By the Numbers

Volume 16, Number 2 The Newsletter of the SABR Statistical Analysis Committee May, 2006

Review

Academic Research: Pitcher Luck vs. Skill

Charlie Pavitt

The author describes two academic studies which try to break a pitcher's record down into luck and skill components.

This is one of a series of reviews of sabermetric articles published in academic journals. It is part of a project of mine to collect and catalog sabermetric research, and I would appreciate learning of and receiving copies of any studies of which I am unaware. Please visit the Statistical Baseball Research Bibliography at www.udel.edu/communication/pavitt/biblioexplan.htm. Use it for your research, and let me know what is missing.

James Albert, A Baseball Hitter's Batting Average: Does It Represent Ability or Luck?, Stats, 42, 2005

James Albert, Pitching Statistics, Talent and Luck, and the Best Strikeout Seasons of all-Time, Journal of Quantitative Analysis in Sports, Volume 2 Number 1, 2006, Article 2

Our fearless and peerless editor Phil Birnbaum alerted me to the recent start-up of the *Journal of Quantitative Analysis in Sports*, in which I found the later of these articles and a reference to the earlier one; the author was kind enough to email me a copy of the *Stats* piece.

I shall begin with the later one, as I read it first.

Its explicit goal is to present a method for evaluating relative pitching performance within given years, allowing their comparison across time, using strikeouts as a specific example. This has, of course, been done many times before. Jim's added twist is to propose an analytic technique conceptually analogous to the types of parametric statistical procedures social scientists work with, which decomposes the total variance in performance across pitchers into two components: the explainable "systematic" variance, in this case representing pitchers' true skill, and the unexplainable "error" variance, in this case representing random fluctuation/luck. Jim presented a table revealing the estimated proportion of random fluctuation for seven pitching-relevant measures for 200, 500, and 1000 batters faced; the proportion for randomness of course went down substantially as batters faced went up, and I will use the 500 number for illustration.

What I found most interesting here is the contrast across different pitching-relevant measures in terms of skill versus luck variance. Strikeouts at 500 BFP were estimated at only 12 percent luck, which is clearly consistent with the Defense Independent Pitching Stats implication that strikeouts are a valid measure of pitching skill. Another DIPS measure, walks, comes in tied for second with runs and earned runs allowed, at 20 percent luck, followed by batting average (37 percent), home runs (56 percent), and hits other than home runs (66 percent). The hits result is again consistent with DIPS, but the home run finding is not; the DIPS concept implies it should be less random than overall batting average. Anyway, z-scores based on the distribution of skill variance shows Dazzy Vance as having four of the all-time top eight strikeout seasons for starters in relative terms (1923

In this issue

through 1926), and Rob Dibble as having three of the top seven and four of the top fourteen for relievers (1989 through 1992).

The earlier one does not include comparisons over time but rather concentrates on the skill variance versus luck variance distinction, this time for batting. As such, it is a good companion piece, although the analyses were performed differently. In this case, Jim found batter strikeouts to be the most strongly based on ability among a set of hitting-relevant measures, followed by home run rate, ability to get walks, OBP, BA on balls in play, total BA, rate of doubles + triples, and, most based on luck, ability to hit singles. This ordering is even more consistent with DIPS than that for pitching.

Jim's piece in the last BTN can be seen as a follow-up to the older of these two efforts (the newer one is easy to find; just google Journal of Quantitative Analysis in Sports). The BTN piece replicates the basic skill vs. luck analysis using a different data set (2005 rather than 2003) and extends the Stats work by exploring the relationship among skill-only estimates for walks, strikeouts, homeruns, and in-play hits. Anyway, as I trust he will be reading this – good job, Jim.

Charlie Pavitt, chazzq@udel.edu ♦

Informal Peer Review

The following committee members have volunteered to be contacted by other members for informal peer review of articles.

Please contact any of our volunteers on an as-needed basis - that is, if you want someone to look over your manuscript in advance, these people are willing. Of course, I'll be doing a bit of that too, but, as much as I'd like to, I don't have time to contact every contributor with detailed comments on their work. (I will get back to you on more serious issues, like if I don't understand part of your method or results.)

If you'd like to be added to the list, send your name, e-mail address, and areas of expertise (don't worry if you don't have any - I certainly don't), and you'll see your name in print next issue.

Expertise in "Statistics" below means "real" statistics, as opposed to baseball statistics - confidence intervals, testing, sampling, and so on.

Member	E-mail	Expertise
Ben Baumer	bbaumer@nymets.com	Statistics
Jim Box	jim.box@duke.edu	Statistics
Keith Carlson	kcsqrd@charter.net	General
Dan Evans	devans@seattlemariners.com	General
Rob Fabrizzio	rfabrizzio@bigfoot.com	Statistics
Larry Grasso	l.grasso@juno.com	Statistics
Tom Hanrahan	Han60Man@aol.com	Statistics
John Heer	jheer@walterhav.com	Proofreading
Dan Heisman	danheisman@comcast.net	General
Bill Johnson	firebee02@hotmail.com	Statistics
Mark E. Johnson	maejohns@yahoo.com	General
David Kaplan	dkaplan@UDel.Edu	Statistics (regression)
Keith Karcher	karcherk@earthlink.net	Statistics
Chris Leach	chrisleach@yahoo.com	General
Chris Long	clong@padres.com	Statistics
John Matthew IV	john.matthew@rogers.com	Apostrophes
Nicholas Miceli	nsmiceli@yahoo.com	Statistics
John Stryker	john.stryker@gmail.com	General
Tom Thress	TomThress@aol.com	Statistics (regression)
Joel Tscherne	Joel@tscherne.org	General
Dick Unruh	runruhjr@dtgnet.com	Proofreading
Steve Wang	scwang@fas.harvard.edu	Statistics

"The Wages of Wins" – Right Questions, Wrong Answers

Phil Birnbaum

The author reviews the recently released book "The Wages of Wins," which applies sabermetric and econometric methods to sports analysis.

Introduction

A good description of "The Wages of Wins" [TWOW] is Alan Schwarz's cover quote: "Freakonomics meets ESPN." Authors David J. Berri, Martin B. Schmidt, and Stacey L. Brook, three academic economists, analyze data from three sports to, in the words of the subtitle, "[take] measure of the many myths in modern sport."

While the subject area is ostensibly economics, only a couple of the chapters deal with traditional sports economics issues. The majority of the text attempts to find performance measures for basketball and football; the subject matter could easily be described as sabermetrics of several sports. Indeed, the authors are serious followers of basketball -- almost half the book is devoted to analysis of that sport.

To its credit, the authors describe the studies and results that lead to their conclusions, unlike "Freakonomics," which discuss the implications of the findings rather than the logic that led to those findings. Some of their regression findings are presented (in suitable simplified form), so readers can connect the studies' results to the authors' conclusions.

The book has garnered good reviews and blurbs, from Alan Schwarz to (famed economist) Deirdre McCloskey, to a long New Yorker review by Malcolm Gladwell. The book is indeed a decent read.

However, as for the details of the authors' studies, in my judgment, many of the authors' conclusions are incorrect.

Wins and Payroll

The chapter entitled "Can You Buy the Fan's Love?", for instance, discusses the effects of MLB team payroll versus performance. The authors regress payroll on wins, and find an r-squared of .176. From this, they argue that "payroll and wins are not strongly linked," because the explanatory power of wins is only 18%.

But an r-squared of .176 *is* a strong relationship between payroll and wins! The statement "payroll explains 18% of wins" is true only in a narrow, mathematical sense -- that the reduction of sums of squares from the regression line after adjusting for salaries is 18% of the total sums of squares from the mean before the adjustment. It doesn't answer the relevant question, which is: what is the relationship between payroll and wins in the baseball context?

For that question, the important number is not the r-squared of .176, but its square root, the correlation coefficient. The square root of .176 is about .42. This means that for every additional standard deviation in salary a team spends, it will improve .42 of a standard deviation in wins. Roughly speaking, 42% of a team's spending will show up in the win column.

That's pretty large, considering how much luck there is in a team's record. A superstar is worth about five wins above an average player. The SD for a team of known talent is six wins. That means that almost half the time, luck is a bigger factor than, say, Derek Jeter. Under these conditions, a correlation of .42 is a fairly large factor.

The authors note that $5 million in payroll buys an additional win. "That's all the bang a team would get for their buck," they write. " … a team would have to add several $10 million players before they could expect to see any real progress in the standings."

But, of course, teams *are* spending that much. In 2005, the Yankees spent about $140 million above the median. The difference, according to the authors' study, should be about 28 wins above average (for an expected 109-53 record). The Red Sox, at $123 million, would start with a 14 win advantage. And the Devil Rays, with a payroll of only $30 million, would have a 7-game disadvantage.

So if the Yankees start with an expectation 35 wins above Tampa Bay, and the Red Sox start 21 games ahead, doesn't that imply that salary is important?

The authors, again, say no, because of the 18% figure.

This might be a nitpick if the authors' argument was a minor one. But it's the primary thesis of the chapter, and the inspiration for the title of the book. The authors start the chapter by arguing , reasonably, that looking at post-season victories is not a good way to check the effects of salaries. They argue, again reasonably, that MLB's "Blue Ribbon Panel" looked only at 1995-1999, the period in which salary had the largest effect. But when it comes to data to actually determine the relationship in the way the authors consider most appropriate, this is the only study they use. And they draw a conclusion that contradicts what the data actually show.

Competitive Balance

There is a fear that, if revenue inequality among teams continues to increase, some teams will buy up all the free agents, other teams won't be able to afford any, and competitive balance will continue to worsen, perhaps fatally to the game.

In Chapter 4, the authors argue against this eventuality. First, they write, studies have shown that attendence rises the more uncertain the outcome of the game. If the Yankees turn into a team of Babe Ruths, and the Royals turn into a team of Danny Ainges, a Yankee victory is almost preordained, and fans won't show up to the game, even in New York. So the Yankees have a vested interest in leaving enough quality players for other teams.

In addition, adding more and more superstars has less and less effect on a team's win totals, and therefore revenue. As a team gets better, there are fewer and fewer losses for it to turn into wins. Derek Jeter might be worth five wins to an average team, but only, say, three wins to the 1998 Yankees. Again, this puts a natural limit on how much a wealthy team will be willing to spend on players, and, by extension, on competitive imbalance.

Not only is there a natural limit to competitive balance, but revenue sharing won't even help. One of the most famous results in economics is the Coase Theorem, which says that resources (players) wind up going to the firms (teams) to whom they have the most value, regardless of who actually owns them at any given time.

The Wages of Wins: Taking Measure of the Many Myths in Modern Sport

By David J. Berri, Martin B. Schmidt, and Stacey L. Brook

Stanford University Press, 304 pages, $29.95 (US), ISBN 0465005969

That is, suppose Derek Jeter is worth $10 million in revenue (increased attendance, TV viewership, playoff chances) to the Yankees, but only $5 million to the Pirates. The Coase Theorem says that even if you force the Yankees to share revenue with the Pirates, and even if you allow the Pirates to draft Jeter in the first place, he will be sold to the Yankees (or traded for cheaper players) simply because he's worth more in New York.[1] This is good stuff and well explained.

Finally, the authors show measures of competitive balance (as measured by W-L records) for fifteen different sports leagues. It turns out that soccer has the most balance, followed by football, hockey, baseball, and, most unbalanced, basketball.

The authors argue that this is the result of the population from which players are taken. Hundreds of millions of people in the world play soccer, so there is an abundance of talent. But basketball is limited mostly to people who are very tall, and there are a lot fewer of those. So, with tall basetball players in short supply, the leagues are filled with inferior players, which allow the best players to dominate.

It sounds plausible, but the problem is that it's the rules of the particular sport that determine competitive balance. A simple thought experiment can show why this is true. Suppose that instead of basketball games being 48 minutes long, they were 480 minutes long. With ten times the opportunity for luck to even out, the better team is much more likely to win the game – perhaps it becomes a .900 team instead of a .600 team. Or suppose the game is shortened to 4.8 minutes. In a game that short, even a markedly inferior team could win; perhaps it becomes a .450 team instead of a .300 team.

[1] One implication, though, which the authors do not address, is that a payroll tax *would* increase competitive balance. If the Yankees have to pay a $10 million tax on Jeter's $4 million salary, but the Pirates don't, he becomes unprofitable to the Yankees ($14 million cost, $10 million benefit) but remains profitable to Pittsburgh ($4 million cost, $5 million benefit).

Competitive balance is partly a consequence of the rules of the game. The more opportunity for luck to even out, the more likely the better team is to win, and the more unbalanced the sport looks.

Let's compare baseball and basketball. I'll oversimplify a bit to make the comparison easier, but the argument will stand even if the details are made more realistic.

- In basketball, each team has 100 ball possessions in which to score; each possession, they score about 50% of the time. In baseball, each team has about 40 plate appearances in which to put men on base; in each plate appearance, they get on base about 40% of the time.

  So basketball has 2.5 times as many chances for the better team to assert its superiority. Since the standard deviation of scoring rate is proportional to the square root of opportunities, we can say that basketball has over 50% more imbalance than baseball in this respect.

- In basketball, the team that scores most wins. In baseball, the team that gets on base more doesn't necessarily win -- it depends *how* it gets on base, and whether those successes are bunched into a relatively few innings.

  So, in basketball, the team with the better success rate wins. In baseball, there's a big "luck" factor (assuming that clutch hitting ability does not dominate) that allows the weaker team to beat the stronger team despite being outperformed in the basic success rate. Again, in this regard, basketball has much more imbalance than baseball.

- In basketball, five players play most of the game. The average player is 20% of his team's performance. In baseball, nine hitters play most of the game. The average player is 11% of his team's offense. (And even a workhorse starting pitcher is only 15% or so of his team's innings.)

  So in basketball, the average player has much more impact on the outcome than in baseball. That allows the superstars to play more, and the average players less, which again means more imbalance in basketball.

- In basketball, a superstar might take 40% of his team's shots. In baseball, every player must bat in turn, so even the leadoff hitter will have no more than about 12% of his team's plate appearances.

  Again, this means more imbalance in basketball, because the superstars can dominate.

Add up these factors, and it becomes evident why basketball has less competitive balance than baseball; the game is structured so that the team with the best players is much more likely to win. These are four powerful theoretical reasons why records should be more extreme in basketball even before considering demographics.[2]

Is there any empirical evidence that can be examined on the question? There is: home field advantage.

In baseball, the home team has a winning percentage of about .540 – 40 points above normal. In basketball, the home team has a winning percentage of around .625 – three times as high.

In 2002-2003, the .488 Seattle SuperSonics were .610 at home (the equivalent of 99-63), but only .366 on the road (the equivalent of 59-103). This can't be explained by the demographics of height.

But it *can* be explained by game structure. Home field makes one team better and the other team worse. In basketball, the game is structured so that the better team wins with high probability. In baseball, the game is structured so that the better team wins with lower probability.

Another way of phrasing the difference is that the structure of basketball gives the illusion that basketball leagues are more unbalanced in terms of talent. They may be; they may not be. But they certainly are more unbalanced in terms of game results.

It's still possible that the authors' "short supply of tall people" theory makes some contribution, but I'm not convinced. It is indeed true that "you can't teach height," which means that the supply of basketball players is limited to a small subset of the population. But, in baseball, "you can't teach sight" either -- if Ted Williams really had 20/10 eyesight, and if extremely good vision is as important in baseball as height

[2] After the first draft of this article was written but before going to press, another review of this book, by Roland Beech, made this same point. That review can be found at http://www.bepress.com/jqas/vol2/iss3/5/ .

is in basketball, you have an analogous situation for baseball. There are probably many important attributes in any sport that are intrinsic and can't be taught, some of which we're not even aware of. (Wayne Gretzky was able to "see" the game in slow motion.) In the absence of any argument that height in basketball is more important than similar attributes in other sports, I remain agnostic.

Basketball

This being a baseball publication, I will summarize the extensive basketball portion of the book only briefly.

In 2003, Dean Oliver's "Basketball on Paper" was published. Of the basketball books I've seen, it's the closest in spirit to the Bill James Baseball Abstracts, and many of Oliver's insights are reminiscent of James. For instance, just as Bill James noted that the out (and not the at-bat) is the currency of a baseball offense, Oliver notes that for basketball, it's points per possession, and not points per game, that's important. That's because every game has a different pace and a different number of possessions, so a team that scores 100 points per game in 100 opportunities per game is a better offense than a team scoring 102 points per game in 105 opportunites per game.

Oliver argues that players should be evaluated per posession. A player would be credited with his team's possession when he ends it -- by shooting, taking free throws, or turning the ball over. A player is credited with the points he is responsible for, but assists are divided between the scorer and the passer.

It's from this base of knowledge that "The Wages of Wins" begins.

First, the authors note that since a possession, on average, scores one point, then a turnover costs one point (being the forfeiture of a possession). A field goal made turns a one-point potential into a two point score, so is also worth one point above average. And a successful 3-point shot is obviously worth two points. The authors list a chart of almost everything that can happen during a possesion, and how many points it's worth. They note that every 30 points equals one win (like 10 runs in baseball), and so they divide by 30 to get a win value for each event in terms of wins. Then, because guards and frontcourt players have different per-game averages, they adjust for position.

And for defense, they take the team's defensive bottom line, and allot it to players based on minutes on the court.

But then, they make what I think is a critical error. In converting player wins into a rate, they divide by minutes played, instead of by possessions used.

This is a problem for several reasons. First, consider two players who always play together; they have exactly identical results per shot, but one takes twice as many shots as the other. Using minutes, one will look twice as good as the other. Second, teams who play a faster-paced game and get more possessions than average will have their players look better than equally-talented players who play on slower teams. Finally, you'll get wrong results when composing hypothetical teams. If the 1995-96 Michael Jordan scored .386 wins per game (48 minutes), does that mean five Jordans would have scored five times that? No -- because if Jordan takes 40% of his team's possessions, five Jordans would need to take 200% of the team's possessions, which isn't possible.

These issues can seriously change conclusions and rankings. Some teams are almost ten percent faster than others (meaning 10% more points can be scored by players on those teams). Further, some players take almost 30% of their team's possessions – one-and-a-half times as much as average. A player could rate some 65% higher than another similar player for these reasons alone.

This flaw, in my opinion, renders three chapters worth of results unreliable, at least those results based on the rate stat.

Clutch and Consistency

In an attempt to find whether some basketball players improve in the clutch, the authors compare players' performances in the regular season to those in the playoffs. They find that performance generally drops. Could that be simply because, in the playoffs, their opponents are limited to the league's better teams? Astoundingly, in eight pages of analysis, the authors don't mention the possibility even once!

Finally, the authors turn to players' consistency. Using a percentile grading system, they find that only 12% of basketball players moved more than two grades (up or down) between seasons. In MLB, 28% moved up or down two grades. But for NFL quarterbacks, the figure was 39%. The authors argue that quarterbacks are "consistently inconsistent," and the difference is caused by team factors such as the quality of the offensive line. That's probably a substantial part of the answer, but I suspect most of the difference is just luck. Quarterbacks have fewer opportunities than basketball players (and about the same as baseball players) – and the standard deviation of outcomes is very high.

Regardless of the reasons, I agree with the authors' conclusions that overreliance on QB statistics is a mistake.

Complexity

Some of the most difficult and interesting problems in the analysis of free-flowing sports is how the players affect each other in ways that are hard to measure.

For instance, suppose player A scores more assists than player B on another team. Is that because A is better at hitting his man near the basket, or is it because his teammates are better at getting open?

If player C is more successful than player D, is it because C is better at shooting, or because the opposing defense is concentrating on covering Michael Jordan, giving C more room to get to the basket?

If player E has a poor shooting percentage, is that because he's not accurate, or because his team often gets him the ball with no time left on the 24-second clock, forcing him to take desperation shots?

All these factors influence a player's stats, but the book doesn't study any. And that's fine -- with the limited statistical record the authors are limited to working with, it probably isn't possible. But what's frustrating is that the authors don't even mention them. Having created and run their model, they proceed as if the problem has been solved.[3]

The authors are very critical of Allen Iverson, arguing that his productivity ranks far below his reputation and traditional scoring statistics. And despite the imperfections of the measures used, the numbers are fairly convincing that the authors are correct. However, given that the numbers highly contradict conventional wisdom, might there be factors the authors didn't consider?

It took me only a few minutes of web searching to find an article suggesting that Iverson, who takes many of his team's shots, saves his teammates from having to take difficult shots by doing so himself. That might explain the disconnect between his high scoring and his low overall rating. It would also suggest that he's not as bad as TWOW suggests, as he's taking a hit to his personal stats to the benefit of his teammates' stats (and hopefully, to the benefit of the team too, if he's the player with the best chance of sinking the hard shots).

Now, I don't know if this is true. But it might be, and the authors don't give it a thought.

Football

TWOW's football chapter can be summarized by one regression result: to evaluate a quarterback's productivity, (1) take his total yards passed plus rushed; (2) subtract 3 for every play; and (3) subtract 50 for every turnover. The result is followed by several pages of player charts and commentary.

I'm sure the formula accurately reflects the regression result. The problem with this method is that it doesn't take the situation into account.

A two-yard rush is worth negative one point by this system, but on third-and-one, it's exactly what's called for, and so an unqualified success. And three consecutive passes of four yards each are worth exactly as much, in real life, as a single 12-yard pass, but the system values them differently. Quarterbacks with different teams who can run more plays (but be equally successful in gaining enough yards to maintain possession) will be underrated by this formula; quarterbacks who lose the ball equally as often after an equal number of yards, but in fewer plays, will be overrated.

Again, I'm not sure how accurate or inaccurate the system is given these difficulties. However, the authors seem unaware of these issues, arguing that it's "just about everything one could want in a performance measure." But a more sophisticated approach was given by Carroll, Palmer, and Thorn in "The Hidden Game of Football" in 1989. That book does not appear in TWOW's bibliography.

---

[3] Actually, the authors do consider whether a player can make his teammates better -- but by "better," they mean "better in their total value statistics." And, with Michael Jordan taking a high percentage of his team's opportunities, other players' stats are not going to improve when MJ shows up.

Overall

There is a common pattern the authors use throughout the book – run a regression, explain the findings, assume the problem is now solved, then dismiss conventional wisdom because it doesn't use regression. On page 7, the authors express dismay at the "laugh test" – the tendency to dismiss analytical findings if they contradict conventional wisdom -- correctly pointing out that research trumps intuition. But the authors go too far the other way. With equal lack of justification, they unthinkingly reject any non-statistical opinion that contradicts their results. Which I think is why they're off the mark so often: they fail to consider that their analysis may be incomplete, or may not completely capture what they're trying to measure. They give opposing views no benefit of the doubt, and their own views get no doubt at all.

And so readers with a decent knowledge of sabermetrics will find this book frustrating -- there isn't much new, the authors are inappropriately immodest, and many of the results, I think, are just plain wrong. One particular frustration is that the authors seem unaware of previous research. On page 41, when claiming that money doesn't buy MLB wins, they suggest that maybe GMs don't know that players decline after age 28. Bill James' study on aging dates back to 1982, and players peaking in their late-20s has become conventional wisdom since then. Not only are the authors unaware of this (their reference is an academic working paper from 2005!), but they blithely assume that general managers know little about the product they're putting on the field.

But having said all that, the book may still be worth a look, if you can get the gnashing of teeth every page or two. The authors write clearly, and they raise interesting questions. If you're looking for the answers, this may not be the place -- but there are many studies that suggest themselves out of the issues the authors raise.

Phil Birnbaum, birnbaum@sympatico.ca ♦

Corrections

In the original version of this issue of BTN, the name of Bryan Reynolds, one of the contributors, was misspelled due to an editing error.

Also, on page 5, the Seattle SuperSonics were mistakenly referred to as the "Seahawks."

Both articles have now been corrected. We regret the errors.

Get Your Own Copy

If you're not a member of the Statistical Analysis Committee, you're probably reading a friend's copy of this issue of BTN, or perhaps you paid for a copy through the SABR office.

If that's the case, you might want to consider joining the Committee, which will get you an automatic subscription to BTN. There are no extra charges (besides the regular SABR membership fee) or obligations – just an interest in the statistical analysis of baseball.

The easiest way to join the committee is to visit http://members.sabr.org, click on "my SABR," then "committees and regionals," then "add new" committee. Add the Statistical Analysis Committee, and you're done. You will be informed when new issues are available for downloading from the internet.

If you would like more information, send an e-mail (preferably with your snail mail address for our records) to Neal Traven, at beisbol@alumni.pitt.edu. If you don't have internet access, we will send you BTN by mail; write to Neal at 4317 Dayton Ave. N. #201, Seattle, WA, 98103-7154.

The Interleague Home Field Advantage

Eric Callahan, Thomas J. Pfaff, and Bryan Reynolds

In 2005, the home field advantage in interleague games was much higher than overall. Is that also the case for other years? Here, the authors investigate.

Introduction

The difference between the National and American Leagues regarding the designated hitter rule creates different styles of play and strategy in the two leagues. In the National League, managers face more situational decisions throughout the course of the game, decisions seldom faced by an American League manager. In the American League, all nine positions in the batting lineup are occupied by reasonably good hitters (compared to the pitchers). On the other hand, in the National League, there is no one place in an American league order that is significantly weaker than the others. Due to this difference, one might expect that the home team would have a greater advantage in interleague play than during intra-league play. Our initial investigation included comparing the home field advantages of interleague and intra-league play in 2004 and 2005. In those two years, each league had a higher home winning percentage against the other league than they did against their own league. In one of the four cases the difference was significant with a p-value of 0.005. Due to this, we chose to examine the interleague winning percentages dating back to 1997, the beginning of interleague play.

Results

Table 1, with data taken from mlb.com and Retrosheet, gives the home winning record and winning percentages for the American League for interleague and intra-league games. The p-value is from a two proportion test with a two-sided alternative comparing the two winning percentages. In only one of the nine tests, the 2005 season, were we able to say that there is a significant difference in the winning percentages. In two of the years, the difference is notable with the p-values around 0.1. In these three cases the interleague record was better that the intra-league record. In five of the nine years, the interleague winning percentage was greater than the winning percentage in regular games.

Table 1 – American League Home Field Record

Year	Total			Interleague			Intra-league			Win % Difference	p-value
	W	L	Win %	W	L	Win %	W	L	Win %		
1997	579	539	.518	61	46	.570	518	493	.512	.058	0.252
1998	596	537	.526	56	56	.500	540	481	.529	-.029	0.561
1999	581	551	.513	61	65	.484	520	486	.517	-.033	0.488
2000	609	523	.538	76	49	.608	533	474	.529	.079	0.090
2001	598	534	.528	75	51	.595	523	483	.520	.075	0.105
2002	608	524	.537	64	62	.508	544	462	.541	-.033	0.487
2003	615	519	.542	62	64	.492	553	455	.549	-.057	0.231
2004	627	506	.553	75	52	.591	552	454	.549	.042	0.367
2005	611	523	.539	82	44	.651	529	479	.525	.126	0.005
Total	5424	4756	.533	612	489	.556	4812	4267	.530	.026	0.103

Table 2 repeats the same analysis for the National League. Similar to the AL, we are only able to conclude that the National League interleague winning percentage was significantly better once, the 1997 season, in nine years. In one other year (2004), the p-value was a notable 0.125. The National League was better against the American League than they were against themselves six times.

Table 2 – National League Home Field Record											
Year	Total			Interleague			Intra-league			Win % Difference	p-value
	W	L	Win %	W	L	Win %	W	L	Win %		
1997	622	501	.554	70	37	.654	552	464	.543	.111	0.022
1998	712	586	.549	54	58	.482	658	528	.555	-.073	0.141
1999	683	612	.527	70	55	.560	613	557	.524	.036	0.440
2000	704	592	.543	66	60	.524	638	532	.545	-.021	0.646
2001	675	621	.521	69	57	.548	606	564	.518	.030	0.525
2002	706	587	.546	67	59	.532	639	528	.548	-.016	0.735
2003	720	575	.556	73	53	.579	647	522	.553	.026	0.576
2004	672	623	.519	74	53	.583	598	570	.512	.071	0.125
2005	695	601	.536	72	54	.571	623	547	.532	.039	0.402
Total	6189	5298	.539	615	486	.559	5574	4812	.537	.022	0.164

Conclusions

Even though the differences in rules of each league might suggest that there is an interleague home field advantage, the results of the proportion tests aren't definitive. Still, the data may suggest a greater home field advantage during interleague play. Consider that of the 18 comparisons 11 showed a higher interleague home field winning percentage. Moreover, of the 7 cases when the interleague home winning percentage was smaller than the intra-league percentage, the average difference was -0.037, interleague home winning percentage minus intra-league home winning percentage, with the worst year being the 1998 National league record with a difference of -0.073, resulting in a p-value of 0.141. On the other hand, of the 11 cases when the interleague winning percentage was better, the average difference was 0.063, with two p-values below 0.05 and three others between 0.09 and 0.125. All five of these p-values were smaller than the smallest p-value of 0.141 from the other 7 cases. Lastly, it is worth noting that the difference in the winning percentage on the totals for the nine years is similar for both leagues with a 0.026 for the American League and a 0.022 for the National league. In fact, if you combine both league totals you get 1227 wins out of 2202 games for interleague play and 10,386 wins out of 19,465 games for intra-league play, which yields a p-value of 0.034.

There are few other interesting results from this data. For one reason or another, the interleague winning percentages versus intra-league winning percentages in the American and National Leagues seem to correlate with one another. In six of the nine seasons, both leagues either have a better interleague winning percentage (four times) or both have a worse interleague winning percentage (twice) when compared to the intra-league home winning percentages. Of the three times they were different, the National League was better twice and the American League was better once. This could be evidence that in those three years, one league was better than the other league. In these years, the better league had a higher winning percentage at home and on the road in interleague games. This means that the particular league won a majority of all the interleague games. In two of those three years, the league that seemed to have been better won the World Series. In six of the nine years of interleague play, the league with the higher interleague home winning percentage won the World Series, including the last three.

Acknowledgements

We would like to thank Jim Conklin, Stan Seltzer, and Steven J. Miller for their helpful comments on pervious versions of this paper.

Eric Callahan, 321 Center St., Apt #2, Ithaca, NY, 14850, NYClar31@hotmail.com
Thomas J. Pfaff, Department of Mathematics, Ithaca College, Ithaca, NY 14850, tpfaff@ithaca.edu
Bryan Reynolds, 214 Coddington Road, Ithaca, NY 14850, breynol2@ithaca.edu ♦

Best of the Ball Hawks

Tom Hanrahan

Who were the top defensive centerfielders in baseball history, by peak, prime, and career skill? Here, the author uses two established and respected fielding measures to figure which score the best.

Of all of the men who have played major league baseball, who was the very best defensive outfielder? The answer is ... given later in this paper. More fun than the answer, though, is the process; how should we attack the problem?

First, how, should we define best? Best at his *peak*? Throughout his *prime* days? Best *career*, which rewards the player's ability to still be on the top of his game while in his mid-thirties? I'll look at each of those. In the end, what I am looking for is this: if I had to construct a team whose goal was to prevent the opponents from scoring, who would I want in the outfield?

"Peak" will mean "who was best at their very best?" The difficulties here reside in small sample size; the fewer years I use, the more reliant the metrics are on irritatingly random elements. I will define "peak" as a player's *best five (not necessarily consecutive) seasons*.

"Prime" will be a bit like peak, except the player needs to be consistently great over a longer period. A few extra boo-boos during one bad season, or a slow healing hamstring will be less forgiving by this metric. I will define "prime" as a player's best *eight consecutive seasons*. If interrupted by military service, I will use fewer than 8 years only if the missing seasons are sandwiched on both sides by the prime period (the same will hold for the CAREER evaluation).

"Career" ideally ought to measure a player's entire career. However, defensive ability declines with age, seemingly even more consistently than hitting ability. And very few men played at a gold-glove level in the outfield when they were in their late 30s, which makes comparisons difficult. I will use a minimum of a player's *twelve best seasons* to define his "career" performance, and will also compare additional years among those very few who played very well for even longer.

Value versus ability

How do we *determine* "best"? Well, this is the toughie, and I'll spend the first part of this paper outlining my methods.

Value is always a combination of effectiveness times playing time, whereas *ability* is only concerned with effectiveness. Players who don't play aren't of any value, no matter how good. But what I am most interested in here is not the total value a player achieved, but how effective he was at a certain task; preventing runs by tracking down fly balls and throwing out base runners. Using playing time introduces the problem that a player may get less time because of his lack of ability in some basic areas; hitting well enough to stay in the game, or not being injured. Again, while these are crucial to providing value, it does not help answer the question of who would I want out in center field when trying to protect a lead.

So, for this paper, I will only use rate statistics, and count each season of at least 100 games played as one full year. Players will obviously need to play well enough (and stay healthy enough) to remain on the field for more-or-less full seasons, but aside from this, a season of 150 games will not count for more than a season of 115, if the level of performance is the same.

So, am I measuring value or ability? By using rate statistics, this paper mostly is measuring ability; however, requiring a minimum playing time ensures that a good deal of value is gained from that ability.

What fielding metrics to use?

Although much progress has been made in recent years, measures of fielding prowess are still a lot fuzzier than for hitting. No one argues that Hank Aaron's 6856 total bases (over 700 ahead of Stan Musial in 2nd place) is an achievement that speaks loudly of his greatness and durability as a hitter; but can the same be said of Willie Mays' 7095 putouts or Tris Speaker's 449 assists? Also, the most promising methods that are being developed today use detailed play-by-play data that is not available for most of MLB's history, and thus cannot be used to compare Joe DiMaggio with Curt Flood.

There are two methods that analyze fielding ability that have been generally well-received by the sabermetric community, are published so all may access them, and give rate statistics that can be used for this study. One is Fielding Win Shares (FWS), created by Bill James in his book *Win Shares*. The other is the set of player ratings found online at Baseball Prospectus,[1] which are under the care of Clay Davenport. I will refer to these systems as WS and BP. Both of these take into account not only the fielder's "range" (putouts), but his arm (assists) and sure-handedness (errors), all adjusted for the team context.

There are cautions to be used with the use of WS and BP. Bill James has indicated that he now considers Win Shares as published a bit "out of date", as he has been working on a major modification to the system; but this update is not ready. Also, the BP numbers are somewhat fluid; the values on the website have been known to change on occasion, as the authors fine-tune their work. The numbers I quote here were valid as of March, 2006.

The timeline, or league strength, could of course be a consideration. BP has a league-quality-adjusted rate (called Rate2 on their web site), and so I entered all of BP's Rate 2 stats to compare to Rate 1, to see how the BP league quality factor might affect the results. But for only one player was there much of a difference – Tris Speaker. Defensive league strength did not change much from 1930 onward, according to BP. Rather than report two different measures, it may be best merely to caveat results, and mention at least in passing where different conclusions could be drawn. I had anticipated that maybe a whole bunch of guys from pre-WWII might show up high on these tables; like so many great hitters or pitchers from the deadball era dominated their peers much more than the post-integration men. But this is NOT the case – Speaker being the only CFer from pre-1925 to make this list, and 60% of the best players coming after integration; so it is less of a concern. More on this near the end.

How to take the raw stats and create measures of effectiveness?

First, I decided to only use centerfielders. As talented as Al Kaline and Roberto Clemente were, very few would seriously believe that any right or left fielder was as good as the best CFers.

I began with a set of CFers who score very well on both (WS and BP) measures. There are 36 outfielders (they were all primarily CFers) who receive an A+ grade from Win Shares. I eyeballed these, and found 19 that also had a high career rate stat from BP (taking into account career length, since playing into your late 30s almost always lowers your career rate). These 19 were, alphabetically:

> Richie Ashburn, Paul Blair, Max Carey, Dom DiMaggio, Joe DiMaggio, Jim Edmonds, Curt Flood, Marquis Grissom, Andruw Jones, Mike Kreevich, Willie Mays, Terry Moore, Hy Myers, Amos Otis, Jimmy Piersall, Kirby Puckett, Tris Speaker, Lloyd Waner, and Sam West.

As I began compiling the data, I found that three players in this group each spent a number of seasons or partial seasons in the middle of their careers in LF or RF. This makes it very challenging to measure them against the others; I decided not to try, and to discard these players from the study. In the end, it certainly seems that if we are discussing the best outfielder ever, that the player in question ought to have played almost all of his years in center. So, as good as Lloyd Waner, Hy Myers (Hy Myers? Who was he?), and especially Max Carey may have been, they were lumped into the same set as Cool Papa Bell and the 19th-century star "Fielder" Jones; off the list for lack of data.

After compiling the data, Amos Otis, Terry Moore, and Kiirrrrrbeeeee (oh, how I luvved that announcer's call) Puckett clearly had the lowest scores of this group, so they were dropped as well. That leaves 13 truly golden CFers, ranging fairly evenly (for a sample of this small size) from the 1910s (Speaker) to the current decade (Jones).

Next, I combined the two measures in a way that puts them on equal scales (apples to apples). The Fielding Win Shares data is provided in terms of "WS Per 150 Games" (where one WS equals a third of a win). BP Rate gives "Defensive Runs Saved per 100 Games." Simple math, using the commonly accepted approximation that 10 runs equal one win, can translate one system into the other. I averaged the scores of the two metrics, and come up with one Statistic that represents defensive quality for each player-season.

[1] For a definition, go to www.baseballprospectus.com, type in a name under the Player Finder box. Click on a statistic used such as Rate, and you will be taken to the Glossary.

Example: The absolute best defensive season among this group (as far as I can tell from an eyeball scan of the data) was Marquis Grissom's 1994, when he had 322 putouts in 109 games, leading the NL by 24.[2]

Grissom's BP Rate for 1994 was 114. This means he was 14 runs above average (average is defined as 100) per one hundred games played.

Grissom's FWS per 150 games that year was 10.87. I subtract out 3.5 FWS per 150 G as the baseline rate; this isn't exactly the "average" of all fielders, but this causes FWS to correspond pretty exactly with the BP Rate. Grissom is 7.37 FWS above average per 150 G, which is also 4.91 per 100 G above average. WS are denominated in "thirds of wins", and typically it takes 10 extra runs to generate 1 win, so one WS equals 3.33 runs. Therefore I multiply 4.91 by 3.33 to get 16.4 runs above average per 100 G. This converts to a rate of 116.4; 16.4 above an average rate of 100.

Combining the two measures and giving them equal weight, Grissom's 1994 year gets a score of 115.2, which I will round to 115 on all graphs and tables. The very best CF year in MLB history was worth 15 runs saved per 100 games played; about 1 run a week.

Results – Peak

Tris Speaker had 5 years better than anyone else's best 5 years. In fact, I get the same answer whether the time period is 3, 9, or a billion years. His best years are very spread out; his top one was at age 31, another great one in his first full season at age 21, and a fine year at age 38. The numbers:

1.	Tris Speaker	111.8
2.	Andruw Jones	111.5
3.	Curt Flood	110.6
4.	Dom DiMaggio	110.2
5.	Marquis Grissom	109.8
6.	Willie Mays	109.7
7.	Jim Piersall	109.6
8.	Paul Blair	109.1
9.	Sam West	107.9
10.	Mike Kreevich	107.5
11.	Richie Ashburn	107.2
12.	Jim Edmonds	107.0
13.	Joe DiMaggio	106.1

Results – Prime

Andruw Jones, from age 21 to 28 (through 2005), has been as good an outfielder as there ever was over an 8 year period. His numbers, nearly surreal early on, have begun to slip, so time will tell if he can continue this pace. Dominic DiMaggio's "prime" was achieved in his age 24-31 seasons, of which he missed ages 26-28 for WWII. None of us know whether he would have been better, or worse, or injured, had he played those years, but it seems eminently reasonable to give him credit for the surrounding seasons as if his play in his true prime was the same as his "prime" defined this way. In other words, he has a case to be called one of best fly-chasers ever, and it is not inconceivable that he was *the* best. Curt Flood was still going strong when he left the game; his "prime" was through his last season at age 31, before his unsuccessful battle against the reserve clause deprived his fans from seeing any more flashes of leather.

[2] Yes, that was the strike year, but he still played over 100 games. If I wanted to use the best full year, there are a bunch of seasons from different players that all look approximately even; and if I was up to giving tie-breaking bonus points among these for post-season play, I would declare the winner to be Willie Mays' 1954.

1.	Andruw Jones	109.6
2.	Dom DiMaggio	109.2
3.	Tris Speaker	108.3
4.	Jim Piersall	108.1
5.	Paul Blair	107.8
6.	Marquis Grissom	107.8
7.	Curt Flood	107.7
8.	Willie Mays	107.1
9.	Sam West	105.9
10.	Jim Edmonds	105.7
11.	Mike Kreevich	105.6
12.	Joe DiMaggio	105.0
13.	Richie Ashburn	105.0

I adjusted the requirements for military service. When the 8 consecutive years straddled the military service time on both ends, I dropped the required number of years by the service time length. Joe and Dominic DiMaggio missed 3 years, and Willie Mays missed 2. I also made an exception for Jimmy Piersall, who only played 91 games in the field in 1959; I credited him with a "typical" season, thus making an 8-year PRIME from 1955 to 1962 for the subject of "Fear Strikes Out".

Results – Career

Only 8 of these 13 men qualified for 12 years of at least 100 games played, even with war credit; and half of these barely made the 12th year! Beyond a dozen, Ashburn had a 13th year, Joe D a 14th (again including 3 yrs for WWII). Only Mays and Speaker went beyond there.[3]

1.	Tris Speaker	108.8
2.	Willie Mays	108.0
3.	Dom DiMaggio	107.1
4.	Curt Flood	106.3
5.	Paul Blair	105.5
6.	Marquis Grissom	104.6
7.	Joe DiMaggio	103.9
8.	Richie Ashburn	103.6

Summaries

Figure 1 (next page) shows the three metrics for each player on one chart. The players are ordered by their peak rate. Players with fewer than 12 qualifying seasons do not have a career bar. There are two cases where a player's prime is actually lower than his career; Speaker and Mays. This is because they had some great years (counted among their best 12) late in their careers, and they didn't have a superlative string of 8 consecutive seasons at any one time. Mays, for example, had his best prime at ages 20 through 27. But his second best eight-year stretch was from ages 31 to 38! This is somewhat speculative, but I wonder if Mays' lesser performance in mid-career could be due to twice switching parks: once from the Polo Grounds to Seals Stadium in 1958, and then to Candlestick in 1960. His average Rate for the four-year period 1958-1961 (the first 2 years in each stadium) was "only" 101.8, lowering his overall career mark by almost one full point.

I also compared the players by ordering all of their individual seasons from best to worst. These are shown in figure 2. Where time was missed for military service, I used the surrounding seasons to create a rating for the missing years.[4]

---

[3] Another technical note - a few players spent part of year in LF or RF. I made adjustments by subtracting a fairly large penalty (8) from the Rate stat (LF or RF converting to CF) when combining part-years and averaging in by games played. Joe DiMaggio in 1936 split time LF and CF. I used his Win Share rate overall. He had a BP Rate of 104 in CF and 108 in LF, which converts to 100. Splitting the two, he earned a 102 for the year. Sam West in 1928 was half RF (rate 113, converted to 105), half CF (rate 95). So he gets 100.

[4] Specifically, I used the six surrounding years when available. When not, I used fewer. For Piersall, who missed time prior to his prime 7, I used the average of the prime. I also decided that I would limit the lower end by the player's average for his entire career. For Joe DiMaggio, whose fielding stats were significantly worse after coming back from WWII, I used his career average.

From this figure, we see Tris Speaker's lead over everyone, regardless of how many years are used (remembering again that these are not consecutive seasons). Willie Mays almost catches him at the "end", since Mays' years from their 14th best seasons and onward were better. Jones and Flood are right next to the top early on, and Dom DiMaggio also comes close for a while as well.

Figure 1

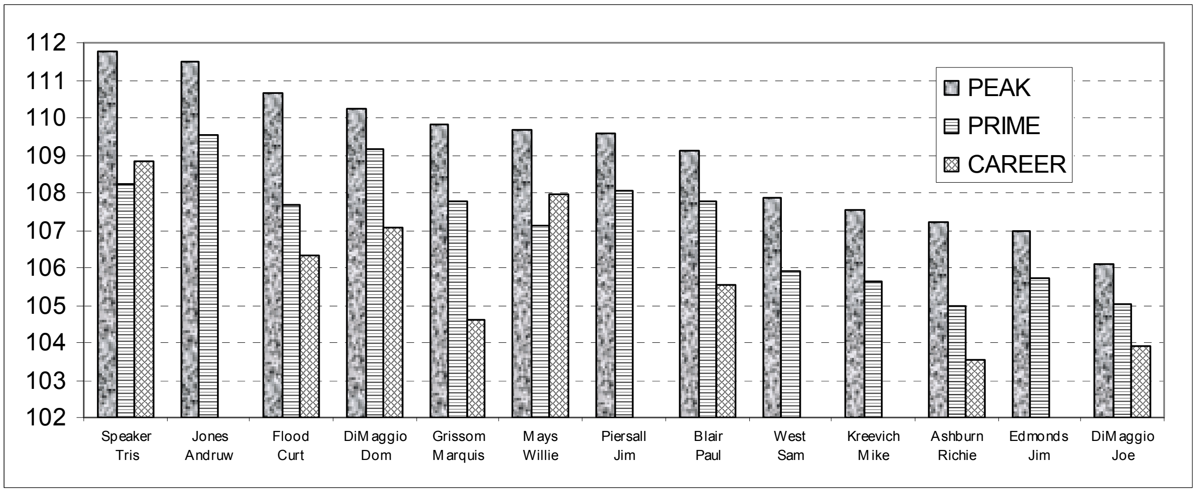
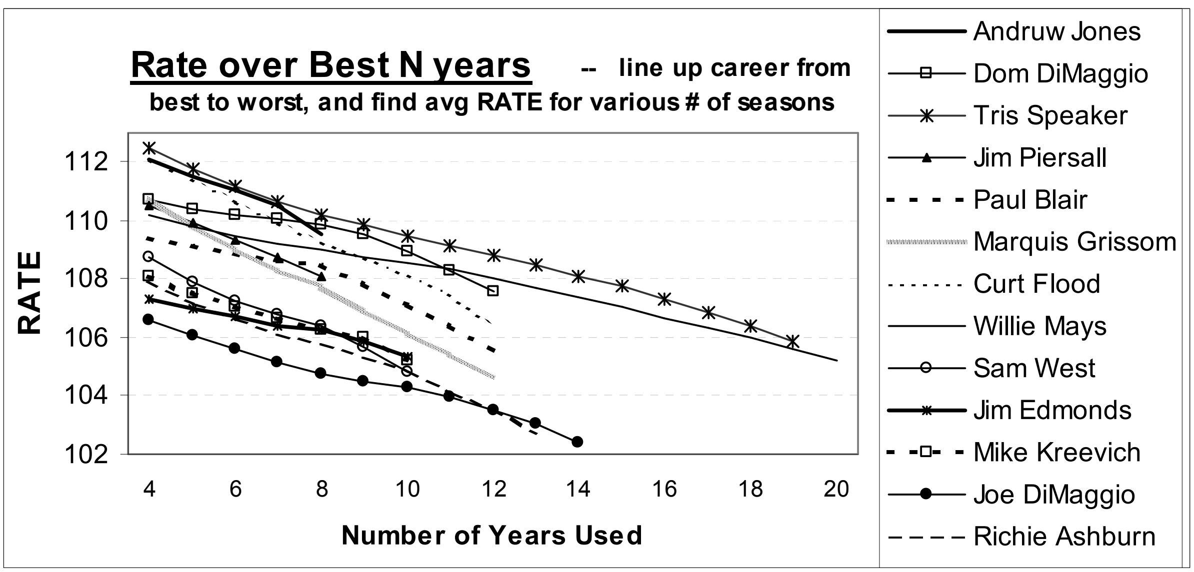


Figure 2



League quality and the deadball era

At this point, I should be moving right into the "conclusions" section, but there is an elephant in the room. Conditions change over time. Integration, expansion, and scouting in Latin America certainly have impacted the game. It is nigh to impossible to adjust for all of these. As I mentioned up front, and seen throughout the study, at least the data did not show a huge imbalance of players from one era, or lack in another. However, Tris Speaker really is an interesting case. He is the only one of the 13 who played when home runs were rare. He thus played much shallower than any outfielder in the modern era, and in fact he played unusually (some would say breathtakingly) shallow even in his day. He was tremendously talented, and it could be argued that his talents were suited particularly for the conditions in which he found himself. Certainly he would not have racked up the huge number of assists (14 years of 20 or more) and unassisted double plays in another type of game. So, one could posit that the Grey Eagle belongs on some different scale, in some other arena, than the other players here. I will not attempt to stake a claim either way.

Conclusions

Tris Speaker dominated outfield play over his contemporaries like no other man who ever wore a glove. Best Peak, Great Prime, Best Career; he has it all.

Dom DiMaggio could have, might have been as good, but those missed years are lost, and to place him among the greatest requires speculation. The old ditty "he's better than his brother Joe....Dom-in-ic DiMaggio" seems to have been true, at least when it comes to defense.

Curt Flood had a chance to be the best ever.

Andruw Jones still *has* a chance to be the best ever.

Willie Mays had a few mediocre years among his great ones, but when accounting for career length, he is still arguably the best ball hawk ever, if you believe it was more challenging to rise above the crowd in the integrated leagues than in Speaker's day. And if I had to pick one player at his very best to cover the vast green expanse for me, I may just choose a young man, cap flying off, sprinting and snagging a liner in the gap or a deep fly, whirling and throwing a strike, in the crisp autumn of 1954.

Tom Hanrahan, Han60Man@aol.com ♦

Submissions

Phil Birnbaum, Editor

Submissions to *By the Numbers* are, of course, encouraged. Articles should be concise (though not necessarily short), and pertain to statistical analysis of baseball. Letters to the Editor, original research, opinions, summaries of existing research, criticism, and reviews of other work are all welcome.

Articles should be submitted in electronic form, either by e-mail or on CD. I can read most word processor formats. If you send charts, please send them in word processor form rather than in spreadsheet. Unless you specify otherwise, I may send your work to others for comment (i.e., informal peer review).

If your submission discusses a previous BTN article, the author of that article may be asked to reply briefly in the same issue in which your letter or article appears.

I usually edit for spelling and grammar. If you can (and I understand it isn't always possible), try to format your article roughly the same way BTN does.

I will acknowledge all articles upon receipt, and will try, within a reasonable time, to let you know if your submission is accepted.

Send submissions to:
Phil Birnbaum
88 Westpointe Cres., Nepean, ON, Canada, K2G 5Y8
birnbaum@sympatico.ca