the majority of games. The latter figure decreased to approximate those for starters per game by 1990.

Carleton, Russell A. (2021). The invasive reliever.
https://www.baseballprospectus.com/news/article/64030/baseball-therapy-invasive-reliever/

The proportion of reliever appearances that lasted exactly three outs, hovering around 20 percent between 1950 and 1980, exploded upward to over 50 percent about 2010 and stayed there through 2020. Compensating for this, those longer than three outs were around 55 percent until 1980, collapsed to about 20 percent around 2010, and then rose a tad to 25 percent in 2020. Those less than three outs were a bit over 20 percent until 1990, rose to about 30 between 1995 and 2016, and then dropped back to about 25 percent in 2020. It will be interesting to see if the recent changes in the latter two are blips or long term trends.

Carleton, Russell (2021). After the invasion.
https://www.baseballprospectus.com/news/article/64205/baseball-therapy-after-the-invasion-short-relievers/

The following data were supplied by Russell Carleton, almost certainly from Retrosheet. The number of relievers used when starters lasted exactly 15 outs fluctuated between 2 and 2½ between 1950 and about 1990 but grew to between 3 ½ and 4 around 2010 and stayed there through 2020. The trends were approximately parallel for 18 outs from the starter (a gradual increase from 1½ to 2 followed by a jump to 3) and 21 outs (1¼ up to a little over 2).
Teams have been able to find more relievers capable of high strikeout rates. Looking only at one-inning appearances, the strikeout rate for those pitching only the sixth, seventh, eighth, and ninth innings were pretty close to one another at around 15 percent until about 1980. Afterwards, it began increasing for all of them, but more quickly for the ninth inning and to a lesser extent the eighth inning until about 2000. Since then, all of them have continued to increase at about the same rate, approximating 30 percent for ninth inning relief appearances, closing in on that figure for eighth inning, and over 25 percent for sixth and seventh. On base averages have followed an analogous if opposite pattern.

The average number of pinch-hitters per game, which jumped from about 1.1 in 1950 to 1.4 in the 1960s as the number of relievers per game rose, returned to about 1.1 in 1973 when the designated hitter came into the American League. It remained there until 2020, when it collapsed for the year to about ½ per game due at least partly due to the National League using the DH that weird season. But looked at more closely, in games with DHs the number, about .8 per game during the 1980s, dropped to about .5 per game by 2000 and stayed there for the next two decades. While overall in the National League pinch-hitting increased during the entire 1950 to 2020 era from about 1.1 per game to 1.8 per game, this rise was fueled by increased PHing for pitchers. In the NL, PHers for non-pitchers dipped analogously to AL usage during the 1980-2020 interim. The proportion of plate appearances in which the team at bat had the standard platoon advantage was below half through much of the 1950s but rose to over 60 percent in the 1980s and early 1990s. As the number of pitchers on rosters rose, leaving less places on the bench for substitutes, it fell to about 55 percent in the rest of the 1990s and stayed there afterward. The proportions were even steeper from the seventh inning on, as teams increased strategic pinch-hitter use, it was up to about 57 percent in 1990; as teams increased strategic reliever use, down to about 40 percent by 2000.

Carleton, Russell A. (2021). How to kill an octopus (of relievers). https://www.baseballprospectus.com/news/article/67949/baseball-therapy-how-to-kill-an-octopus/

For relievers entering the game beginning in the seventh inning and , strikeout percentage rose steadily between 1950 and 2020 except for the dip caused by rule changes after the “year of the pitcher” (1968). Russell's point here is that, although the rate for relievers pitching a second inning paralleled that for relievers pitching the first of a multi-inning appearance, that rate has always been about three percent lower. This is an argument in favor of one inning appearances under the current (2021) rules.

Carleton, Russell A. (2021). The importance of sharing. https://www.baseballprospectus.com/news/article/70770/baseball-therapy-importance-of-sharing-reliever-workload/

When this webpost was offered, starter usage was nosediving, down to 56.7 percent of PAs in 2021; it was in the mid or upper 60s mid 2000s until mid 2010s. During that earlier period, starters averaged around 18 outs, and was down to 15.07 in 2021. The average number of pitchers per game was around four from mid 2000s to mid 2010s, but by 2021 was closing in on 4½.

Based on 2015-2019 data for games with fewer than 30 pitches and with no period of more than 10 days without pitching, Russell performed a multiple regression for factors impacting reliever workload. Variables that did NOT matter include number of rest days (which doesn't mean it isn't important, but managers don't let it get out of hand), pitches during the last outing, amd back-to-back usage. An important predictor was total seasonal pitch count up until the relevant appearance, and it affected of all things batting average on balls in play (not K or BB). Sixty extra pitches equalled 1 additional OBA point. In other words, relievers wear down over the season.

Carleton, Russell A. (2022). A solution to the strikeout problem? https://www.baseballprospectus.com/news/article/75619/a-solution-to-the-strikeout-problem/

Given the decision to swing, the following is relevant to the decision concerning the type of swing, specifically whether to swing for the fences or just try to make contact. The following is from Russell's charts for 1993-2011: Contact rate when swinging on first pitches was around 80 percent from 1993 to about 2010 but down to around 74 percent from 2019 to 2021. Contact rate when swinging with two strikes was almost the same, but started its downward trend a bit earlier, about 2005. Slugging average on contact on first pitch was around .500 mid-1990s to mid-2010s and then over .550 after that through 2021. Slugging on contact with two strikes was about the same early on but drifted down to .450-.475 mid-2000s to-mid 2010s and up to .500 or higher after that through 2021. Putting these together, it seems that batter decisions on the type of swing have not been much different with two strikes than on the first pitch all the way back to 1993 if not earlier.

These data imply that when batters did make contact, they seem to have gotten more bang for the buck starting in the mid-2010s. But here is counterevidence – with two strikes, overall RE24 in linear weight terms varied between about –.06 and –.07 again 1993 to 2015 or so, but decreased below –.07 afterward. This implies that the increase in absence of contact aka strikeouts since then has outweighed the slugging on contact increase, making the swing-for-the-fences strategy an overall loser.

Russell also did a scatterplot of annual HR per fly ball rate pitted against K rate 2002 to late June 2022. Although not the case for the first 10 years – in fact the relationship then might be negative – there seems to be a positive relationship for the 10+ years since. This illustrates the swing-for-homers vs. strikeout tradeoff nicely.

Other data Russell presented here: In 2021, the average number of pitches per PA were 4.71 on strikeouts, 5.73 on walks, and 3.32 on “everything else,” which would mostly be batted balls. Finally, irregardless of whether the batter beats it out, the proportion of bunts with runners on that successfully advanced the runner were between 85 and 95 percent from 1950 to about 2000, then drifted down about linearly to about 80 percent by 2021.

Carleton, Russell (2022). What if contact is the #NewMoneyball? https://www.baseballprospectus.com/news/article/76630/baseball-therapy-what-if-contact-is-the-newmoneyball/

More on the same issue as above. Contact rate decrease really started in 2014, and since then .01 drop in RE24 per PA with two strikes, showing that power vs strikeout tradeoff with two strikes is a loser for batters. With more than 3000 PAs that reach two strikes per team per season, that adds up to 30 runs scored lost.

Carleton, Russell A. (2022). So you've decided to give up. https://www.baseballprospectus.com/news/article/75885/baseball-therapy-position-players-pitching/

The issue is when should a team give up and use a non-pitcher on the mound for mopping up. An MLB rule now says that with the exception of players that are designated as "two-way," which requires previous experience (see https://dodgerblue.com/mlb-changes-rule-for-position-players-pitching-any-scenario-during-2020-season/2020/06/25/ for details, non-pitchers can toe the mound either in extra innings or with at least a six run lead or deficit.
at both this can only occur deficit must be at least 6 runs for this to occur. Russell studied this issue using probably Retrosheet data from 2008 through July 11th, 2022. During that period, non-pitchers have given up a slash line of .336/.408/.654. Assuming that the other team needs at least two batters to get on base before three outs, the odds of winning while giving up a .408 OBA are 46.1 percent with a one run lead, and 66.8 percent with a two run lead. He didn't say anything about a three run lead, but after that:

Size of lead	Success rate
4	89.6%
5	94.5%
6	97.1%
7	98.6%
8	99.3%
9	99.7%
10	99.8%

Looks worth using a non-pitcher is worth the risk with a four run lead given that if the non-pitcher gets into trouble you can get a real pitcher in there. After that, it is definitely worth it.
The following shows "the percentage of games where a team enters that half inning as the pitching team and is down by 7 or more runs."

Half Inning	Down 7+ runs
Bot 7	4.3%
Top 8	4.4%
Bot 8	5.5%
Top 9	5.3%

It actually happens quite often. Russell didn't show analogous data for 6 or more runs down, which would be more informative given the rule.
Turning to a different topic, this is for the average FIP for pitchers on the mound in the top of the 8th. Russell said that other innings have the same trajectory.

Down 6-8 runs, 4.1
Down 5 runs, 4.05
Down 4 runs, 3.95
Down 3 runs, 3.9
Down 2 runs, 3.75
Down 1 run, 3.65
tied, 3.5
Up 1 run, 3.45
Up 2-3 runs, 3.5
Up 4 runs, 3.6
Up 5 runs, 3.7
Up 6 runs, 3.8
Up 7 runs, 3.85
Up 8 runs, 3.9

Note that the current rule allowing a save to be credited with a three-run lead results in pitchers with a FIP equally low three runs ahead as when tied.

Here are win probabilities for team in the field from 2003-2021 almost certainly Retrosheet data. The point of this work is to determine the run deficit at which the team in the field should "give up" and use a non-pitcher on the mound.

Run Diff	Bot 7	Top 8	Bot 8	Top 9
Tied	37.6%	53.2%	36.2%	52.1%
-1	17.9%	23.4%	10.5%	14.6%
-2	8.4%	12.7%	4.4%	6.4%
-3	4.0%	5.5%	1.4%	2.4%
-4	1.9%	3.0%	0.8%	1.0%
-5	0.8%	1.6%	0.2%	0.7%
-6	0.3%	1.2%	0.1%	0.5%
-7	0.1%	0.1%	0.0%	0.0%

Carleton, Russell A. (2022). The New England Journal of Baseball. https://www.baseballprospectus.com/news/article/74629/baseball-therapy-the-new-england-journal-of-baseball/

This is in response (40 years late) to an article published by McLean and Ciurczak in the *New England Journal of Medicine* (1982, Vol. 307 No. 20, pages 1278-1279) in which these authors claim that a 20 point advantage in BA for batters who both bat and throw lefthanded over those who bat left and throw right and those who are pure righties is due to lefties having less lateralized i.e. less specialized brains, resulting in less overt handedness and perhaps greater overall dexterity. Stephen Jay Gould in an essay originally published in 1983 and reprinted in his 2003 book *Triumph and Tragedy in Mudville* disputed this claim, preferring the more prosaic but probably more accurate explanation, consistent with a later report (Grondin et al., 1999, *Journal of Experimental Psychology: Human Perception and Performance,* Vol. 25 No. 3, pages 747-754), that mixed lefties were in general born righthanded and learned to bat lefty due to its competitive advantages in baseball, but as it is a bit unnatural are not as successful overall.
Gould's account gains support from the fact that the proportion of righties in general society is far greater than that in baseball. Russell noted that in 2021, 40.3 percent of PAs were lefty hitters, about 4 times the overall proportion of lefties in the general population. And 66.8 percent of lefty hitters that year were righty throwers. (Incidentally 20.2 percent of lefty pitchers who batted that year batted righty). To compare their performance with pure lefties and pure righties, one has to control for position, because lefty throwers can't be infielders (other than first) or catchers and are more likely to become pitchers. Anyway, controlling for position, in 2017-2021 data (probably Retrosheet), the BL/TR had slightly higher Three True Outcome figures than lefty hitters and righty throwers overall, In addition, using 1976-1980 data as did McLean and Ciurczak, they had slightly higher BAs with the position control in place.

Carleton, Russell A. (2022). Ghost of the bunt. https://www.baseballprospectus.com/news/article/73675/baseball-therapy-the-ghost-of-the-bunt/

Between 1993 and 2019, home-field advantage in extra inning games was 52.1 percent, just a bit less than the overall 53.7 percent during that period of time. However, in 2020 and 2021, with the runner-on-second rule, it dropped to 47.2 percent against the overall 54.2 percent. Russell Carleton (2022) argued that the presence of the "ghost runner" on second was responsible for the decline. Here are some relevant data he displayed from 2017 to 2021 for run expectancies with a runner on second and no outs (the relevant circumstance for ghost runners) across all situations:

Runs scored	… if bunt	… if swing away
0	31.8%	39.3%

1	41.7%	31.8%
2+	26.5%	28.9%
Average	1.16	1.15

As Russell noted, the overall average is basically the same, but the distribution differs substantially, showing the huge advantage of bunting as a one-run strategy. Even with batters who bunted two or fewer times a year, whom we will assume are much better at hitting than bunting, one-run percentages were 43.7 and 31.8, respectively. These numbers imply that if either the visitor does not score or scores two or more in the top of the inning, the home team ought to swing away, but if the visitor scores one, the home team is better off bunting (the total odds of scoring any runs are 68.2%) than swinging away (60.7%). Russell noted that the home team bunted only 20 percent of the time in this situation, which was probably the wrong choice most of the time given these figures.

Carleton, Russell A. (2022). Who will catch that deep drive into left field? https://www.baseballprospectus.com/news/article/73522/baseball-therapy-who-will-catch-that-deep-drive-into-left-field/

Does better hitting result in better fielding? Russell Carleton (2022; 2003-2021 Retrosheet data) took individual player's basic outcome for a PA (on base or out both individually and over 10 PAs to see whether that outcome affected whether the player's odds of making a play on a ball in their territory was higher or lower than the player's overall odds. Russell found nothing either way.

Carleton, Russell A. (2022). What my thirteen-year-old taught me about bunting. https://www.baseballprospectus.com/news/article/76475/baseball-therapy-what-my-thirteen-year-old-taught-me-about-bunting/

This webpost featured the history of sacrifice hits, 1894-2021. Per team per game, they were at about 0.7 in the 1890s, then well over 1 through about 1925 (as high as 1.3 in the 1900s), then back down to 0.6 to 0.8 until the late 1940s, then a fairly steady drift down to 0.3 around 2010, followed by a steeper drop to 0,1 by the end of the 2010s. The following is 2017-2021 run expectancies.

Situation	Run expectancy	Score at least one
Runner at 1st, no out	0.92	42.3%
Runner at 2nd, one out	0.71	40.5%

In general, still a bad overall strategy.

Situation	Run expectancy	Score at least one

Runner at 1st, no out, 9th hitter up (non-pitcher)	0.96	43.0%
Runner at 2nd, one out, leadoff hitter up	0.78	42.5%

This is relevant to the universal DH game. A bad overall strategy even with #9 hitters.

Situation	Run expectancy	Score at least one
Runner at 1st, no out, 8th hitter up (non-pitcher)	0.85	39.2%
Runner at 2nd, one out, 9th hitter up (non-pitcher)	0.75	41.6%

An example in which the sac but works as a one-run strategy.

Situation	Run expectancy	Score at least one
Runner at 1st and 2nd, no out	1.51	62.1%
Runner at 2nd and 3rd, 1 out	1.42	67.6%

Two bases is worth an out if you need one run.

Number of Runs	Runner at 2nd, no out	Runner at 3rd, 1 out
0	39.1%	34.3%
1	32.3%	46.8%
2	14.7%	10.9%
3	7.5%	4.5%
4	3.5%	2.1%
5	1.6%	0.8%
6	0.7%	0.4%
7	0.3%	0.2%
Total Average	1.15	0.99

Another good one run strategy, but not otherwise. The webpost also includes detailed win expectancies for runner on first/no out and runners on first and second/no out; see the webpost.

Carleton, Russell A. (2022). Were the two leagues ever really that separate?
https://www.baseballprospectus.com/news/article/77086/baseball-therapy-were-the-leagues-ever-really-that-separate/

This webpost first provides a history of the amount of pinchhitting from 1950 to 2022. Pinchhitting for other than pitchers for both leagues was about 500 per season in 1950s, cruising up in the N.L. to about 1000 around 1990 and then down to maybe 800 by 2021. In the A.L, when the DH started, it exploded to well above 2000 during many years in the 1980s and early 1990s, almost certainly because without pitchers hitting there was opportunity to use them for weak-hitting non-pitchers. It fell afterward but stayed well above 1000, and was perhaps about 1200 in the 2010s.
In the N.L., pinchhitting for the pitcher occurred maybe 800 times in 1950, and then up about linearly to about 2000 around 1990 and then faster to 3000 by the mid 2000s and may still have been rising when the DH was finally approved.
Second, base stealing attempts: From 1950 to 2021. Attempts per team per (I assume) plate appearances went up in tandem for both leagues from about .05 to about .10 when the DH stated early 1970s, and then continued up to about .125 in early 1980s in the NL whereas the AL remained about constant. The NL came back to about .10 in early 1990s, then the leagues went down in tandem to about 0.6 in 2021. Second, Although highly variable from year to year, from 1901 to 1920 a good 80 to 90 percent of players, entire careers were in one league. After that, this proportion, although remaining variable, dropped approximately linearly to about 40 percent in the early 2020s.
On a second topic: Although highly variable from year to year, from 1901 to 1920 a good 80 to 90 percent of players, entire careers were in one league. After that, this proportion, although remaining variable, dropped approximately linearly to about 40 percent in the early 2020s.

Carleton, Russell A. (2022). Remember when pitchers used to bat?
https://www.baseballprospectus.com/news/article/74470/baseball-therapy-pitchers-hitting-dh-small-ball/

This webpost covers situations during 1993-2021 (not 2020) with no outs and a runner on first, the classic sacrifice bunt situation. Pitchers attempted bunts about 90 percent of the time and were successful about 70 percent of the time, with a lot of year-to-year variability. They also struck out about 25 percent of the time, and probably the other 5 percent were mostly groundouts. For non-pitchers, sacrificing decreased from about 12½ percent (15 percent when playing with A.L. rules) to (guessing from a graph) 3¾ percent, and was successful maybe 75 percent of the time when attempted.
Other included data: From 1993-2021 (not 2020), with zero or one out/runner on first and second clear, attempted steals went down from about 10 percent of the time to about 6 percent. With runners on, fewer than two outs, and the batter making on a fair territory batted ball, runners advanced around 33 percent at the start of period but wwnt down to about 30 percent at the end of it. For 2003-2018 and 2021 with runners on, counts other than 3-2, situations in which runners are often put in motion and the ball put into play (the definition of hit and run) were very rare, between 1.6 and 2.1 percent of the time 2003 to 2014, then dipping to around 1 percent in 2021.

All of this occurred parallel with and was likely affected by rises in strikeouts (from about 15 to 22 percent during interim) and HRs (from about 2½ to 3½ percent).

Carleton, Russell A. (2022). Ghost of the bunt.
https://www.baseballprospectus.com/news/article/73675/baseball-therapy-the-ghost-of-the-bunt/

As mentioned in the Situation chapter, home teams won only 47.2 percent of extra inning games in 2020 and 2021, much lower than both the overall 54.2 overall percentage in that period and the 52.1 percent of extra inning games between 1993 and 2019. Russell Carleton (2022) argued that the presence of the "ghost runner" on second was responsible for the decline. Here are some relevant data he displayed from 2017 to 2021 for run expectancies with a runner on second and no outs (the relevant circumstance for ghost runners) across all situations:

Runs scored	**… if bunt**	**… if swing away**
0	31.8%	39.3%
1	41.7%	31.8%
2+	26.5%	28.9%
Average	1.16	1.15

As Russell noted, the overall average is basically the same, but the distribution differs substantially, showing the huge advantage of bunting as a one-run strategy. Even with batters who bunted two or fewer times a year, whom we will assume are much better at hitting than bunting, one-run percentages were 43.7 and 31.8, respectively. These numbers imply that if either the visitor does not score or scores two or more in the top of the inning, the home team ought to swing away, but if the visitor scores one, the home team is better off bunting (the total odds of scoring any runs are 68.2%) than swinging away (60.7%). Russell noted that the home team bunted only 20 percent of the time in this situation, which was probably the wrong choice most of the time given these figures.

Carleton, Russell A. (2022). Is DHing bad for polar bears?
https://www.baseballprospectus.com/news/article/73232/is-dhing-bad-for-polar-bears/

Using 2003-2021 (presumably Retrosheet) data, Russell Carleton (2022) estimated an overall DH penalty of six OBA points overall, with a lower BABIP and no fewer homers despite more fly balls, implying less sharply batted balls. But, when 75 percent or more of PAs were as DH as opposed to on the field, there was no penalty.

Carleton, Russell A. (2022). Were the two leagues ever really that separate? https://www.baseballprospectus.com/news/article/77086/baseball-therapy-were-the-leagues-ever-really-that-separate/

Russell Carleton (2022) gave us the following historic trends: Pinch-hitting for batters other than pitchers across both leagues was at about 500 per season in 1950s, in the N.L. cruising up to about 1000 around 1990 and then down to maybe 800 by 2021. In the A.L., when the designated hitter began and PHs were not needed for pitchers any more, it exploded to well above 2000 in many years in the 1980s and early 1990s, then down but well above 1000, perhaps at about 1200 in 2010s. In the NL, pinch-hitting for the pitcher was at maybe 800 in 1950, and then up about linearly to about 2000 around 1990, faster to 3000 by mid-2000s as starters pitched fewer innings, and may still have been rising until the DH hit the league in 2022.
Although highly variable from year to year, from 1901 to 1920 a good 80 to 90 percent of players, entire careers were in one league. After that, this proportion, although remaining variable, dropped approximately linearly to about 40 percent in the early 2020s.

Carleton, Russell A. (2022). Looking for quality in the modern era. https://www.baseballprospectus.com/news/article/76956/baseball-therapy-looking-for-quality-in-the-modern-era/

As has been previously noted, about half of starts qualified as Quality from 1950 to the mid 2010s, but nosedived to 30 percent by 2020 as the mean number of innings per start dropped to about five. Between 1993 and 2012, home teams getting through the top of the sixth giving up three or fewer runs won 68.7 percent of the time; keep in mind that this includes games in which the starter did not last six innings, which is why this figure is so much higher than the percentage of Quality Starts. The corresponding number for the road team through the bottom of the sixth was 63.0 percent, the difference being home field advantage. So, as Russell did, if we round these figures to two-thirds, then quality starts could be defined as starts in which the starter's team wins about two-thirds of the time. Here is a table of relevant winning averages for those years:

Outs Gotten	**Runs Allowed**	**Visitor Win%**	**Home Win%**
9	0	62.6%	68.6%
10	0	64.4%	70.1%
11	0	66.9%	72.4%
12	1	63.8%	69.5%
13	1	64.7%	70.4%
14	1	66.5%	71.9%
15	2	63.2%	69.0%

16	2	64.5%	70.2%
17	2	66.5%	72.0%
18	3	63.0%	68.6%
19	3	63.6%	69.2%
20	3	65.6%	70.8%
21	4	62.0%	67.8%

Being a mite conservative, one could say that getting 11 outs with no runs allowed, 14 outs with one run allowed, 17 outs with two runs allowed, and 20 outs with three runs allowed could then qualify as Quality Starts. With this definition, about 60 percent of starts were Quality between 1950 and the mid-2010s, with the exception of the pitching-dominant late 1960s when it jumped to about 70 percent, and with the late 2010s dip to a more reasonable 50 percent. Another approach would be to assign responsibility to the pitcher getting the most outs, as such replacing openers with bulk pitchers, the corresponding figures were 65 to 70, 75, and 55 to 60 percent.

Carleton, Russell A. (2022). What my thirteen-year-old taught me about bunting. https://www.baseballprospectus.com/news/article/76475/baseball-therapy-what-my-thirteen-year-old-taught-me-about-bunting/

Russell Carleton (2022) computed the following run expectancies and odds of scoring for sacrifice bunting for runner on first/no out for 2017 through 2022. It remained a poor overall strategy. The first two columns overall, and the following sets of two columns assuming a DH world with, in turn, a non-pitcher #9 hitter followed by the #1 hitter and a #8 hitter followed by a #9 hitter, both non-pitchers:

	Overall		#9 non-pitcher to leadoff		#8 to #9, non-pitchers	
Situation	Run expectancy	Score at least one	Run expectancy	Score at least one	Run expectancy	Score at least one
Runner at 1st, no out	0.92	42.3%	0.96	43.0%	0.85	39.2%
Runner at 2nd, one out	0.71	40.50%	0.78	42.5%	0.75	41.6%

Here is one that works as a one-run strategy:

Situation	Run expectancy	Score at least one
Runner at 1st and 2nd, no out	1.51	62.1%
Runner at 2nd and 3rd, 1 out	1.42	67.6%

And runner on second, no out:

Number of Runs	Runner at 2nd, no out	Runner at 3rd, 1 out
0	39.1%	34.3%
1	32.3%	46.8%
2	14.7%	10.9%
3	7.5%	4.5%
4	3.5%	2.1%
5	1.6%	0.8%
6	0.7%	0.4%
7	0.3%	0.2%
Total Average	1.15	0.99

Everything in this chart refers to a situation with a runner on second and no outs.

Scenario	Expected Win% if swing away	Expected Win% if bunt
T7, visitors down 1	40.6%*	37.8%
T7, tied	59.9%*	58.9%
T7, visitors up 1	76.5%*	75.9%
B7, home down 1	50.4%*	48.7%
B7, tied	71.8%*	71.6%
B7, home up 1	86.4%	86.4%
T8, visitors down 1	38.9%*	35.7%
T8, tied	62.7%*	62.4%
T8, visitors up 1	82.1%*	81.8%
B8, home down 1	48.5%*	46.5%
B8, tied	75.6%	76.5%*
B8, home up 1	92.3%	92.5%*
T9, visitors down 1	36.4%*	32.6%
T9, tied	67.1%	68.1%*
T9, visitors up 1	89.8%*	89.7%
B9, home down 1	45.8%*	44.1%

Same basic idea, but first and second, no outs:

Scenario	Expected Win% if swing away	Expected Win% if bunt
T7, visitors down 1	46.4%*	45.9%
T7, tied	63.5%	64.0%*

T7, visitors up 1	78.7%	79.0%*
B7, home down 1	55.3%	55.9%*
B7, tied	74.2%	75.2%*
B7, home up 1	87.4%	88.0%*
T8, visitors down 1	45.6%	45.6%
T8, tied	66.0%	67.2%*
T8, visitors up 1	83.8%	84.2%*
B8, home down 1	54.0%	55.5%*
B8, tied	77.2%	79.0%*
B8, home up 1	92.8%	93.3%*
T9, visitors down 1	43.9%	44.9%*
T9, tied	69.5%	71.7%*
T9, visitors up 1	90.7%	91.1%*
B9, home down 1	51.7%	54.8%*

Carleton, Russell A. (2022). Are the playoffs really different? https://www.baseballprospectus.com/news/article/78078/baseball-therapy-are-the-playoffs-really-different/

Data on how post-seasons have differed from the regular seasons, mostly from 2017 through 2022. First, more pitchers were used per game post-season than in season.

	Regular Season	Playoffs
Number of pitchers used per game	4.36	5.09
Number of outs recorded by starter	15.68	13.70
Percentage of relief appearances that start mid-inning	31.5%	42.3%

Between 1950-2021, reliever usage patterns revealed much more variation across years due to far fewer games, but the same overall decrease in outs per appearance. Summarizing over a graph; Relievers were used for three outs around 20 percent of the time in 1950, which went up to around 40 percent in 2020. Appearances lasting fewer than three outs about also about 20 percent of the total in 1950, then maybe 35 percent in 2020. Consequently, relievers recording more than three outs comprised the remaining 60 percent in 1950 and 25 percent in 2020. The average length per start also differed; at 20 or so batters in 1950, the figure decreased to about 14 in the late 2010s through 2020 and only to 12 in 2021.

Again limited to 2017 to 2021, fielding metrics looked better in the post season.

	Regular Season	Playoffs
BABIP	.298	.282
Out percentage on GB	73.7%	74.5%
Out percentage on GB to infielder	88.2%	88.6%
Out percentage on OF flies	88.6%	89.90%

Games were far longer (218.3 minutes vs. 188.6 minutes in regular season).
Strikeouts were up (25.3% vs 22.6% in the regular season)
Walks were pretty close (8.5% vs 8.2% in the regular season).
Teams behind 1 or 2 runs in the 7th inning came back to win more often (25.9% vs. 24.2% in the regular season).
In contrast, when 1 or 2 runs behind in the 9th inning, teams came back to win less often (7.1% vs. 11.1% in the regular season), a testiment to relief pitching.
Sinoyara home runs? About the same (3.3% vs 3.5% in the regular season)
Games ended with strikeouts more often (relief pitching again; 34.6% vs 32.8% in the regular season)

Greater use of bench players:

	Regular Season	Playoffs
Average number of pinch runners	0.15	0.21
Average number of pinch hitters	1.83	2.00
Total players entering the game	14.7	15.9

Differences for "small ball": Mixed for usage:

	Regular Season	Playoffs
Bunts by non-pitcher with runner on 1st/no out	2.8%	2.6%
Taking an "extra" base on a hit	42.2%	43.8%
Stolen base attempt rate (2nd base)	6.2%	5.1%
Hit and run rate	0.8%	0.9%
"Productive" out rate	33.6%	32.3%

But except for caught stealing, small ball was more helpful:

	Regular Season Win%	Playoff Win%
More sac bunts	65.9%	68.2%
More SB	64.4%	66.0%

More caught stealing (as runners!)	56.5%	57.9%
More Extra bases taken on hits	75.3%	80.0%
More "productive" outs	64.9%	65.8%
More pitches seen per PA	45.1%	41.5%

"Big ball" was bigger:

	Regular Season Win%	Playoff Win%
More HR	77.5%	83.8%
More Extra Base Hits	80.1%	88.8%
More (hits + walks + HBP)	81.3%	84.7%
More strikeouts	38.1%	38.4%

Carleton, Russell A. (2023). Is bunting a lost art?
https://www.baseballprospectus.com/news/article/79811/baseball-therapy-is-bunting-a-lost-art/

Russell Carleton (2023) offered the following Retrosheet-based data concerning bunting, all read off of graphs: The proportion of PAs with a runner on first, no outs, and a non-pitcher batting during which the batter bunted dropped from between 10 and 12 percent from 1950 to 1990 to 2 or 3 percent 2020-2022. The proportion of players with at least 250 PA who were asked to bunt at least once went from 90 percent in 1993 to below 6 percent in 2022. These drops came primarily from the 25 players who bunted the most often, from about 60 a year to about 40, with analogous but less extreme decreases from the next three groupings of 25 players. Success rates in terms of getting a bunt down in fair territory, between 50 and 55 percent 1993 to 2000 for both pitchers and non-pitchers, dropped to around 45 percent for non-pitcher and a bit below that for pitchers by 2015-2022. The non-pitcher figure was not affected by how often the player was asked to do so. One reason for these dropoffs has been increased pitch velocity; there has been a direct negative relationship between velocity and success rate; 56 to 58 percent for 84 to 86 mph down to 44 percent for 96 and 97 mph (Statcase data, 2015-2022). Given a bunt, successful runner advancement, about 90 percent for non-pitchers and 80 percent for pitchers, dropped about 5 percent for both in 1984, the year in which Retrosheet began using Project Scoresheet data; this measurement change was probably responsible for the drop. Along the same line, the success rate for non-pitcher bunts with nobody on, clearly attempts at hits, was at about 60 percent 1950-1983 and 45 percent 1984-2022.

Carleton, Russell A. (2023). Scoop there it is(n't).
https://www.baseballprospectus.com/news/article/79960/baseball-therapy-scoop-there-it-isnt/

We have no good method for measuring the ability of first basemen to make plays on errand throws from other infielders. Russell Carleton (2022b) made an admittedly very approximate attempt at formulating one. He worked with 2018-2022 (likely Retrosheet) data for infield grounders (no bunts) for which the other infielder was not charged with an error and there was no reason to believe that a throw went to another base, controlling for specific other infield position to which the ball was hit and (not very precisely) baserunner speed. A regression analysis showed that, if getting past baserunning speed as a predictor, the responsibility for whether the batter was safe at first broke down to 92 to 8 in favor of the other infielder; and Russell replicated that ratio with 1993 to 1999 aka pre-infield-shifting data. A second regression, again controlling for grounder location plus fielding and batter-as-baserunner success rates, Russell concocted a plus-minus metric for the batter being safe versus out as a stand-in for first baseman scooping-up-throws performance. The metric was not trustworthy, with year-to-year correlations for specific first basemen bouncing between + and –0.2 between 2003 and 2022. It was, however, the case that emergency first basemen (i.e., those stuck there with no previous experience at the position) were demonstrably worse for a period of time.

Carleton, Russell A. (2023). The shift was framed.
https://www.baseballprospectus.com/news/article/80353/baseball-therapy-the-shift-was-framed-strikeouts-the-baseball/

Russell Carleton (2023) defined a pulled batted ball by a righty(lefty) batter as either one in play fielded by the third baseman or left fielder (first basemen or right fielder), with shortstop (second base) getting half credit so to speak, or as a homer hit to left (right) field no including left-center (right-center). Based on Retrosheet 1950-2022 data, pull rate, at about 42 percent in the 1950s, went down steadily to about 39 percent in the early 1980s and, from then through 2022, jumped around between 37 and 39 percent.

Carleton, Russell A. (2023). The swarm of relievers.
https://www.baseballprospectus.com/news/article/82606/baseball-therapy-the-swarm-of-relievers/

The main point of this webpost is to demonstrate the greater responsibility accorded to secondary relievers starting in the early and mid 2010s, based on what has to be Retrosheet data shown in a series of graphs for data from 1950 to 2022. To begin, the average number of pitchers used by teams in a month, at about 10 in 1950, reached maybe 12 in the 1970s and about 15 around 2010, but went up faster to maybe 18 by 2022. Most of this growth was in relievers, pretty steadily up from about 5 1950-1970 to 10. Those only starting were only at about 2½ in the 1950s but drifted up to close to 5. The loss was in swing men, maybe 3-4 early on but maybe 1-2 later. Looking at 1997-2022 only, the150 most used relievers in the league actually pitched fewer innings, down from about 10,000 to about maybe 9200; but innings from other relievers, about

4000 in 1997, went up to the 5000s from the late 2000s to early 2010s, and after that to over 8000. Consistently with this, returning to 1950-2022 data, the number of relievers per game jumped from up from 2 to almost 3½. Whereas overall back-to-back game usage rose only a bit, from maybe ⅓ to ½, in high leverage times (8th or 9th innings with one to three run lead) the percentage of back-to-back appearances went up from low 20s in mid 1950s to around 35 in 2000s, but down since then to the upper 20s; Russell would interpret the drop as due to the availability of more competent relievers.
Here are additional data making a related point; the percentage of games of various length handled by starter, 1993-2012:
8th inning, down from about 20 percent to less than 10.
7th inning, down from over 40 percent to about 20, with most of the drop starting about 2014.
6th inning, pretty steady in the mid to high 60s until 2014, then down to less than 50
5th inning, steadily over 80 percent until 2014, then down to about 70 percent.
4th inning, about 90 percent again until 2014, then down but not much, maybe 85 percent.
Note that the big dropoff was just about the time when the abundance of relievers became obvious, evidence that starters were pitching fewer innings after that because of that abundance.

Carleton, Russell A. (2023). And boom went the strikeouts.
https://www.baseballprospectus.com/news/article/83651/baseball-therapy-strikeouts-boom-or-bust/

Russell constructed three league average hitters using 2022 data, one actual (BA of .243, SLG of .395), one simulated high strikeout slugger (BA of .205, SLG of .445), and one low strikeout singles hitter (BA .285, SLG .360), and simulated run production with lineups consisting exclusively of each type. Their productivity was almost the same: real hitter 4.11 runs per game with standard deviation of 2.75, slugger 4.15 runs per game with standard deviation of 2.75, singles hitter 4.18 runs per game with standard deviation of 2.77. The moral of the story is overall player production is responsible for team performance no matter the manner in which it is achieved. Incidentally, the actual 2022 runs per game figure for 2022 was 4.28; as Russell noted, real life lineups include different production levels with batting orders designed to cluster the best hitters together.
In addition, Russell took five year periods starting with 1950-1954 up to the time of the webpost and correlated individual player (minimum 250 PA) strikeout rates with linear weight production per plate appearance, and found figures to vary across periods in the range of +/– 0.15, evidence that there really is none. In contrast, from 2018 to 2022, the home run/ production correlation was 0.561, and Russell claimed that it was similar across the period. Home runs beget runs.

Carleton, Russell A. (2023). Weren't pitch limits supposed to reduce injuries? https://www.baseballprospectus.com/news/article/84441/baseball-therapy-werent-pitch-limits-supposed-to-reduce-injuries/

The average plate appearances starters faced dropped from about 29 in 1950 to about 26 around 2010 and then more quickly to about 21 in the early 2020s; the mean pitches from 93 to 96 1950 to 2010 but down to low 80s in early 2020s. Also, the standard deviation of batters faced dropped from about 9 in 1950 to about 4.25 between 2013 and 2022, a sign of the standardization of starter workload. Yet, the number of UCL reconstructions, between 10 and 15 2000-2011, went up to between 20 and 30 2012-2022. This roughly parallels a stark change in starter usage patterns that truly began in 2014, in which they faced fewer batters but were expected to go full out at all times. The latter is what Russell believed responsible for the increase in injuries; data from elsewhere is supportive of this belief.

Carleton, Russell A. (2023). Is the ninth inning really harder?
https://www.baseballprospectus.com/news/article/84617/baseball-therapy-is-the-ninth-inning-really-harder/
NOT IN BIBLIOGRAPHY, IN REFEREMCES

Russell Carleton's (2023) goal in this webpost was to see if there has been something special about "proven closers." Using what is probably Retrosheet data from 2003 through 2022, he began by estimating the odds of strikeouts and homers in a given plate appearance based on seasonal K and HR rates for the pitcher and hitter. He then compared performance for pitchers in save (defined for this purpose as ninth inning or later, ahead by one or two runs) and hold (eighth inning, same lead). For relievers with at least 25 percent of batters faced in save situations, the ninth inning resulted in more of both K's and HR's. For relievers with at least 25 percent and at least 10 percent of batters faced in hold situations, HR's in save situations were up, but not K's. The implication is that there is a difference, a piece of evidence that closers are indeed a special breed. Russell's very speculative explanation was that batters try to hit homers in save situations, and true closers are able to exploit that tendency whereas other relievers are not.

Carleton, Russell A. (2023). The firemen will save us. https://www.baseballprospectus.com/news/article/81152/baseball-therapy-the-firemen-will-save-us-relievers-rule-changes/

The average number of mid-inning pitcher substitutions (Russell did not say, I am guessing per team per game) went up from less than 1.2 in the early 1950s to over 2 and still climbing after 2000 (as high as 2.3 in, I think, 2019), but the three-batter requirement dropped it to maybe 1.9 in 2021 and 2022.

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

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

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

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

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

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

Pickoff	24346	0.000	0.095	-0.197	-0.017	-0.119
Wild Pitch	56520	0.000	0.265	-0.001	0.000	0.263
Passed Ball	15238	0.000	0.259	-0.001	0.000	0.257
Balk	9624	0.000	0.253	0.000	0.000	0.253
Other advance	2502	0.000	0.063	-0.298	-0.040	-0.276
Foul Error	3284	0.000	0.000	0.000	0.000	0.000
Walk	607110	0.244	0.061	0.000	0.000	0.305
Intentional Walk	59403	0.185	0.004	0.000	0.000	0.189
Hit By Pitch	49877	0.251	0.078	0.000	0.000	0.329
Interference	918	0.254	0.109	0.000	0.000	0.364
Error	90717	0.288	0.205	-0.002	-0.001	0.490
Fielder's choice	26606	0.304	0.181	-0.371	-0.152	-0.037
Single	1252776	0.260	0.207	-0.003	-0.002	0.461
Double	314183	0.415	0.332	-0.002	-0.001	0.745
Triple	44499	0.590	0.430	0.000	0.000	1.020
Home Run	178776	1.000	0.404	0.000	0.000	1.404
Double play	192350	0.002	0.023	-0.325	-0.041	-0.341
Triple play	210	0.000	0.003	-1.015	0.000	-1.012
Total	8076015	0.114	0.083	-0.018	-0.034	0.15

Carleton, Russell A. (2023). Lost in the shuffle.
https://www.baseballprospectus.com/news/article/83225/lost-in-the-shuffle/

For players in at least 81 games during full seasons between 1950 and 2022, the average number of batting order positions in which they appeared increased from about 4 to 4½ in the 1960s, went back to around 4 in the 1990s, and then increased to 5½ and still rising in the early 2020s. The number of changes in the average lineup card, below 2 ½ in the early 1950s, went up sharply to 3 in the late 1960s, eased up to 3½ around 1980 and stayed there until about 2010, and then jumped to 4½ and still climbing in the early 2020s. For these players, controlling for average performance, pitcher, league, and handedness, there was a decrement in individual batting production when moved up in the lineup, worth about 8 OPA points, but no impact on moving down. Russell proposed two possible mechanisms. One was that moving up actually had a negative impact and moving down had none. The other was that managers moved players up when they were hitting particularly well and, as batting streaks basically occur randomly, the players then returned to normal production and were returned to their previous spot with no further change from normality.

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

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

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

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

Brian did strikeout rates similarly. In this case, 69.4 percent of variance was accounted for by real K rates but a better 73.5 percent by regressed rates, with 66, 18, and 16, or relative projection weights of 8/2/2. Note that not only are both the walk and strikeout seasonal weights markedly different from 5/4/3, they are also quite different from one another.

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

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

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

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

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

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

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

The first part of this two-part study of relative performance examined 108 players from the "mid-1990s" through 2008 with experience playing both in Japan and elsewhere, with U.S. data from Retrosheet. Brian took their Major League Equivalent figures (which would include minor league play) outside of Japan and compared them with Nippon Professional League performance. They were divided into the following five categories based on MLE HR percentage: A, greater than 0.65; B, 0.50-0.65, C, 0.30-0.50, D, 0.16-0.30, and E, less than 0.16.

U.S. totals.

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

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

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

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

Factor	AA	A	B	C	D	E	F
0.30	0.52	0.58	0.40	0.31	0.37	0.18	0.36
0.40	0.60	0.56	0.50	0.47	0.47	0.45	0.34
0.50	0.69	0.59	0.58	0.59	0.53	0.52	0.34
0.60	1.20	0.69	0.69	0.54	0.66	0.55	0.51
0.65	0.79	0.77	0.67	0.66	0.64	0.72	0.75
0.70	0.92	0.79	0.75	0.69	0.68	0.68	0.71
0.75	0.75	0.83	0.77	0.75	0.76	0.72	0.76
0.80	0.80	0.86	0.83	0.85	0.80	0.77	0.75
0.85	0.96	0.93	0.89	0.86	0.83	0.93	0.79
0.90	0.98	0.91	0.92	0.96	0.92	0.92	0.81
0.95	1.00	1.00	0.98	0.95	0.96	0.96	1.00
1.00	0.97	0.97	1.03	1.04	1.07	1.04	0.95
1.05	1.05	1.12	1.05	1.05	1.10	1.10	1.07
1.10	1.01	1.07	1.11	1.14	1.15	1.18	1.36

1.15	1.11	1.11	1.20	1.16	1.20	1.23	1.46
1.20	1.12	1.16	1.12	1.33	1.29	1.29	1.61
1.25	1.23	1.08	1.19	1.32	1.34	1.44	1.63
1.30	1.17	1.35	1.27	1.34	1.35	1.46	2.21
1.40	1.15	1.23	1.43	1.36	1.59	1.86	1.21
1.50	1.32	1.12	1.43	1.51	1.80	2.14	2.27
1.60	1.56	1.45	1.25	1.83	1.85	1.45	4.05
1.70	1.38	1.63	1.71	1.60	1.75	1.89	3.33
1.90	1.29	1.59	1.93	1.58	2.68	2.90	3.08

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

Cartwright, Brian (2009). So how long does it take for BABIP to become reliable?https://statspeakmvn.wordpress.com/2009/01/
Carleton, Russell A. aka Pizza Cutter (2007). DIPS and handedness. https://statspeakmvn.wordpress.com/2007/07/

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

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

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

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

A number of people have examined outfield throwing by using play-by-play data to compute the proportion of baserunners who advanced an extra base on a hit to a given outfielder along with the proportion of baserunners who were thrown out. Calculating the proportions for each outfielder allows the analyst to compare outfielder arms to one another. In addition, comparing run expectancies for before and after the play, these percentages can be turned into runs saved when a baserunner is thrown out or runs given up when baserunners take the extra base. Most likely the first such method was

Clem Comly's Average Run Equivalent Method (ARM), based on Retrosheet 1959 to 1987 data. Clem limited his analysis to singles with runners on first and/or second. The best annual figures in Clem's data were about 10 runs saved and the worst about 7 runs lost.

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

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

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

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

Two models for predicting specific game outcomes using Retrosheet data.