

Summaries of Research Articles and Books Using Retrosheet Data By Charles Pavitt

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.

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

Beyond base-out situation, the risk of attempting a steal (along with other speed-related moves such as taking extra bases on hits) depends on the specific abilities of the player making the attempt. Obviously, some players are better basestealers and/or baserunners than others, and the risk is lower the better the player is on the basepaths. Through a simulation based on the "team based on a given player" method for evaluating offense and using 2007-2009 Retrosheet data, Baumer, Piette and Null (2012) examined the expected outcomes of such attempts for 21 players purposely

chosen for their variety of capabilities as hitters and baserunners. Their results suggest that taking the risk of the steal or extra base is more productive long-term to the extent that the player is a good baserunner and a less productive hitter. This is because the cost of an out on the attempt is unsuccessful is greater for a better hitter than a poorer one. Although they interpret this in the context of the chosen individual players, the real implication is that attempting the steal or extra base makes more sense when the next batter is weak, as that next batter could use the help of the extra base for driving the baserunner in.

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

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

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.

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

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

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

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

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

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

retaliation hypothesis than those previous through use of Retrosheet 1973-2003 data. Based on game-by-game information, they first (2006) noted evidence for both hypotheses in predictive model allowing for determination of the order of importance of associated variables. The variable most strongly associated with hit-by-pitches was whether the game had designated hitters, with this effect occurred in interleague games including NL teams, evidence against the idea that HBPs are just idiosyncratic to the AL but perhaps due to pitchers not batting. However, the difference between leagues largely disappeared in the 1990s. On the other side of the dispute, the second most associated variable was total runs scored, evidence that when teams are hitting well the other side finds less reason not to hit batters. Further, home runs by the other team were also associated, more evidence that a HBP against a powerful batter would be considered less harmful. Finally, and not surprisingly general pitcher wildness was also correlated. In their second (2007) paper, Bradbury and Drinen determined whether a hit-by-pitch in one half inning increases the odds of retaliation in the next. According to two analyses, one for 1969 combined with 1972 through 1974, the other for 1989 through 1992, it does, as does a home run by the previous batter in the more recent data set; both of these findings support the retaliation hypothesis. Consistently with the second argument, higher OPS was positively associated with HBP whereas pitchers were less likely to be plunked than everyone else; both of these results suggest the "less harm" hypothesis. In addition, large score differentials increase HBP, likely because there is less harm when such a differential leaves less doubt concerning which team will probably win the game. Again, wilder pitchers are, not surprisingly, more likely to hit batters.

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

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

authors conclude with the interesting speculation that the reason that the overall league difference in HBP has disappeared may partly be due to the fact that the number of relievers used in a game has increased markedly.

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

John Charles Bradbury and Douglas Drinen (2008) is one 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.

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

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

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

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

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

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

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

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

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

		Balls			
		0	1	2	3
	0	28%	41%	40%	8%
Strikes	1	46%	40%	59%	56%

	2	49%	58%	65%	74%
--	---	-----	-----	-----	-----

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

The odds of a called strike if no swing

		Balls			
		0	1	2	3
Strikes	0	42%	40%	47%	63%
	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.

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.

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, 2008, 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.

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

A well-publicized paper by a University of Pennsylvania student named Elan Fuld that unpublished but easy to access online (search for "Elan Fuld clutch") claims that

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

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

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

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

A well-publicized (at least within the sabermetric community) study claiming support for the existence of hot and cold streaks can be found in an unpublished paper by Green and Zwiebel, based on Retrosheet data from 2000 through 2011. In essence using the "conditional probability" method for measuring streakiness, Green and Zwiebel wanted to see if the outcome of a particular plate appearance for both batters and

pitchers could be predicted more accurately using the outcomes of the previous 25 at bats than using overall performance for the given season, minus a 50 at bat window around the plate appearance under question. They provided various operational definitions for hot and cold streaks. Some of these definitions seem to bias the study in favor of finding streakiness; these established criteria based on the assumption that the average player is hot five percent and cold five percent of the time, which strikes me as out of bounds given that it presumes streakiness exists. A more defensible definition required the batter to be hot or cold if in the upper or lower five percent of a distribution based on his own performance. Their equations also controlled for handedness and strength of opposing batters/pitchers and ballpark effects. Unfortunately, the latter was poorly conceived, as it was based solely on raw performance figures and did not control for relative strength of the home team (i.e., a really good/bad hitting home team would lead to the measure indicating a better/worse hitting environment than the ballpark is in truth). The authors' results indicated the existence of hot/cold streaks for all examined measures: hit, walks, home runs, strikeouts, and times on base for both batters and pitchers.

Green and Zwiebel's work elicited a lot of critical comment. Along with the ballpark problem, which Zwiebel acknowledged in email correspondence with Mitchel Lichtman, one comment was that subtracting the 50 at bat window biased the study in favor of finding streaks. Here's an example showing why: let us assume that a player is a .270 hitter. If a player happens to be hitting .300 or .240 during that window, then the rest of the season he must be hitting say .260 or .280 to end up at that .270. In this case, the .300 and .240 are being compared to averages unusually low and high rather than the player's norm. But it strikes me this would only be a problem if hot and cold streaks actually existed – if not, it would be .270 all the way. It is the case that subtracting the 50 at bat window lowers the sample size of comparison at bats, increasing random fluctuation and again adding a bias in favor of finding streakiness. Whether this loss of 50 at bats is catastrophic during a 500 at bat season for a regular player is a matter for debate.

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

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

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

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

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

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

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

runs a year, whereas the worse cost 20 runs, when compared to the average.

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

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

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

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

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

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

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

Larrick, Timmerman, Carton and Abrevaya (2011), using all games with

Retrosheet data from 1952 through 2009 which included game temperature and controlling for pitcher control, discerned that the odds of a hit batsman increased as an interactive function of temperature and the number of teammates hit by the opposing team, such that more hit teammates resulted in more plunking of the opposing team, with this effect accentuated by hotter weather.

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

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

Occurrence	Result	Sample Size	Total	Left Field	Center Field	Right Field
Single with runner on first	Runner to third	31132	31.3%	19.1%	34.6%	49.4%
Single with runner on second	Runner scores	18399	65.3%	68.4%	82.6%	71.7%
Double with runner on first	Runner scores	6997	53.6%	40.5%	58.6%	37.7%

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

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

Of the several reasons proposed for the home field advantage in baseball, which is consistently measured at 53 or 54 percent, the most strongly backed by research is the presence of fan support, as home field advantage increases with rising attendance. Indirect corroboration comes from work by Lei and Humphreys (2013). They proposed a measure of game importance (GI), based on either how far a team leading a divisional or wild-card race is ahead of the second place team or how far a team not leading is behind the team that is. Smaller differences imply higher GI scores. Unfortunately, as

the authors note, their measure is not weighted by how far in the season a game occurs, so that GI will be the same for a team one game ahead or behind after the 1st as the 161st game. Anyway, in Retrosheet data from 1994 through 2010, GI was positively related with both attendance and home team winning percentage, with the latter implying that home field advantage rises as games become more important. The authors did not know to relate all three, but we can conjecture that game importance raises attendance which increases home field advantage in turn.

McCotter, Trent (2008). Hitting streaks don't obey your rules. *Baseball Research Journal*, Vol. 37, pages 62-70.

Pavitt, Charlie (2009). Hitting streaks and psychology. *Baseball Research Journal*, Vol. 38 No. 1, pages 6-7.

McCotter, Trent (2009). Reply. *Baseball Research Journal*, Vol. 38 No. 1, pages 7-8.

Albert, Jim (2008). Long streaks. *Baseball by the Numbers*, Vol. 18 No. 4, pages 9-13.

MCCOTTER/ALBERT DEBATE BRJ 39(2)

McCotter, Trent (2010). Hitting streaks don't obey your rules. *Chance*, Vol. 23 No. 4, pages 52-57.

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

Although nobody challenged Trent's analysis as such, there has been some criticism of other aspects of his work. His first attempts at explaining these patterns (batters facing long stretches of subpar pitching or playing in a good hitting ballpark, and streaks occurring more often in the warmer months) were proposed, found no evidence for the first, and claimed the second and third to be unlikely, but never empirically evaluated (although all could be). He instead opted for untestable speculations concerning a change in batter strategy toward single hitting and just the existence of a hot hand. I called him on these, and he responded with helpful analyses inconsistent with the second and third of the testable explanations and basically punted on the untestable ones. Jim Albert (2008) lauded the method and replicated it, but this time restricting the sample to five seasons of Retrosheet data studied separately (2004 through 2008). Again, real streaks occurred more often than in the random permutations, but only three out of twenty comparisons (for 5 or more, 10 or more, 15 or more, and 20 or more, for each of the five seasons) were significant at .05 and a fourth at .10, leading Jim to question the practical significance of Trent's results.

LATER DEBATE

A later paper (McCotter, 2010) added nothing substantive to the debate.

Mejdal, Sig (2000). The recipe for a stolen base. *By the Numbers*, Vol. 10 No. 3, pages 20-22.

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

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

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

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

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

71-79.

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

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

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

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

respectively for the latter two methods. Anyway, Rockoff and Yates estimated that allowing streaks to carry over across two seasons would increase the overall number by about ten percent.

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

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

Outs	Run Potential			Odds of Scoring		
	0	1	2	0	1	2
All runners	-.005	-.014	+.031	+.053	+.031	+.043
Fast runners	-.014	-.045	+.030	+.060	+.018	+.047

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

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

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

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

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

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

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

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

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

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

The most strongly supported explanation for the consistent 54% home field advantage for baseball is the impact of fan support. In one piece of relevant evidence Smith and Groetzing (2010) combined data for the years 1996 through 2005 from the

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

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

Tom Tango (2008) proposed a creative method for evaluating catcher ability to prevent passed balls and wild pitches, thwart attempted steals, and pickoff runners. For a given catcher:

- 1 – Choose a pitcher he caught.
- 2 – Count how many WPs, PBs, and innings occurred with that pitcher/catcher combination.
- 3 – Count how many WPs, PBs, and innings occurred with that pitcher and other catchers, and then use the ratio of WPs and PBs per inning to estimate the number that would have occurred if the other catchers had caught that pitcher the same number of innings and the catcher under examination.
- 4 – Comparing the results of steps 2 and 3 reveals how much better or worse the catcher under examination was than the other catchers for the given pitcher.
- 5 – Repeat these steps for all other pitchers the catcher under examination caught, and sum the results for an overall index.

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

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

I begin with an editorial comment: This book belongs on the shelf of anybody who seriously studies quantitative baseball data. The entire book is based on

sophisticated analysis using Retrosheet data (different seasons for different analyses, so I will skip the details on what seasons were employed). I will only list the themes, as describing all the findings would take too long:

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

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

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

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

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

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

platoon advantage if the pinchhitter is considerably better than the batter replaced.

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

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

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

addition, one should not use the same strategy all of the time as the other teams will respond accordingly with their defensive alignment, so randomly placed variation to decrease predictability will help.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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