How Teams Score: A Picturebook

by Dave Studeman

August 31, 2004

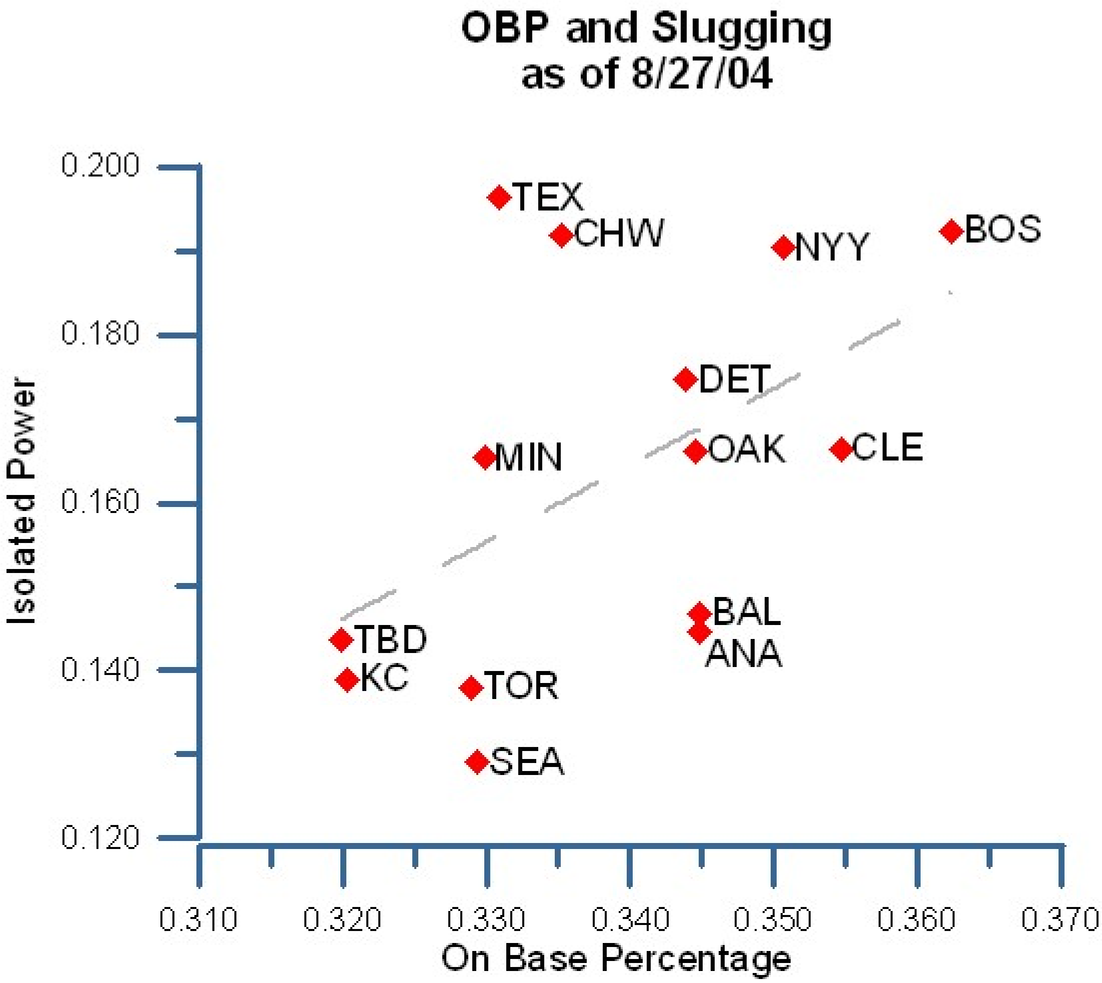
There are lots and lots of ways to score runs, but there are probably even more formulas for calculating how runs are created. **This very mathematical article** includes an analysis of 39 different run estimation formulas, and I don't think the author covered them all. Bill James, for instance, has even created at least fourteen different versions of his Runs Created formula. We use the most recent, and complicated, version **on our site**. And it seems that every baseball analyst has their own take on the best run estimator.

I find the whole debate interesting, but I'm the kind of person who better understands things through pictures rather than numbers. If you're that kind of person too, I've created several graphs that you might find interesting. If you'd rather read tables of numbers, then go nuts on the links in the first paragraph.

The first graph is our standard "runs created" graph here at THT, because it's a simple graph of On Base Percentage (OBP) and Slugging Percentage (SLG). These are the two fundamental components of Runs Created, which is why people pay so much attention to OPS, and which is why it makes sense to plot them on two axes of a graph.

Actually, I like to use Isolated Power (ISO, which is SLG minus Batting Average) instead of SLG on the graph because it better illustrates the differences between teams and players. Remember, I'm not trying to create formulas with the best results; I'm trying to create pictures that best illuminate what's happening.

Here's the American League graph. Teams to the upper right of the graph have the best offenses, vice versa for the lower left. Teams above the line are relatively better at slugging, while those below the line are relatively better at getting on base.



According to the graph, the Red Sox have the best offense in the league, and, lo and behold, they have in fact scored the most runs in the league. Now, it's important to remember that these stats aren't adjusted for ballpark factors, and the Red Sox play in a decidedly hitter-friendly park. But they are still the best offensive team in the league.

Other things of interest are the Rangers' and White Sox's relative reliance on slugging average (both factors influenced by ballparks) and the reliance on OBP by the Indians, Angels and Orioles. It's also interesting that the Angels and Orioles pretty much occupy the same space on the graph, but the Angels have scored about thirty more runs. So this graph is good, but it's not perfect. Anyway, here is a link to **the National League version of this graph**.

That's the Runs Created approach, but there are other ways to picture an offense. One approach I like is to graph three simple components, which I estimate account for over 90% of all runs scored. I'm not saying this is the best mathematical way to estimate runs or anything. I'm just saying this provides an interesting, useful take on the subject:

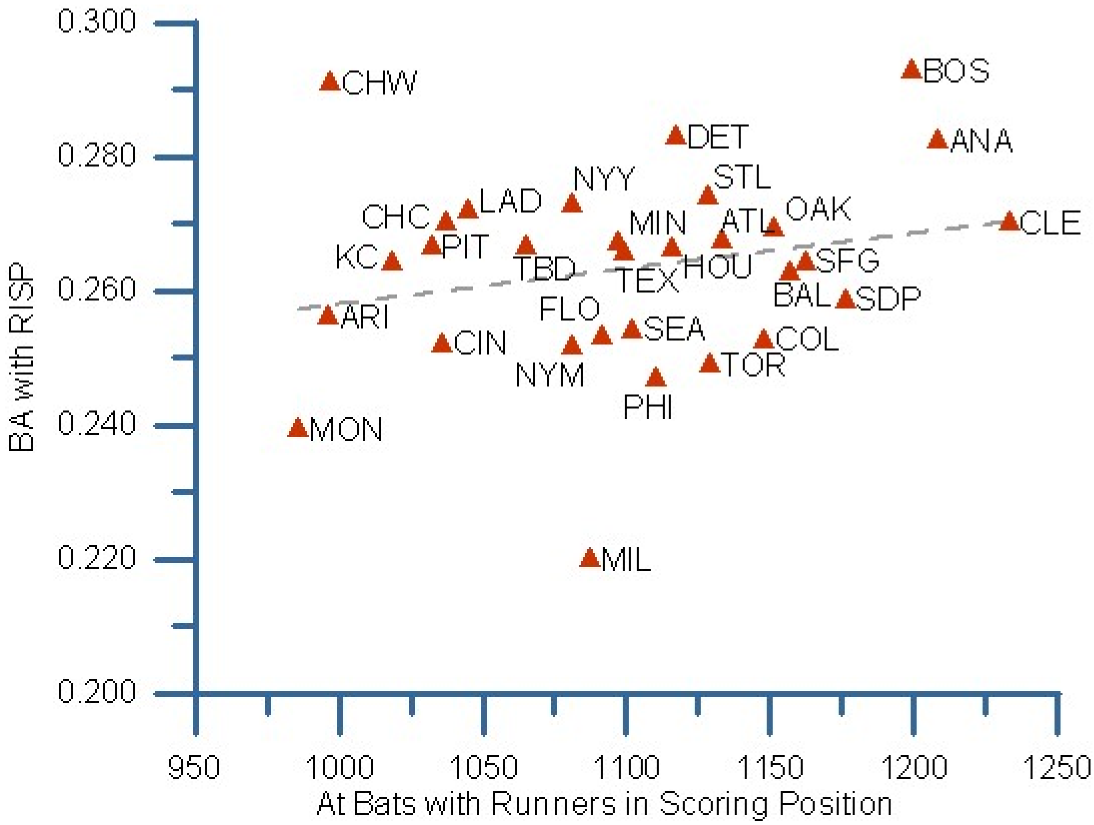
First, hit a home run. Whenever you want, as often as you can.
Second, if you don't hit a home run, get yourself into scoring position (second and third base).
Third, hit with runners in scoring position. Drive them home.

To show you what I mean, let's take a look at each of these components. First, here's a list of the number of home runs hit by all AL and NL teams:

```
American League            National League
===============            ================
Teams       HR's           Teams        HR's
NYY         191            CHC          187
CHW         188            STL          169
TEX         188            COL          167
BOS         179            PHI          165
OAK         158            LAD          160
DET         153            CIN          158
MIN         147            SFG          148
CLE         143            NYM          146
BAL         127            ATL          139
ANA         125            HOU          135
KC          117            MON          117
TBD         112            FLO          116
SEA         107            ARI          115
TOR         106            PIT          115
                           MIL          104
                           SDP          100
```

Ballparks are a significant factor in these ratings, too. **As noted earlier**, Chicago is the place to be if you want to hit home runs. Nevertheless, the Cubs are clearly the class of the NL in hitting home runs, while four teams (Yankees, White Sox, Rangers and Red Sox) are at the head of the AL class. Notice, too, that the Angels and Orioles have hit about the same number of home runs.

By the way, sorry about the table. Let's move onto the second two components of scoring runs, and let's also get back to pictures. Here's a graph of the number of times major league teams have batted with runners in scoring position (RISP), along with their batting average with RISP.



Ah, I love this graph. Again, the leading offenses tend to be on the upper right part of the graph (Red Sox, Indians) and vice versa (Expos, Diamondbacks). Also, there's a natural upward slope to the data (as drawn by the line) because teams that are good at getting into scoring position are teams that bat well.

This graph also separates the Angels from the Orioles. The Angels have scored more runs than the O's because they've gotten more runners into scoring position AND they've hit better with runners in scoring position. The White Sox are the truly odd team on this chart — they lead the league in batting with RISP, but they have the third-least total number of runners in scoring position.

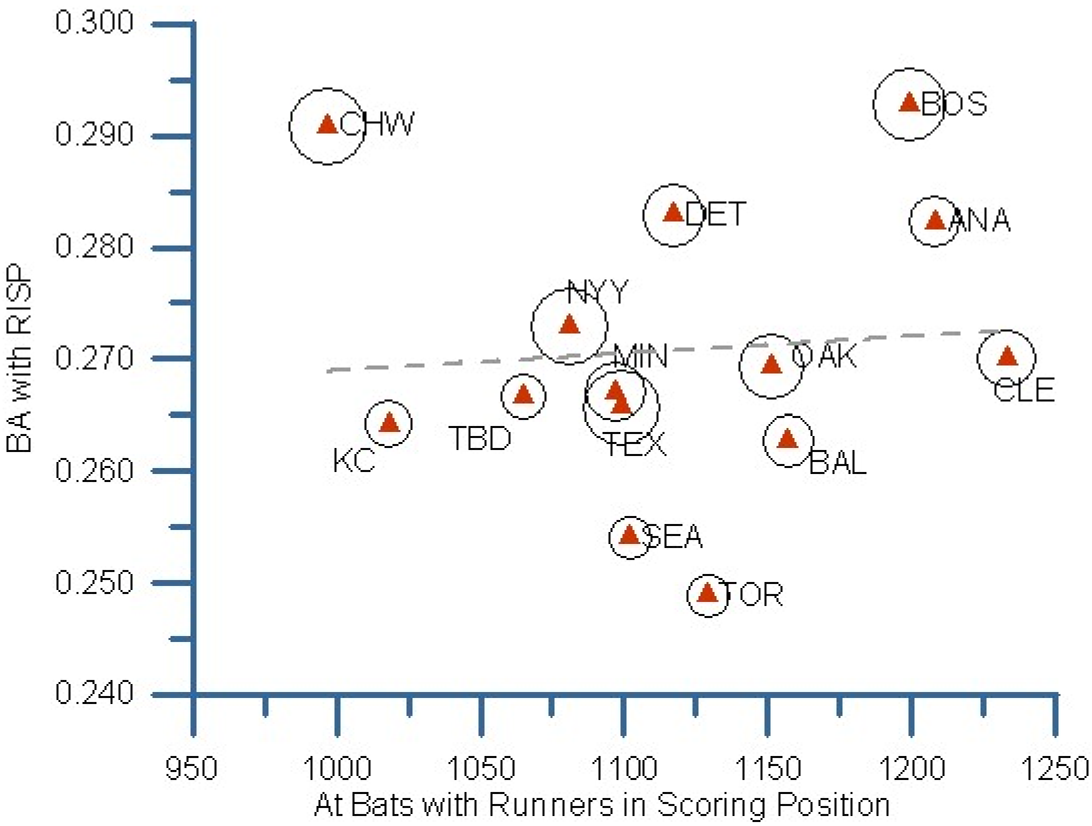
BA with RISP tends to be equal to overall BA. In the AL, teams are batting .271 overall, and .271 with RISP. But the Sox are batting .291 with RISP vs. .268 overall. In other words they've been lucky by hitting well in clutch situations, despite what you may have heard about injuries to their top hitters.

In general, runners get into scoring position one of three ways:
– 20% get there by reaching first (via a single or walk) and moving on by stealing a base or by a teammate's "productive out".
– 30% get there directly by hitting a double or triple.
– 50% get there by reaching first, and subsequently moving on via a positive

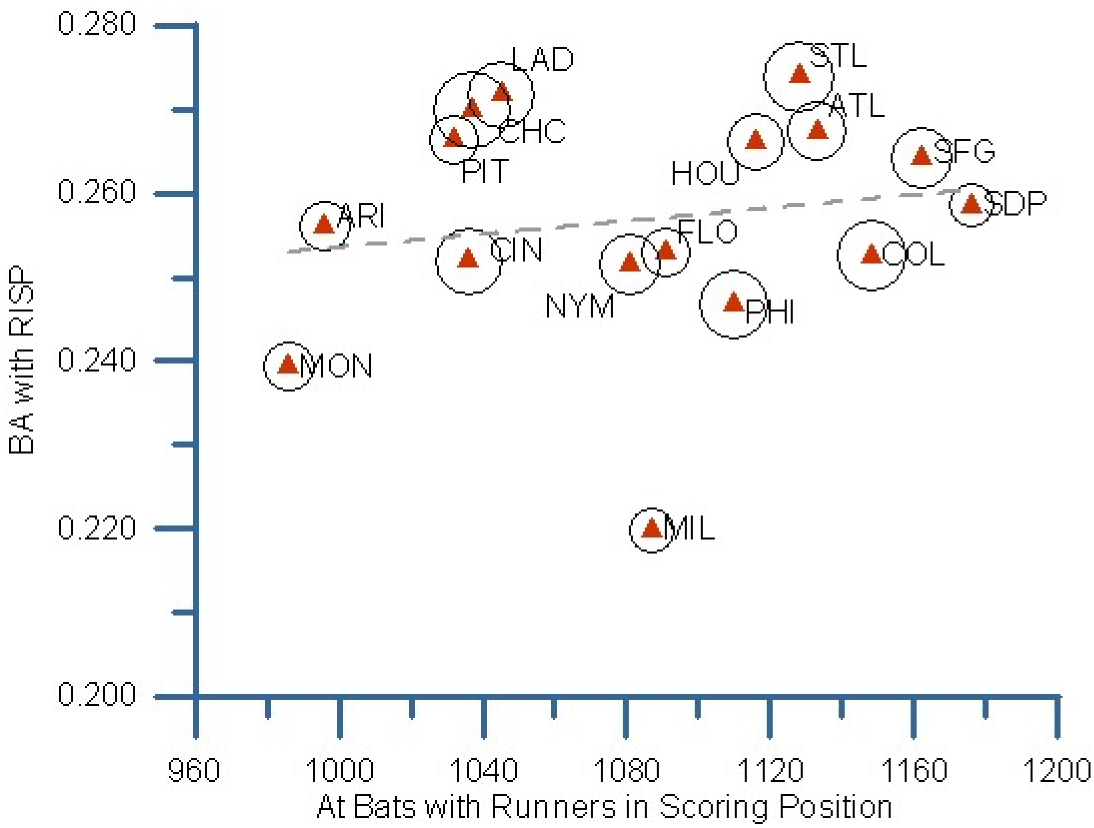
contribution (hit, walk, etc.) from a teammate.

When first hired, White Sox manager Ozzie Guillen talked a lot about getting his runners in motion, but the Sox don't reach first base enough in the first place; they are third from last in singles, walks and HBP combined. They've batted well with RISP and hit a lot of home runs, but reaching scoring position has been their problem.

Speaking of home runs, let's make this chart complete by adding home run information. I'll add a circle to each data point that is proportional to the number of home runs each team has hit. Here's the American League:



I think this paints a fairly complete picture of each team. For instance, you can see how the Yankees' relatively average position on the graph is bolstered by their home runs (circle size), which makes them the third-best offense in the league. Instead of adding additional comments, I'll let you reflect on your favorite teams. Here's the National League.



In case you were wondering how San Diego was scoring, now you know. And consider Milwaukee, the unclutchiest team in the NL. Pictures are indeed worth a thousand words. I'll mention that to my editor.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

But I Regress…

by Dave Studeman

January 4, 2007

Do you know that thing that statisticians do called **regression analysis**? It's when they look at two (or more) numbers to determine how closely correlated they are. To use a couple of examples I've seen recently, **education is correlated with health** and the presence of a Led Zeppelin bumper sticker is correlated with **the likelihood of that vehicle containing a controlled substance like marijuana**. I first learned regression analysis back in the days when you had to compute it by hand; now all you need is a computer with Excel. It's a neat tool, perhaps a bit too easy to use for some.

But something's always bugged me: why is it called regression analysis? Why isn't it called correlation analysis? I mean, when you run a regression analysis, the main output is the correlation between the variables, right? So why is it called regression? Huh? Haven't you wondered the same thing? Even once?

Okay, perhaps you're not as geeky as I am. But you'll be happy to know that I think I found the answer while reading a biography of the guy who invented regression analysis, **Sir Francis Galton**.

Galton was an amazing, quirky guy; one of those classic Victorian gentlemen with lots of time on their hands and lots of things to discover. He traveled the Nile and explored parts of Africa that hadn't been seen by white men before. He published a book on survival in the wild, parts of which are still included in survival guides. He invented some silly things (one of my favorites: the gumption-reviver machine, which simply dripped water on you until you were thoroughly soaked) and some very important things (weather maps; the system for categorizing fingerprints still used today). Most of all, he counted things.

Galton was an obsessive counter. He determined a precise formula for preparing the perfect cup of tea. He counted beautiful women in different parts of England to deduce his own "beauty map." And when his cousin, Charles Darwin, invented a little something called evolution, he threw himself into the task of counting hereditary traits.

He was convinced that things like criminal behavior, intelligence and genius were linked to heredity. His beliefs stood in contrast to many of his critics, who also cited environment. In fact, it was Galton who first turned the phrase "nature/nurture" to describe the argument. Along the way, he decided the best thing to do would be to collect statistics on people and measure them. So he set up shop in a Public Health exhibition and asked people if they would like to be measured (height, armspan, breathing capacity, eyesight, etc.). After a year, he had collected measurements on over 10,000 people.

Statistics was still in its infancy, and Galton certainly didn't have a computer back then. But he decided to analyze these numbers as best he could. He took the heights of 205 sets of adults and their children and (much to my delight) laid them out in a scatterplot graph. He saw that the points moved together: the taller the parents, the taller the children. However, the points didn't line up perfectly.

So he drew a line that seemed to best fit the relationship between the points, and measured its slope. The result was two-thirds. As Galton thought it through, he realized that children were two-thirds as likely to be as "extreme" as their parents. He called the remaining one-third "regression." Actually, he called it "regression to mediocrity," which we have modified to regression to the mean.

This was actually a blow to Galton, who wanted to believe that heredity was absolute. But it was a huge step forward for the field of statistics. Galvin went on to refine his technique, developing correlation coefficients and lots of other things. But the very first thing he noticed, the thing that the graph showed him, was regression. And that's why we call it regression analysis. I think.

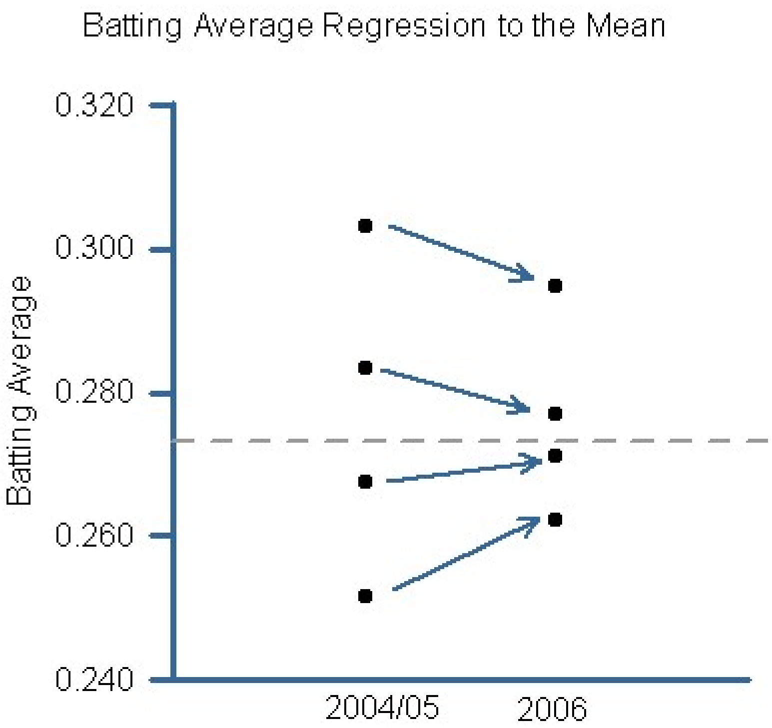
Regression to the mean is everywhere in baseball. Sophomore slump? Regression to the mean. Seattle's 93-69 record after going 116-46 in 2001? Regression to the mean. **Luke Scott's** Slugging Average in 2007? Regression to the mean.

Let me show you another graph. This graph plots batting average in 2005 and 2006. What I've done is to split up the 2005 batters into quartiles, and then plotted how those same batters performed in 2006. I used a minimum of 300 at bats in 2005 and included the player in in both years if he played in 2006 at all. This is what regression to the mean looks like:

As you can see, each one of the four quartiles moves closer to the average (that gray line) in 2006. The first quartile of batters batted .305 in 2005 and .294 in 2006. The lowest quartile batted .245 in 2005 and .263 in 2006. Each group moved closer to the mean.

There is probably some selection bias in that lower quartile. The worst batters played less in 2006, which skews the overall results higher. So regression to the mean isn't quite as strong as it appears in that lower quartile, but it's still pretty strong.

What we're really after is understanding the difference between a player's "true talent" and the overall league average. The problem is that one year isn't enough data to establish a player's true talent. So let's see what happens when we include two year's batting average (2004 and 2005) in the initial quartiles:



If you compare the two graphs, you'll see that the lines aren't as steep when you have two years' worth of data to begin with. In this case, the first quartile moved from .303 in 2004/05 to .295 in 2006, a little less than the one-year sample. The bottom quartile migrated from .252 to .262, a lot less than the one-year sample. If you have more years in your baseline, there is less regression to the mean.

Why do I bring this up now? Because lots of people are producing forecasts for the 2007 season, and one of the first things every decent projection system will do is regress a player's performance to the mean. In fact, there is one system that does nothing other than regress each player's performance to the major league average as a basis for its 2007 projection. It's called Marcel, because it's so simple that even a monkey can do it. (Marcel, from **Friends**. Get it?)

You can read more about the Marcel system from its current caretaker, Tangotiger. Tango's specific calculations are **laid out in this thread**—he essentially takes each player's previous major league performance and regresses it to the mean. That's it; no park adjustments, minor league stats or anything like that. The amount to which he regresses each player depends on how long the player has been in the majors. If he's only been in the majors a year or two, Tango regresses his performance a lot. He also regresses a pitcher's performance more strongly than a batter's, because pitchers are typically more random.

Chone/Sean Smith found that Marcel had a .66 correlation with batters' actual performance last year. The best correlation he found was PECOTA's, at .74. Nate Silver of **Baseball Prospectus** has worked tremendously hard to make PECOTA a cutting-edge system and has succeeded. But even his model only gains a smidgen of accuracy over Marcel. That is the power of simple regression to the mean.

You can **download the 2007 Marcel projections from Tango's site**. Just for the heck of it, I downloaded them and compared them to each player's 2006 performance. Here is a list of the batters who are most likely to see an increase in their batting average, based on Marcel and regression to the mean (minimum at bats in 2006: 300. Minimum batting average in 2006: .240):

Last	First	06BA	mBA	Diff
Gonzalez	Luis A.	.242	.285	.043
Cantu	Jorge	.249	.281	.032
Izturis	Cesar	.245	.276	.031
Ellis	Mark	.249	.278	.029
Mueller	Bill	.252	.279	.027
Duffy	Chris	.255	.281	.026
Kubel	Jason	.241	.266	.026
White	Rondell	.246	.271	.025
Crisp	Coco	.264	.289	.025
Casey	Sean	.272	.296	.024
Lopez	Javy	.251	.276	.024
Peralta	Jhonny	.257	.280	.024

In general, you won't see many predicted improvements for first- or second-year players, because there's not enough history to regress to. But Cleveland fans should feel good about seeing **Jhonny Peralta** on this list.

Here's a list of players whose batting averages are most likely to decline next year:

Last	First	06BA	mBA	Diff
Redmond	Mike	.341	.291	-.050
Scott	Luke	.336	.292	-.044
Bard	Josh	.333	.293	-.041
Ozuna	Pablo	.328	.290	-.038
Ward	Daryle	.308	.269	-.038
Cirillo	Jeff	.319	.281	-.038
Jones	Chipper	.324	.286	-.037
Helms	Wes	.329	.293	-.036
Coste	Chris	.328	.294	-.034
Jeter	Derek	.343	.311	-.033

You shouldn't really be surprised by any of the players on this list. Let's switch to On-Base plus Slugging Average (OPS). Here's a list of players most likely to improve next year by regressing to the mean:

Last	First	06OPS	mOPS	Diff
Clark	Tony	.643	0.826	.183
Gonzalez	Luis A.	.625	0.764	.139
Guillen	Jose	.674	0.800	.126
LaRue	Jason	.663	0.763	.101
Peralta	Jhonny	.708	0.803	.095
Lee	Derrek	.842	0.934	.092
Cantu	Jorge	.699	0.789	.090
Lopez	Javy	.683	0.767	.084
Hermida	Jeremy	.700	0.782	.082
Niekro	Lance	.673	0.754	.082
Crisp	Coco	.702	0.783	.081
Varitek	Jason	.725	0.806	.080
Navarro	Dioner	.687	0.767	.080

Here's a list of players most likely to decline:

Last	First	06OPS	mOPS	Diff
Scott	Luke	1.047	0.872	-.175
Ward	Daryle	.926	0.782	-.144
Ross	Dave	.932	0.788	-.144
Helms	Wes	.965	0.831	-.134
Dye	Jermaine	1.006	0.879	-.128

Thome	Jim	1.014	0.900	-.114
Beltran	Carlos	.982	0.875	-.107
Anderson	Marlon	.866	0.765	-.102
Bard	Josh	.926	0.826	-.100
Saenz	Olmedo	.927	0.828	-.099

Is Marcel saying that each of these players will regress to the mean? Absolutely not. Some of them won't. But enough of them will regress to the mean to validate the entire approach. Marcel doesn't predict breakout seasons; by definition, those are nearly unpredictable. It predicts what you can most likely expect from a player.

Projection systems start with regression to the mean, but they differ significantly in what they regress to. Marcel simply regresses to the overall major league average (with one exception for pitchers in the American League), while PECOTA regresses to the average of similar players (based on height, weight and other things). As another example, **this thread includes a fine discussion** of how to regress players who have only been in the majors a year or two.

Sir Francis Galton would be proud of the way baseball fans and analysts have incorporated regression to the mean in their thinking. I can also think of a few players who could use that gumption-reviver machine.

References & Resources
The biography of Galton is called ***Extreme Measures: The Dark Visions and Bright Idesa of Francis Galton*** by Martin Brookes. The New Yorker **reviewed the book** a couple of years ago.

Correlation and regression analysis were a tremendous contribution to mankind, but Galton's other legacy is the field of **eugenics**. Galton envisioned eugenics as a utopian way to build the best human species. In his conception, eugenics was relatively innocent and naive. Adolf Hitler turned eugenics into a nightmare.

I want to credit John Burnson's **2006 Graphical Pitcher** for the graphical inspiration of regression to the mean. John used it to show the extreme regression to the mean of home runs per fly balls among pitchers.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic

tweets **@dastudes**.

Comments are closed.

Ten Things About One-Run Games

by Dave Studeman

June 30, 2005

There's been a lot of discussion lately about one-run games and the **Pythagorean Formula**. I know, I know. **Boring**. But so much has been **written about the subject**, and so many **ideas have been kicked around**, that I thought it would be good to have one article that lists some basic facts regarding one-run games and the Pythagorean formula. And in compiling this list, I learned a few new things too.

Here goes…

Won/loss records are primarily driven by a team's average number of runs scored and runs allowed.

This is old news to most of you, but it bears repeating. The basic elements of the Pythagorean formula are runs scored and allowed, and it "explains" over 90% of the variances in won/loss records over the last five years. For you math types, I got an R squared of .91 when I regressed the basic Pythagorean formula against the last five years of team data. Others have gotten closer to 94%, depending on their sample and the exact formula they use.

The big question everyone wants answered, however, is how to explain that remaining 9% variance, which can be much more than 9% for some specific teams.

Run distribution patterns explain only some of the unknown variance.

I wrote an article two days ago about "run distribution patterns," showing that sometimes you can learn a bit more about a team's offense or defense by looking beyond just the average. After I wrote the article, I went back to the data and sort of combined each specific team's distribution pattern (in runs scored and runs allowed) to see how much of the unknown variance they could explain.

The answer was 1.5 percent. By combining the two, I increased the R squared of my projection to 92.5. Put another way, isolated run distribution patterns explained a little less than 20% of the variance not covered by the Pythagorean formula. That's good, but it's not enough.

The rest of the variance is explained by the outcomes of close games.

In a broad mathematical sense, there are two things that affect the unexplained variance:

- Specific run distribution patterns, as discussed above, and
- the timing of those patterns in specific games.

Here's an example of the second point. Let's say your team scores four runs the "average" number of times and allows five runs the "average" number of times. But let's further say that all the games in which they score four runs happen to be the same games in which they give up five runs. They'll lose every one of those games by one run, of course, despite following the average distribution.

You can capture this second point by analyzing the outcomes of close games, because when run patterns match up in odd ways they usually produce close games. In fact, when I add a team's record in one-, two- and three-run games to my formula, I explain nearly all of the team-specific variances from the Pythagorean formula, for an R squared of .99.

Here's a list of how much each factor explains the 9% of unexplained Pythagorean variances:

```
Record in one-run games:          4.0%
Record in two-run games:          1.5%
Overall runs/game distribution:   1.5%
Record in three-run games:        1.0%
Other stuff:                      1.0%
```

Now you know why everyone talks about one-run games. But it's also important to remember that one-run games account for less than half of Pythagorean variances. In **our team stats**, for instance, we track teams' records in both one-run and two-run games, in the category called "Close." We think it adds a bit more insight to the stats.

The 2003 Detroit Tigers, one of the worst teams in history, won over 50% of their one-run games.

A lot of people think that the outcome of one-run games is pretty much random. As an example, the 2003 Detroit Tigers, one of the worst teams of all time, actually won over 50% of their one-run games. Here's a breakdown of the Tigers' record, by margin of the game:

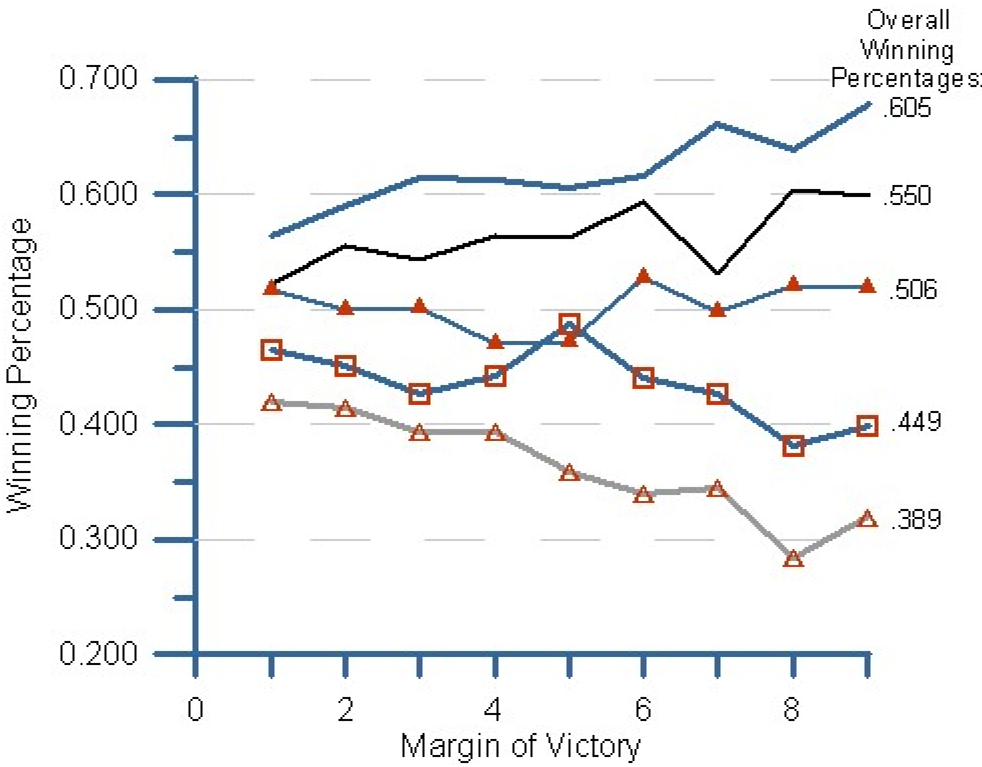
Margin	Win%
1	.514
2	.189
3	.208
4	.294
5	.286
6	.000
7	.250
8	.000
>9	.125

As they say, one of these things is not like the other. For the record, **much more sophisticated study of this** in 1997 and reached the same result.

It's this kind of analysis that leads some people to make the "strong form" of the argument: that the outcomes of one-run games are completely random. But that's not really true.

Good teams win more one-run games.

Here's a graph of the winning percentage of all teams in each of the past five years, according to the margin of victory in each game. I've combined seasonal teams into five different groups based on their overall record.



As you can see, a team's true talent emerges as the margin of a game increases. One-run games do tend to bring all teams closer to .500, but the best teams still win one-run games more often than other teams.

Bill James published **an article three years ago** in which he reviewed Tom Ruane's article, and added the useful insight that a team's record in one-run games can be projected by the ratio of its runs scored to runs allowed, each raised to the power of .865. In other words, he used the Pythagorean formula, but used .865 instead of 2 as the exponent.

So, in essence, the Pythagorean formula actually captures the notion that good teams generally win more one-run games. But it obviously won't capture unexpected swings in one-run game outcomes. And as we've said, wild swings do occur.

You might say that some teams seem to have a particular talent for winning slightly more one-run games.

This is what James covers in the article linked above. In a nutshell, James found that there is some evidence that some teams display an ability to perform better or worse in one-run games independent of their overall talent level. The teams that show this ability have two fundamental traits: they play small ball (sacrifice hits, stolen bases, fewer home runs, etc.) and have good pitching. However, you have to apply this finding VERY carefully. For example, here are the teams that most outperformed their expected record in one-run games from 2000 to 2004:

Team	Actual	Proj.	Diff
LAN	.586	.515	.070
CIN	.534	.474	.060
MIN	.559	.503	.056

You can certainly argue that the Dodgers and Twins had many of the attributes Bill listed, but the Reds? Sometimes, things just happen.

By the way, here's a list of the three relatively worst teams:

Team	Actual	Proj.	Diff
KCA	.395	.471	-.076
BOS	.474	.532	-.058
HOU	.470	.521	-.051

Yes, the 2003 Tigers won proportionately more of their one-run games than last year's World Champions.

Bullpens may have an impact, too.

Seeing the Dodgers on the top of that list, and the Twins third, infers the point many others have suggested—that teams with strong bullpens may win more one-run games. As an example, **here's a list** of which teams have the best one-run records over the last three years, compared to their performance in saves and holds.

The logic is unescapable. If your team is leading by one run entering the ninth inning, and your bullpen (think pre-injury Eric Gagne) can shut down the opposition, you will win that game. It could be that the White Sox's bullpen has been more responsible for its record in one-run games this year than its offensive strategy.

As I say, the logic is unescapable. But I'm not totally convinced. By definition, teams that win more one-run games will have more saves and holds. And there are many teams with great bullpens who have had relatively poor one-run records.

Perhaps the only analysis that can truly answer this question is a complete **Win Probability** analysis of all one-run games. Let me know if you've got some free time.

A single run scored in an inning may be more valuable than one of the runs in a two-run or three-run inning.

There may be a natural conclusion to all of this: that a single run scored in an inning is worth slightly more than a single run scored as part of a two-run or three-run inning. **As I said earlier this week**, the first two to five runs scored in a game are worth more than the runs that pad your total after about seven. Therefore, it may be true that single-inning runs may be more valuable because they tend to occur in games with lower scores.

So if you want to make that argument, I won't disagree.

But if you play for one run, you will only score one run.

Two runs are still worth more than one run. And, as Earl Weaver famously said, *if you play for one run, you will only score one run*. If you sacrifice a runner to second base, you are still giving up an out. I don't think any of the "things" in this article suggest that there should be a much greater use of one-run strategies.

But I'm not totally against sacrifice bunts either, for many of the **reasons outlined in this article**. And maybe our list of one-run things can help you better identify situations in which one-run strategies are appropriate.

There will always be a lot of fog.

In ***Underestimating the Fog*** (link is to a PDF document), Bill James listed eight "strong form" conclusions that may, in fact, not be supported by the analysis. One of them was:

> "Winning or losing close games is luck. Teams which win more one-run games than they should one year have little tendency to do so the next year.

As we've already seen, James believes that "winning or losing close games is probably not *all* luck." I mention this because the "fog debate" continues, and **you can read it in this article**. It's an excellent discussion for those of you who enjoy that kind of thing. Thanks to BTF for posting it.

By the way, if you made it through this entire article without yawning, then you're ready for the **ultimate test**.

References & Resources

You can improve the Pythagorean formula marginally by taking a different approach to the exponent in the formula. **Here is Baseball Prospectus's approach**, here's

another good article by the follks at Prospectus, and **here is an excellent article by US Patriot** that uses his formula for an exponent: ((RS/G + RS/G) ^ 0.28). Finally, **here's a chart by Tangotiger** comparing all the different approaches.

An article in **this version of *Baseball By the Numbers*** includes a home/road adjustment for the Pythagorean formula.

Tangotiger has written a program that calculates the expected distribution of runs scored and allowed per game, and then calculates an expected winning percentage. You can **download it from his site (at the bottom of the page)**.

All stats were courtesy of **Retrosheet**, the best thing on the Internet. Retrosheet's David Smith has also written two articles about this subject (PDF files):
Do Good Teams Really Win More Close Games?
Patterns of Scoring and Relation to Winning

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Of runs and wins

by Dave Studeman

December 14, 2012

The Orioles caused a bit of consternation on the Interwebs this season. It's been an article of faith from the Beginning of Sabermetrics (approximately 1980, or 1 A.J. (After James)) that teams tend to win relative to how well they outscore their opponents. If they score a lot more runs than they allow, they win a lot of games. If they don't, they don't.

The Orioles, however, won 93 games and lost only 69 last year, despite scoring only seven more runs than they allowed. This wasn't a record for outperforming a team's run differential, but it was close. As the season progressed, saberists kept insisting it couldn't continue. Yet it did, right up until the end of the season. Anti-saberists seemed to enjoy a certain amount of schadenfreude when the O's made us sabermetric types look silly. O's fans, of course, were delighted regardless.

So what are we to think? Is this Article of Sabermetric Faith wrong? Are we fooling ourselves when we pay too much attention to runs scored and allowed, and not enough to basic wins and losses?

To answer these questions, let's go back to the basics. What's more, let's go back to the data.

I decided to compare two consecutive months of a team's win/loss record to each other, within a specific season. By comparing in-season months, I mostly avoided the hassle of personnel turnover and that sort of thing. To make sure I had enough data, I collected all teams and months from 1970 through 2012, a total of 5,747 team/month/year combinations.

Next, I grouped the teams into 21 buckets, based on how they performed in the first baseline month. (A winning percentage of .000 to .050 was bucket 0, .050 to .100

was bucket 1, and so on. Bucket 19 included winning percentages between .950 and 1.000; bucket 20 was a winning percentage of exactly 1.000.) Some of these months consisted of only one or two games, so I weighted each team/month comparison by the lesser number of games in the two months (if there were 25 games in the baseline month but only one in the next month, the stats were prorated as if there was only one game in both months). That way, I didn't have to exclude any months by an arbitrary baseline, and I also didn't have to worry about those pesky strike years.

For the following analysis, I included only groups with a baseline winning percentage (the percentage in the first month) higher than .200 and lower than .800—12 groups representing 6,546 team/month/year combinations in all.

You may not have followed all that. It's okay. You'll get it when we dive in.

Regression toward the mean

Before we get to the runs scored and allowed thing, we have to do something else. We have to regress toward the mean. **Everything regresses to the mean**, as someone once said. The Mariners win 93 games after winning 116. Norm Cash hits .243 after hitting .361. I write a lousy column after writing a great one. Sometimes.

The basic rule is this: Before you compare two things to each other, make sure you regress them toward the mean first.

In the table below, you'll see that all 12 groups regressed toward .500 in the second month of comparison. Teams in our lowest group, for instance, had a composite .224 winning percentage in the baseline month. In the next month, their winning percentage jumped up to .422. At the other end of the standings, teams that averaged .762 in the baseline month had a .545 winning percentage in the next month. There remained a difference between the groups, but it wasn't nearly as extreme.

That is as stark an example of regression toward the mean as you're likely to see today. Here are the data for all 12 groups.

From...	To...	Number	Avg. Win%	NextWin%
.200	.250	49	.224	.422
.250	.300	191	.277	.461
.300	.350	363	.329	.446
.350	.400	610	.376	.468

From…	To…	Number	Avg. Win%	NextWin%
.400	.450	1041	.426	.481
.450	.500	835	.472	.496
.500	.550	1202	.519	.506
.550	.600	931	.571	.516
.600	.650	675	.621	.531
.650	.700	374	.670	.543
.700	.750	195	.719	.557
.750	.800	80	.762	.545

Teams in every category—above-average teams and below-average teams—moved closer to .500 in the second month. Much closer. In fact, these findings lead us to a useful rule of thumb:

If the only thing you know about a team is its winning percentage in a single month and you want to predict how it will perform in the next month, add these two things: {exp:list_maker}25 percent of its winning percentage in the first month, and .375 (which is 75 percent of a .500 record). {/exp:list_maker}In made-up technical English, a team will regress 75 percent toward average in its second month.

Now let's talk about that runs scored and allowed thing.

Pythagorean variance

Here's what I did next. I took the number of runs scored and allowed per game for each team in its baseline month and used those data to calculate its **pythagorean record**. The pythagorean record, which is an estimate of a team's record based on its runs scored and allowed, was developed by Bill James around 1 A.J. The basic formula is RS^2/(RS^2+RA^2). RS means Runs Scored and RA means Runs Allowed. That little hat means "raised by," or squaring the number. You may remember a similar formula from your geometry class.

I varied the "squared" part of the formula for each grouping, based on the **the run environment of each team**. It adds a little more precision to the mess.

So how do we factor the pythagrorean record into our regression? Well, first I created a second group (a subgroup, if you will) of teams, based on their pythagorean record in the baseline month. For instance, if a team's pythagorean record was better than its actual won/loss record by two games (this would be an "unlucky" team in standard sabermetric parlance), it was were placed in group 2. If their won/loss record was two games better than their pythagorean record (a "lucky" team), there were placed in group -2 (negative two). The number of the group represents the pythagorean difference from reality in games won.

Once again, looking at the data output may help you understand what I did. In the table below, I regressed each group's winning percentage toward the mean (using the 75 percent rule) to predict how it would perform in the second month. Then I broke them into pythagorean subgroups to see how each type of "lucky" team performed relative to its regressed projection.

Bottom line: The higher a group's pythagorean difference, the more they outperformed their projected record in the second month. The results are dramatic.

Pyth Diff	Projection Diff
-7	-.125
-6	-.064
-5	-.057
-4	-.078
-3	-.024
-2	-.008
-1	-.003
0	.012
1	.003
2	.018
3	.015
4	.020
5	.046
6	.235

Picking on one example, teams that were three games better in their pythagorean "runs record" than their actual record beat their regressed projection in the second month by .02 percentage points. Their runs scored and allowed in the baseline month made an impact on the outcome of the second month.

What I'm saying is that baseball analysts are still right. Runs scored and allowed still matter. In fact, if you know nothing about a team except how it performed in one

month, add these two things:
{exp:list_maker}30 percent of its pythagorean record in the first month, and .350 (which is 70 percent of a .500 record). {/exp:list_maker}If you know these two things, **knowing the team's actual won/loss record won't help you one bit**. You'll do a better job of predicting a team's future if you ignore its actual won/loss record and just use its "runs record."

When James first wrote about these things, he developed **a number of sabermetric forces**. One was called the Plexiglass Principle, which today we call regression toward the mean. The other was called the Johnson Effect, which is what he called it when teams that had extreme pythagorean variances in one year tended to relapse in the next year.*

**This shows how far we've fallen as sabermetric writers. No one coins terms than James does.*

The problem these days is that people tend to throw the two forces together. When people say that teams tend to "fall back" toward their pythagorean record, they're really combining the two ideas. Pythagorean variance has become sort of a lazy man's regression term. Today, I've tried to separate the two more distinctly for you.

Still, there are more questions. There are always more questions, aren't there? What if we have multiple months of a team's record? Is there a point at which its actual won/loss record is more important than its runs record? Is at two months? Three? Four? Are things different these days than how they used to be? Has bullpen usage changed things at all? Do season-to-season effects still hold?

I'll be back.

Bonus Table

Wow. I'm impressed you're still here. As a bonus, I'm going to break out both types of groups in the following table. The top row lists the Winning Percentage groups (from a .200 level to an .800 level, by .050) and the left column list the Pythagorean Difference subgroup. The data in the table is the difference between each group's second-month projection, based on simple regression to the mean, and its actual record in the second month.

Observe and enjoy.

Pyth Diff	4	5	6	7	8	9	10	11	12	13	14	15	Total
-7											-.125		-.125
-6					-.104						-.034	-.053	-.064
-5					-.046	-.003	-.062	-.036	-.081	-.015	.005	-.220	-.057
-4				-.316	-.162	-.008	-.024	-.068	-.054	.005	-.015	-.058	-.078
-3			.042	-.088	-.028	-.049	-.036	-.026	-.014	-.014	.007	-.030	-.024
-2		.056	-.063	.005	-.022	-.028	-.008	-.015	-.009	-.010	-.013	.019	-.008
-1	-.017	-.010	-.002	-.011	-.016	-.001	.000	-.004	.008	.010	.013	-.001	-.003
	-.023	.006	-.013	-.002	.007	.008	.009	.023	.013	.013	.036	.072	.012
1	-.025	.017	-.003	.011	.007	.017	.017	.013	.023	-.028	-.016		.003
2	.030	.020	-.016	-.016	.031	.031	.002	.009	.006	.084			.018
3	-.009	.025	-.014	.051	.008	.069	.034	-.005	-.028				.015
4	-.051	.087	-.010	.068	.005								.020
5	.056			.036									.046
6		.235											.235
Total	-.006	.055	-.010	-.026	-.029	.004	-.007	-.012	-.015	.006	-.016	-.039	-.009

Feel free to ask questions, point out faulty logic and generally make fun of my math in the comments below.

References & Resources

Here's **a very mathematical examination of why the Pythagorean Formula works**.

All data courtesy of **the spectacular folks at Retrosheet**.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Why wOBA Works

by Dave Studeman

April 17, 2014

Calculating wOBA for players like Mike Trout or fictional cohort Al Trout is very simple (via Bryan Horowitz).

Johnson's Scale

In the *1985 Bill James Baseball Abstract*, Bill included an article from a guy named **Paul Johnson** who had developed his own version of Runs Created. From a sabermetric perspective, this was an important event for several reasons, but for me it all came down to the math.

Johnson called his formula Estimated Runs Produced and it's a simple construct, really: positive batting events minus negative batting events (also known as outs). The trick is in the weighting. The formula goes like this:

> “(Two times (total bases plus walks plus HBP) plus hits plus stolen bases… minus (.605 times (at-bats plus caught stealing plus GIDP minus hits))) times 0.16

I know the first part of the formula looks complicated, but it really isn't. As Johnson explained in the article, he had found that batting events follow a natural scale relative to each other in terms of their impact on run scoring. He just played around with math to find a way to replicate that scale in a formula.

The scale starts at 9 (the last single digit) and touches down on every odd number, throwing in "2″ right before the end. In other words, it's 9, 7, 5, 3, 2, 1. Each number stands for the relative weight of a different type of batting event.

Specifically…

- Home run: 9
- Triple: 7
- Double: 5
- Single: 3
- Walk: 2
- Stolen base: 1

Johnson's scale had a big impact on me and I have never forgotten it. If you remember Johnson's scale, you'll remember that a home run is worth about three times more than a single, or that a walk and a stolen base are worth about as much as a single. What's more, Johnson's scale is simple and easy to communicate. I've used it many times to explain how offense really works, including **this article**.

Most important, it's true.

Linear Weights

Another reason Johnson's article represented an important sabermetric landmark was that it directly addressed an argument that dominated sabermetrics for over two decades: what is the best way to estimate the number of runs a batter has contributed to his team? The argument raged with a white-hot intensity for a long time, but a winner eventually emerged. We call it linear weights.

I don't want to go into the entire history or rationale for linear weights (**here is a good history**), but you should know that it is behind most of our advanced stats today. **wOBA** is based on linear weights, as are **wRC** and **wRC+**. **RE24** and **WPA** are based on the same logic as linear weights. **UZR** and **DRS** are essentially linear weights applied to fielding. **WAR** and WARP are based on linear weights, too.

So I'm going to multiply each number on Johnson's scale by 0.16 (as in Johnson's formula) and then compare the results to the linear weights that Tom Tango and friends published on page 26 of *The Book* (and which were based on the years 1999 to 2002). As you'll see, there is virtually no difference between the two.

Johnson's Scale Vs. Linear Weights			
Event	*Scale*	*Times 0.16*	*Linear Weights*
Home Run	9	1.44	1.40
Triple	7	1.12	1.07
Double	5	0.8	0.78
Single	3	0.48	0.48
Unintentional Walk	2	0.32	0.32
Stolen Base	1	0.16	0.18

Paul Johnson's scale mimics linear weights almost perfectly (don't forget that linear weights change somewhat from year to year). Next time you need to remember the relative value of batting events, just remember Johnson's scale. Leave the linear weights to the spreadsheet.

I like to remind people of Johnson's scale because I've noticed that some folks are getting confused about linear weights and the relative value of baseball events. One reason they're getting confused is due to an invention that Tango introduced two pages later in *The Book*.

But first, we need to talk about the value of an out.

The Value of an Out

Go back and look at Johnson's formula to see how he valued an out. He multiplied each out by .605, which means that an out equals -0.6 (that's negative 0.6) on his 1-9 scale. Allow me to draw the comparison here in a way that mimics the table I just showed you. Following is the value of an out in…

The Value of an Out	
Source	*Value*
Johnson's scale	-0.605
Multiplied by 0.16	-0.097
In Tango's Linear Weights table	-0.299

In this case, there is a big difference between Johnson's scale and Tango's linear weights. In linear weights, the negative impact of an out is three times greater than in Johnson's formula. You may wonder why this is.

Johnson wanted a metric that followed the same scale as a team's total runs scored. In linear weights, however, everything is based on average. Generally speaking, if you apply Johnson's formula to a league's stats, you'll get the total number of runs scored by the league. If you apply linear weights to a league's stats, you'll get zero.

There's another way to explain the difference. There are essentially three types of negative impacts of making an out:

1. Removing a runner who is on base, which occurs during a caught stealing or double play.
2. Decreasing the value of a runner on base, because he now has fewer outs in which to score during the rest of the inning.
3. Reducing the potential number of runs a team can score in a game, by reducing the number of outs left in the game.

The third aspect—the "ticking clock" part of making an out—is calculated by simply dividing runs by outs. For instance, in the major leagues last year, there were 20,255 runs scored in 43,653.1 innings pitched, or 130,960 outs. Divide 20,255 by 130,960 and you get 0.154. Make it negative and you have the "ticking clock" value of an out.

This is ignored in Johnson's formula. He just includes the impact of the first two aspects of making an out. This isn't something he did on purpose—I didn't understand it myself until Tango talked about the value of outs **in this seminal post**. But breaking apart the value of the out is a key to understanding different run estimation formulas like Estimated Runs Produced and wOBA.

wOBA

To create wOBA (which is basically a linear weights rate stat), Tom did something very clever. He ignored batter outs and instead added the positive value of an out (or the negative of the negative value) to each positive batting event. For example, he added 0.299 (the positive value of an out) to 0.48 (the value of a single) to get a new value for a single: 0.779. Let's round up to 0.78.

Here is a table of the linear weight and wOBA weight of each batting event:
(technical footnotes: I am not including the impact of the wOBA multiplier, which Tango uses to make wOBA follow the same scale as OBP. It's not really necessary for the discussion. Also, FanGraphs' implementation of wOBA doesn't include stolen bases.)

Batting Event Weights		
Event	*Linear*	*wOBA*
Home run	1.40	1.70
Triple	1.07	1.37
Double	0.78	1.08
Single	0.48	0.78
Unintentional walk	0.32	0.62
Stolen Base	0.18	0.48

As you can see, each wOBA weight is exactly 0.3 runs more than its linear weights value—0.3 being the positive value of the out. To calculate wOBA, simply multiply each batting event by its wOBA multiplier, add them up and divide the total by plate appearances. Voila, the perfect rate stat.

Why wOBA Works

Let's say you have a player, let's call him Al Trout, who has hit six singles in 10 plate appearances (making an out every other time) and the league, on average, hits three singles every 10 plate appearances (again, making an out every other time). To use linear weights to figure out how many more runs Al contributed above the league average, you'd…

1. Calculate the extra runs Al contributed by hitting more singles, which equals the difference in singles times the run value of a single, or (6-3) times 0.48,

or 1.44. Then you'd…

2. Calculate the extra runs that Al contributed by making fewer outs, which equals the difference in outs made times the run value of an out, or (4-7) times -0.30 (that's a negative 0.30), or 0.9. Then you'd…
3. Add the two together. 1.44 plus 0.9 equals 2.34. In his 10 plate appearances, Al contributed 2.34 runs more than the average player.

Okay, I have to add a technical footnote here. I have shown you this way of calculating linear weights because it will make it easier to understand the wOBA formula But, in reality, you only have to multiply Al's singles and outs by the appropriate linear weights of that league and year to calculate the number of runs he contributed above average.

Anyway, that's the hard linear weights way to calculate the difference. Here's the wOBA way:

1. Multiply the difference in singles times the wOBA multiplier, or (6-3) times 0.78, or 2.34.

That's it; one simple step. wOBA shows that Al contributed 2.34 runs more than the average player, the same outcome as the linear weights.

Why does this happen? Because wOBA weights include the impact of the hit AND the impact of turning an out into a hit. When you keep the number of plate appearances even, you're not just adding a hit to the hit total. You're also reducing an out from the out total. wOBA captures the impact of both event changes.

wOBA fundamentally works because it is a rate stat. Its divisor is plate appearances. When you compare two players' wOBA you have equalized their total plate appearances.

Incremental vs. Changed Baseball Events

I hear and see this type of discussion all the time: what's the impact of giving up a walk? Well, according to our linear weights table, it's 0.32 runs, but according to our wOBA weights table, it's 0.62 runs. So which is it?

The answer is: Can you repeat the question? Because it really depends on what you're asking. If you're talking about adding a walk to a batter's line—and also adding a plate appearance to account for the walk—then you're adding 0.32 runs. However, if you're talking about adding a walk and keeping total plate appearances the same, that means that you're adding a walk AND subtracting an out. The difference is 0.62 runs.

One kind of event is incremental; it's added to the total. The other kind of event is a changed event; the number of plate appearances doesn't change. To add an event you have to reduce the opposite kind of event.

Next time you get caught in one of these discussions, keep the distinction in mind.

Relative Value

So, then is a home run three times more valuable than a single (Johnson's scale and linear weights) or 2.2 times more valuable (wOBA weights)? There's no equivocation here: It is three times more valuable. There is only one right way to ask and answer the question.

The wOBA weights don't really speak to the relative value of baseball events; they speak to the number of plate appearances needed to make the tradeoff between the events and an out. In other words, "you have to convert 2.2 plate appearances from an out to a single to have the same impact as converting one plate appearance from an out to a home run." Not many people think or speak in those terms.

So be very careful anytime you use wOBA weights as part of your thinking. In fact, just stay away from wOBA weights and stick to the Johnson scale or the actual linear weights. You'll be less confused.

Converting wOBA to Total Runs

I have a bias. I like run impact scales that add up to team and league totals. I like player run impact totals that are similar to their Runs Scored and/or Runs Batted In totals. I like being able to say that **Paul Goldschmidt** created 128 runs last year; not that he created 50 runs above average.

It's just a thing of mine.

Thankfully, with Tango's help, FanGraphs carries a number like that. It's called wRC and it's a simple derivative of wOBA. It works by taking the league average runs scored per plate appearance, adds in the player's relative performance (according to wOBA) and then multiplies the total by the player's plate appearances. It's kind of cool, really.

The exact formula is…

> (((wOBA – lgwOBA) / wOBAScale) + (lgR/PA)) * PA

wOBAScale is what I call the wOBA multiplier. You need it for the math, but it doesn't impact the concepts we're discussing here.

By adding back in runs per plate appearance, FanGraphs is adding some of the negative elements of an out, but not the "ticking clock" value of the out, just as Johnson did in his formula. Last year, major league teams scored 20,255 runs in 184,873 plate appearances. That's 0.11 runs per plate appearance (for the value of an out, we'd make it negative), which is pretty much the same multiplier that Johnson used in his formula. If you don't believe me, go back to the top of the page to take a look.

And now you know a quick-and-dirty way to calculate the run impact of an average out. Just add together runs scored per out (the "ticking clock" portion) and the runs scored per plate appearance (the impact on baserunners).

> R/PA plus R/O

The formula says that the average negative run impact of an out last year was -0.154 plus -0.11 = -0.264. **Tango's Markov Calculator** returns a value of -0.258. Pretty close and much easier, right?

wOBA replacement level

Maybe you have your own bias. Maybe you're okay with runs against average, or maybe you're a fan of replacement level. Sadly, there is no replacement level version of wOBA, but I'm going to show you how to make your own.

First, pick a target replacement level winning percentage—I'm going to pick .300—and then use the **Pythagorean formula** to figure out a corresponding percentage decrease in runs scored. For a .300 winning percentage, I found that a replacement level offense would be 65 percent of the league average (assuming that defense is average).

In that case, the replacement level wOBA is this:

> League wOBA minus (0.35 times league Runs per plate appearance times the wOBA multiplier)

The 0.35 is the result of subtracting 65 percent from one.

To apply this level to a player, take the wRC formula and replace the league wOBA with the replacement-level wOBA. The formula looks like this:

> (((wOBA – ReplwOBA) / wOBAScale) + (lgR/PA)) * PA

To help you along, **here is a list of all wOBA weights by year**—the first column is the wOBA multiplier. Remember, ignore all the other weights because they'll just confuse the issue. Stick with Johnson's scale.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Ten Things About Momentum in the Postseason

by Dave Studeman

September 29, 2005

In 1969 the Mets routed the Orioles in the World Series, 4-1. The Mets had gone 100-62 that year, for a winning percentage of .617. Pretty darn good, but not as good as the great Orioles' record of 109-53, .673.

You might say, however, that the Mets had momentum in their favor. In the month of September, the Mets' winning percentage was an incredible .767 (ask the Cubs), .137 points higher than the Orioles'. In fact, that September record was the best of any team that has won the World Series in the last 35 years. It would seem that the Mets' momentum meant more in the Series than the Orioles' consistent excellence.

I've been reading a lot lately about Wild Card teams winning the last three World Series. It's a significant trend, to be sure, and disturbing in some ways. Some say that the Wild Card format gives hot teams a chance to sneak into the playoffs based on late-season momentum, which then gives them an advantage against teams that have been consistently excellent throughout the season but not particularly hot at the end. That feels like an unfair advantage, doesn't it?

Exhibit Number One is the **Marlins team of 2003**, which had a .692 record in September vs. .562 overall and went on to win the Series crown. In fact, I sense a creeping baseball truism sneaking into the e-mails and blogs I read: Momentum can overcome Good in the postseason.

So my question is, has anyone shown this to be true? I personally haven't seen a thorough analysis of momentum in postseason play anywhere. So, with my best sabermetric hat on my little pointy head, I loaded all the postseason results beginning in 1969 into my Excel spreadsheet to see what I could learn about Good and Momentum.

I chose 1969 because that was the first year of the two-division format; the Wild Card format wasn't fully introduced until 1995. By going back to 1969, I've got 35 years, 184 teams and 145 different series (excluding 1994, when baseball canceled the greatest show on earth). I looked at two winning percentages for each team: winning percentage for the year and winning percentage in September only.

I found out a lot more than 10 things. But that's the name of the column, so let's jump in.

Neither a team's overall record nor its record in September matters a whole lot.

In the 35 years from 1969 through 2004, the team with best overall winning percentage won the World Series only eight times. Let me emphasize: **the team with the best regular-season record has won the World Series only 23% of the time**. The winners include some of the best and best-known teams of our time: the 1970 Orioles, the Big Red Machine in 1975 and 1976, the Mets in 1986 and the 1998 Yankees.

How about the teams with the best September record? The answer is exactly the same: they won eight World Series, too. Same impact. In six of the eight examples, however, the team with the best September record was also the team with the best overall record. So there's a lot of overlap between the two groups.

In fact, in about half of the last 35 years (17, to be exact), the team with the best regular-season record was also the team with the best record in September. Of those 17 teams, six won the World Series. Even teams that were Good and had Momentum won it all only 35% of the time.

On average, teams that qualified for the postseason had a higher winning percentage in September.

Added together, these 184 teams had an overall winning percentage of .589 and a .612 winning percentage in September. This makes sense. These are the teams that qualified for postseason play, and a number of them made it on the wings of their late-season drives.

Here's a breakout of the overall and September records for all postseason teams over four-year periods.

Years	Overall	Sept.	Diff
1969-1972	0.608	0.627	.019
1973-1976	0.590	0.605	.015
1977-1980	0.595	0.625	.030
1981-1984	0.568	0.583	.015
1985-1988	0.592	0.636	.043
1989-1992	0.584	0.607	.023
1993-1996	0.581	0.607	.026
1997-2000	0.587	0.580	-.007
2001-2004	0.597	0.642	.045
Grand Total	0.589	0.612	.023

In the last four years, teams entering the postseason have had a significantly better September record than most previous years. This stands in stark contast to 1997-2000, when postseason teams actually had a worse record in September than their overall record, and it may contribute to the perception that momentum matters. We'll look at this more closely below.

Overall, Momentum is less important than being Good.

Now, here is the same table, but for World Series winners only:

Years	Overall	Sept.	Diff
1969-1972	0.621	0.700	.079
1973-1976	0.608	0.579	-.029
1977-1980	0.599	0.683	.084
1981-1984	0.597	0.573	-.024
1985-1988	0.584	0.594	.010
1989-1992	0.588	0.595	.007
1993-1996	0.593	0.622	.029
1997-2000	0.604	0.502	-.103
2001-2004	0.586	0.620	.033
Grand Total	0.598	0.607	.009

Compare the two tables. You can see that the World Champs, on average, had a better regular-season percentage (.598 vs. .589 for all postseason teams). Good does matter. But the champs actually had a slightly lower September record (.607 vs. .612) than the average postseason team. The most recent four-year period, with three Wild

Card winners, yielded a lower overall record for the winners than the initial entrants (.586 vs. .597), but the September records were lower too (.620 vs. .642).

Momentum was King from 1977 through 1980.

If ever Momentum was king, it was from 1977 through 1980 when the September record of World Series champs was .084 greater than their overall record. 1977 was the year the Yankees returned to glory, winning their first World Series since 1962. This was the **Billy Martin/Reggie Jackson** fight year, the year that made George Steinbrenner famous for buying his teams. Reggie Jackson established himself as "Mr. October" with his three-homer game in the 1977 Series. But it was their 41-12 record after August 10 that won the pennant.

1978 was even more dramatic, one of the most memorable pennant races of all time. The Yankees were seven games back on August 30, but the Red Sox subsequently went 3-14 (including six losses against the Yankees), to put the Yanks up 3.5 games on September 16. The Red Sox turned it around and won their last eight games to force a one-game playoff, which the Yankees won 5-4 on **Bucky "F—ing" Dent's** three-run homer. The Yankees played .733 ball that September (the third-highest total of any World Series winner) and went on to beat the Royals and Dodgers (same teams they beat in 1977) for the World crown.

1979 was the year everyone talked like the **Pirates**, as in *We Are Family* and **Willie Stargell**. The Pirates were a .500 team in early July, but they went 61-30, .670 over the last three months to win an exciting pennant race against the Expos. They swept the Reds (who were no longer quite so Big and Red) and won a fine 4-3 Series against the Orioles.

And then came 1980, the only year the **Philadelphia Phillies** have ever won the World Series. Like the Pirates the year before, the Phillies had a tight division race against the Expos, winning it in a final three-game showdown in Montreal. The

Phillies weren't a great club; their 91-71 record was the lowest of the four teams that qualified for postseason play. But they played .655 ball in September and won perhaps the closest League Championship Series ever against the Astros, when the last three games all went into extra innings. They then beat the Royals in the postseason to win it all.

Yet the team with the most Momentum of all didn't even win one series.

The Kansas City Royals haven't always been bad. It just seems that way. In 1977, the Royals had their best season ever, going 102-60, .630. These were the **Brett/White/Otis** Royals, with **Whitey Herzog** at the helm. As of September 1, however, the Royals were only playing .585 ball, 2 1/2 games ahead of the White Sox.

After September 1, the Royals went 25-5, .833 to run away with the division. This is the best September record of any team to make the postseason over the past 35 years. So what happened in the postseason? The Yankees beat them 3-2, particularly by scoring four runs in the last two innings of the last game to win 5-3. Which perhaps proves that nothing will stop momentum faster than a faulty bullpen.

Turnaround is fair play, however. In 1980, the year of the Phillies' victory, the Yankees actually had the best September record of any postseason team (.750) after winning a very tough race against the Orioles. Yet the Royals swept them in the League Championship Series.

From 1997 through 2000, Momentum was bad.

Each World Champions from 1997 through 2000 had a worse record in September than the season as a whole. In fact, two of those teams had September records below .500. The **1997 Marlins** made it to the postseason through the Wild Card slot and became the first Wild Card team to win it all. However, you can't really say they had momentum on their side, as they played only .444 ball in September, the second-lowest September record of any World Series winner in the past 35 years.

The worst September record for a World Series winner was .433; the **Yankees in 2000**. In fact, the 2000 Yankees lost 15 of their last 18 games but still managed to finish first in the American League East. They then went on to beat Oakland, Seattle and the Mets for the world title.

By the way, the fourth-worst September record of any World Series champion belongs to the **2001 Diamondbacks**, who played .476 ball the last month of the season. But the arms of **Johnson** and **Schilling** pulled Arizona through each postseason series until they had won it all.

The irony here is that, for all the recent focus on Momentum and the postseason, September records meant almost nothing a few years ago.

Specific matchups follow the same pattern.

Up to now, this analysis has been relatively superficial. Postseason analysis is complicated because teams don't all play each other. If a hot team matches up with a good team and another one matches up with a relatively worse team, a top-level view may not capture this nuance. So I looked at every postseason series from 1995 through 2004, the Wild Card years, and compared the overall and September records of the teams in each series.

Here's an example: the 2003 Marlins (.562 overall, .692 in September) first played the Giants in the postseason, who had a better overall record (.621) and the same September record (.692). They then played the Cubs, who actually had a worse overall record (.543) and a better September record (.704). They beat them anyway. In the World Series, they beat the Yankees, who had a better overall record (.623) and a slightly lower September record (.667). So, compared to their postseason competition, the Marlins were 102 points worse in the regular season, but 14 points better in September.

Now check out this table of the results of every series from 1995-2004 (counted twice, for each team's perspective):

Win% Diff	Series	Win%
<-0.120	5	.200
-0.120--0.04	25	.480
-0.040-0.040	80	.500
0.040-0.120	25	.520
>0.120	5	.800
Grand Total	140	.500

The left column is the difference between two teams' regular-season winning percentage. A difference of .120, for instance, might be the difference between a .640 team and a .520 team. Reading the bottom line of the table, teams with a winning percentage at least .120 better than the opposition has won four of five series.

By the way, the data is perfectly symmetrical because it includes the viewpoint of both teams. For every five teams with a record over 120 points higher, there were five teams with a record 120 points lower.

The key point here is that the results are not only symmetrical, they're logical. As the difference between a team's regular-season record and its opponent's moves from negative to positive, so does the team's chance of winning. And most series (130 of 140) are close enough that the outcome is relatively random; if .500 is completely random, these results range from .480 to .520.

Even on a matchup level, Momentum doesn't help.

I apologize, because the next table is kind of complicated (and this article is already long enough!). But the above table got me wondering: could Momentum be the difference in winning and losing when opponents are relatively well-matched?

To find out, I took the results from above and added a wrinkle: each team's relative advantage in their September record. Following is the table of how well each team did (expressed as the series winning percentage) when evenly matched but with differences in September play:

	Difference in September Records			
Win% Diff	Worst	Bad	Better	Best
-0.12--0.04	.500	.444	.625	.250
-0.04-0.04	.800	.536	.469	.200
0.04-0.12	.750	.333	.545	.500
Grand Total	.722	.500	.500	.278

To help interpet the table, look at the Grand Total line. What it's saying is that when teams had less Momentum (the "worst" difference in September records) they actually won 72% (.722) of the time. On the other end, teams that had the biggest advantage in September records only won 28% (.278) of the time.

That's right. There has actually been a negative correlation between Momentum and winning in the postseason the last 10 years. I can't explain it; it's probably just random variance. But momentum has meant less than nothing in the postseason over the past 10 years.

Did something happen in 2002?

To be fair, the story has been somewhat different the last three years. In 2002, 2003 and 2004 Momentum has actually been positively correlated with postseason success. One good example would be the 2003 Cubs (.543 winning percentage, .703 in September) beating the Braves (.623, .538) before going on to lose to the Marlins. But this is a recent phenomenon, against a background of 32 previous years in which momentum didn't really matter. Have things changed?

My theory is that we're confusing cause and effect here. Wild Card teams have won the past three World Series; during the same time, Wild Card teams have had great Septembers, with an average September winning percentage of .681. This is why it suddenly looks like Momentum matters. It just so happens that the high-momentum teams won.

Counterpoint: in 2001, the two Wild Card teams (Oakland and St. Louis) had an average September record of .791!! Yet neither one even made it to the World Series.

What this means for 2005.

What does all this mean for this year? It means that some people will pick the hottest team (say, the Indians or Yankees) to win it all, because they have momentum on their side. Based on this study, I'd say that the postseason is still a relatively random event. You're better off using the **guidelines developed over a year ago by our own Vinay Kumar**.

For instance, look at what the Baseball Savant has done, **analyzing the hottest starting pitchers entering the postseason**. This sort of analysis is more likely to yield something useful.

Let the real games begin.

References & Resources

The winning percentages and postseason records were calculated with data from the **Retrosheet Game Logs**, **the Lahman database**, and **Baseball Reference**. For my historical research, I leaned heavily on my *Sporting News Baseball Guides*, *Baseball's Pennant Races: A Graphic View* by John Warner Davenport and Leonard Koppett's *Concise History of Major League Baseball.*

Let me also add the following caveats to this analysis:
– Teams don't play the same schedule, so their won/loss records aren't directly

comparable. I used winning percentages here as a proxy, but it can certainly be improved upon.
– September records may not be the best definition of “momentum.” But I can’t think of a better one.
– It may be that “momentum”, when combined with other factors, does have an impact on postseason play. That would be another study for another day.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

Inside the Mind of Brian Sabean

by Dave Studeman

November 17, 2004

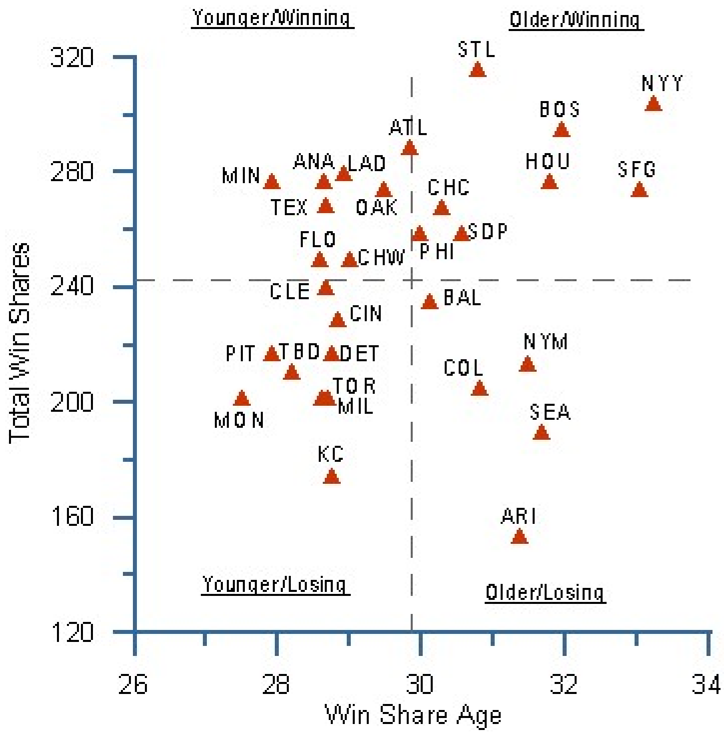
As **Aaron reported earlier**, and **as commented upon by others**, the San Francisco Giants signed Omar Vizquel to a three-year deal worth more than $12 million this past Monday.

Omar Vizquel will be 38 at the beginning of his three-year deal, which means he'll "make" more than $4 million at the advanced baseball age of 40 (more on the structure of the contract later). Want to know how many players made more than $4 million at the age of 40 this year? Just one — the player **whose contract represented the best deal in baseball last year**, Mr. Barry Bonds.

Now, I'm not saying Omar Vizquel will be a great value for the Giants in three years — far from it. But Brian Sabean is probably the only General Manager who would even think about offering a player in his late 30's a three-year contract. At least, I think he is. The question is, why? Why does Brian Sabean like old guys?

Due to Sabean's free agent signings and the phenomenal Bonds, the Giants are a pretty old team. How old are they? Well, a nifty way to measure team age is to take the average age of its players, weighted by the number of Win Shares each contributed to the team. Presented by Bill James in the **original Win Shares book**, this stat allows you to understand how old a team is, based on the actual ages of those who contributed the most.

Here's a graph of each team's Win Shares age on the horizontal axis, along with total Win Shares on the vertical axis:



I've added dotted lines for each of the two averages, which splits the graph into four quadrants. It's safe to say that you want to avoid the lower right quadrant; older, losing teams don't have much of a future. Pity the poor Dback, Rockies, Mets and Mariners fans. They're all virtually in the same boat, though some of them have better young players than others.

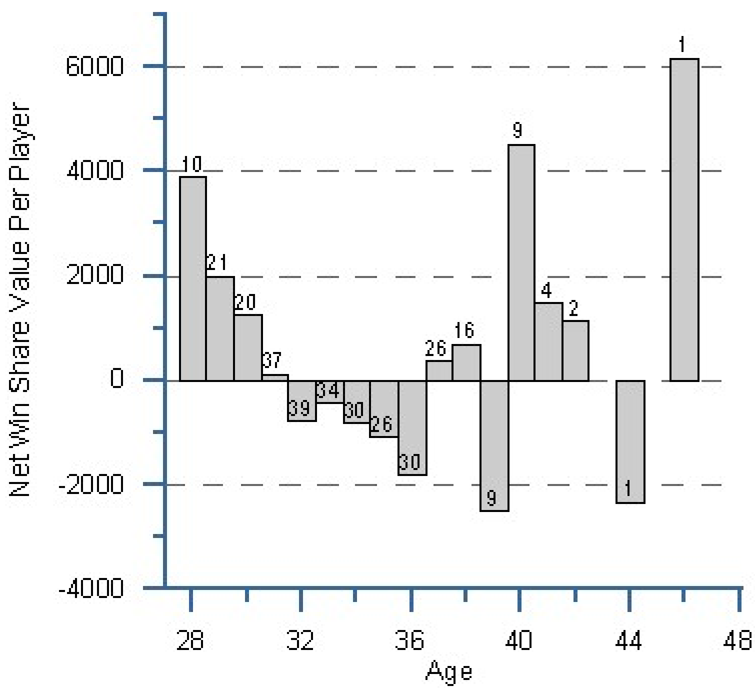
You do want to be in the upper left corner — younger and winning — because young teams improve and tend to be less expensive. In fact, the future looks bright for many of these teams, such as the Twins, Rangers and Angels.

The lower left quadrant is dominated by "small market" teams struggling to find the winning formula. But how about that upper right quadrant? Winning is good no matter how old you are, right? In fact, did you notice that this year's "final four" teams are all in the upper right quadrant? The four teams that didn't advance to the League Championships were the four younger teams (Braves, Twins, Dodgers and Angels). Think that's something **Billy Beane would want to know?** This year, at least, age seemed to help in postseason play.

There are actually five teams solidly in the upper right quadrant; San Francisco was the only one that didn't make it to the League Championships. In fact, they didn't even make the playoffs. San Francisco is getting older, too. In 2003, their Win Shares age was 32.5. This year it was 33.0. With Vizquel on board, no prospects in line and Bonds (and everyone else) continuing to age, their Win Shares age could creep close to 34 next year. If so, I wonder if they will be the oldest successful team in Major League history.

Other major changes from last year include the Yankees increasing from 31.9 to 33.2 and the Dodgers declining from 31.0 to 28.9 (thank you, Kevin Brown and Adrian Beltre). You can view a list of all 2003 team Win Share ages over at **the Baseball Graphs website.**

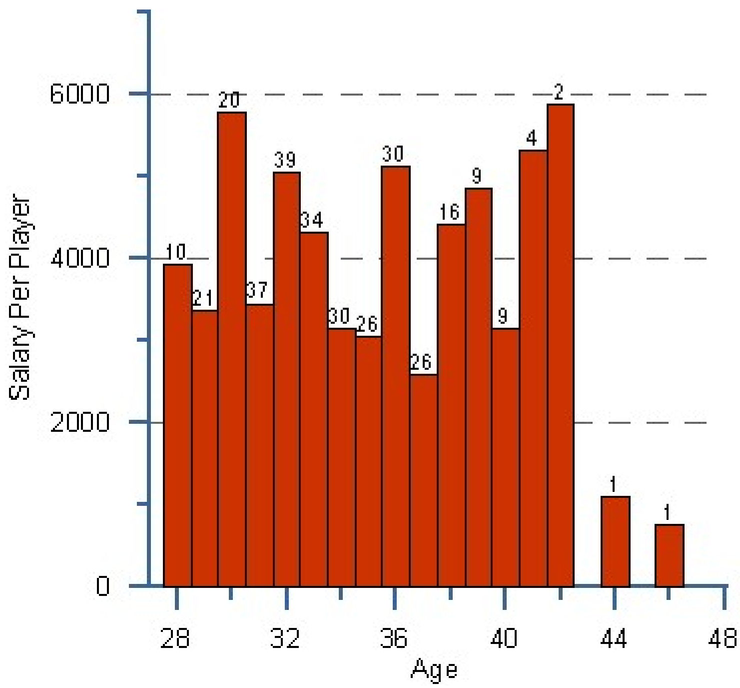
So Brian Sabean likes to sign old players. Really old players. So, you ask, are old players a good investment? Well, here's another graph — this one is a bar graph of the total 2004 Net Win Shares Value per player by age bracket. Net Win Shares Value is a measure of the value of a player's contract, as **explained in this article** and included in the ***Hardball Times Baseball Annual***. This particular graph includes all players signed as free agents, and the number above each bar represents the number of players in that age category:



First of all, there's a natural progression downwards from the age of 28 to 36. This is a reflection of a number of things, such as the effect of long-term contracts for players who age (think 36 year-old Mike Piazza or Sammy Sosa), a market that has declined in the past several years, so that younger players who have signed more recent contracts represent better deals (think 28 year-olds Vlad Guerrero and Miguel Tejada), and the structure of long-term contracts that pay out less in the beginning and more at the end (all of the above).

After age 36, things change, however, and there are a number of age brackets that represent very good value. In general, this graph resembles a "U" shape, with the bottom of the "U" occurring at age 36 and then turning positive (with a couple of exceptions) thereafter. This second part of the "U" is Sabean territory.

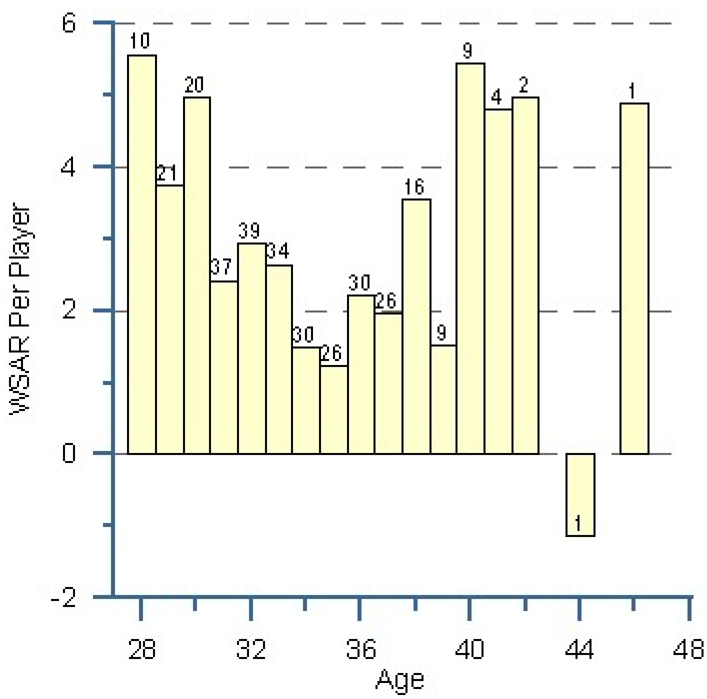
I'm going to show you two more graphs to try and explain why this happens. First, here's a graph of average salary by age:



See that jump in salary at the age of 36? 36-year-olds made $5 million a player last year, which was higher than most of the age brackets on either side of it. Those are some old, bad contracts that owners now regret. Salaries dropped 50% at age 37, to $2.5 million, which is the key reason that Net Win Shares Value improved so much at that age.

Some of the key value players at age 37, such as Tom Gordon, Reggie Sanders and the Giants' Marquis Grissom, signed deals in last year's bear market. But the Value "U" continues past the age of 37, into 38 and, after a big drop at 39, into the 40's.

Some of the valuable 38-year-olds were guys who had signed longer term contracts and were subsequently paid more in 2004 ($4.4 million per player). But these guys, including Moises Alou and Curt Schilling, delivered the Win Shares and Net Win Shares Value. Age 39, however, was a completely different story. Several of these guys were also signed to long-term contracts a long time ago (Kevin Brown, Al Leiter) but they just weren't that good in 2004, as measured by Win Shares Above Replacement. To illustrate, here's a graph (the last one, I promise) of Win Shares Above Replacement by age:



See, there's a general "U" shape in this graph, just like the Net Win Shares Value graph. Players hit bottom around age 35 and then, as the talent pool weeds itself out in the mid 30's, the older players are literally more valuable. Perhaps the average good player in his late 30's loses it more slowly than the average good player in his early 30's; I don't know if that's true, but it makes intuitive sense to me. That, plus the short-term contracts many of these players sign, make players over the age of 36 generally good values.

Who were the stars at age 40? Well, Bonds of course, but also B.J. Surhoff, Barry Larkin and Kenny Rogers. 41-year-olds included Randy Johnson and David Wells, and the key 42-year-old was Roger Clemens, whose value more than offset Jamie Moyer's terrible year.

Those two one-player samples at ages 44 and 46 are the Francos. John Franco, at the end of a fine career, was not really worth the contract he was given in 2004. But Julio Franco, who had **probably the second-greatest season of anyone at 46**, gives a nice flair to the end of the "U."

Give Brian Sabean a little credit. He has a feel for this phenomenon, and he's taken advantage of it. In a crazy way, this is the "Moneyball" philosophy. Approach the market in a different manner, and take advantage of values that others don't recognize. Billy Beane and Paul DePodesta valued OBP before others did; Brian Sabean values really old guys.

Now, this business of signing free agents before the deadline on purpose to give up a draft choice is simply suicidal. And I'm still not saying the Vizquel contract is a good one — two years should have been enough — but it isn't as bad as it appears on the surface. Vizquel will be paid a base salary of $2 million next year, $2.5 million in 2006 and $4 million in 2007. His bonus will be paid in four installments from 2005 to 2009 (skipping 2007), which lowers the overall cost of the contract. And Vizquel has an excellent chance to be better than **Cristian Guzman for at least the next two years**.

So maybe there's a method to Sabean's madness. Just for fun, try this thought exercise: If Bonds and Randy Johnson continue their elderly brilliance and the Giants continue to contend in 2005, what will Sabean offer Johnson when he becomes a free agent next year? Your answer will indicate just how "mad" you think Sabean is.

References & Resources

When computing age, I subtracted the player's birth year from the current year (2004), which differs from the typical baseball age (which is computed as of July 1st of the year). By doing this, I essentially round the player's age, measured in years and months as of July 1st, to the nearest age, measured in years.

As an example, Barry Bonds was born July 24, 1964, so his "baseball age" is 39. But, actually, he was 39 years and 11 months old on July 1st, so my method rounds his age up to 40, which seems more appropriate for an analysis like this. Thanks to Tangotiger for pointing this out (and I highly recommend **you read his aging analysis**).

I apologize for any confusion.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

In Good Standings

by Dave Studeman

August 2, 2005

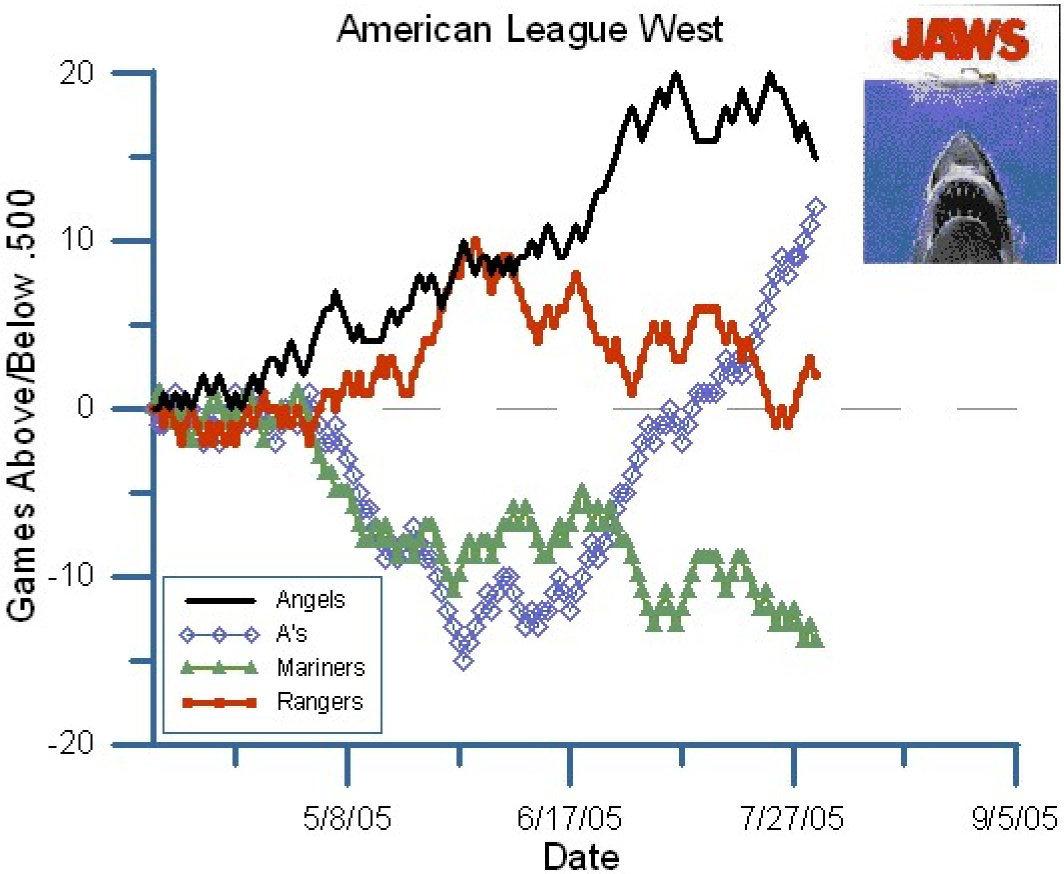
This has been one crazy year. In a **July 25 ESPN Insider article**, ESPN baseball analyst Rob Neyer talked about the number of teams that remain in contention for a postseason berth this year:

> “At dawn this morning, there are 21 teams within five games of either first place or the wild card. Yes, five games is an arbitrary standard–and just leaves out the Dodgers, who are five-and-a-half games behind the Padres–but hey, one has to draw the line somewhere. And 21 teams are a lot. Last July 25, there were 18 teams within five games of first place in one standing or another … and that was a lot.
>
> From 1996 through 2003, the number of “contenders” on the morning of July 25 was amazingly stable from season to season.
>
> ```
> 1996 -- 16
> 1997 -- 15
> 1998 -- 13
> 1999 -- 14
> 2000 -- 15
> 2001 -- 13
> 2002 -- 13
> 2003 -- 14
> ```
>
> So in all those seasons, roughly half the teams in the majors were close enough on July 25 to think, without drifting into Fantasyland, about postseason glory. But last year it was 60 percent, and this year it’s 70 percent.

70 percent of teams are in contention for a playoff spot! Between the six divisions, we’re seeing combacks, fades, competition both stellar and terrible. Here are some of the highlights …

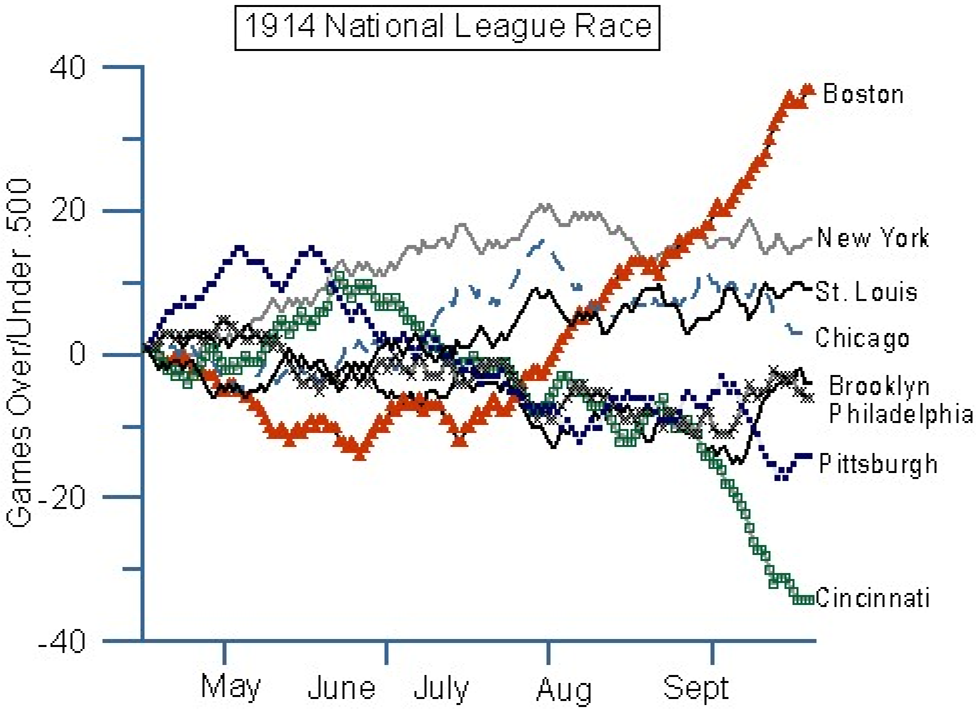
The Turnaround

I can't help it. When I look at the divisional standings graph for the American League West, I think of a certain Steven Spielberg movie:



Like an angry shark climbing from the depths of the ocean, the Oakland A's have been surging to the top, and they now have the Angels in their sights. THT writer Aaron Gleeman already did **a nice job of deconstructing the Athletics' surge,** but it's worth mentioning that there are very few historical precendents for this remarkable run.

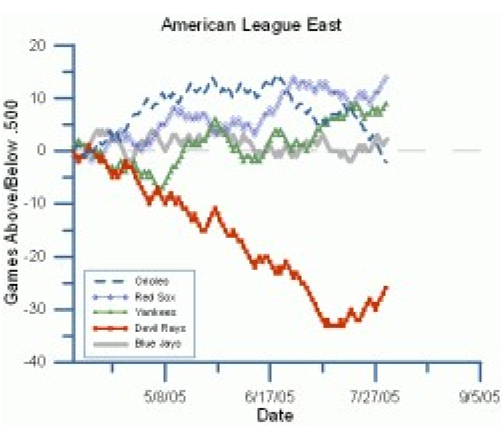
Here's a graph of the greatest "miracle" in baseball history, the worst-to-first surge of the 1914 Boston Braves:

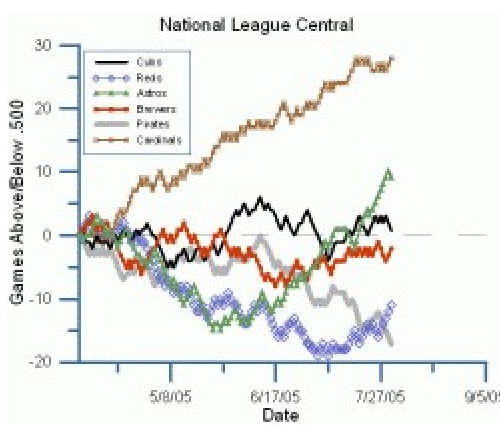


The Boston Braves, led by starting pitcher **Bill James** (36 Win Shares and a 1.90 ERA), among others, leapt from last to first in 37 days starting July 16, and proceeded to sweep the mighty A's (then stationed in Philadelphia) in the World Series. This year's A's can't top that—they were last in last place on June 24, 38 days ago and they only have to pass three teams instead of seven. But if no one stops this juggernaut, it will be one of the greatest bottom-to-top surges ever.

Mirror Images

The Tampa Bay Devil Rays are to the American League East what the St. Louis Cardinals are to the National League Central. Only in reverse.

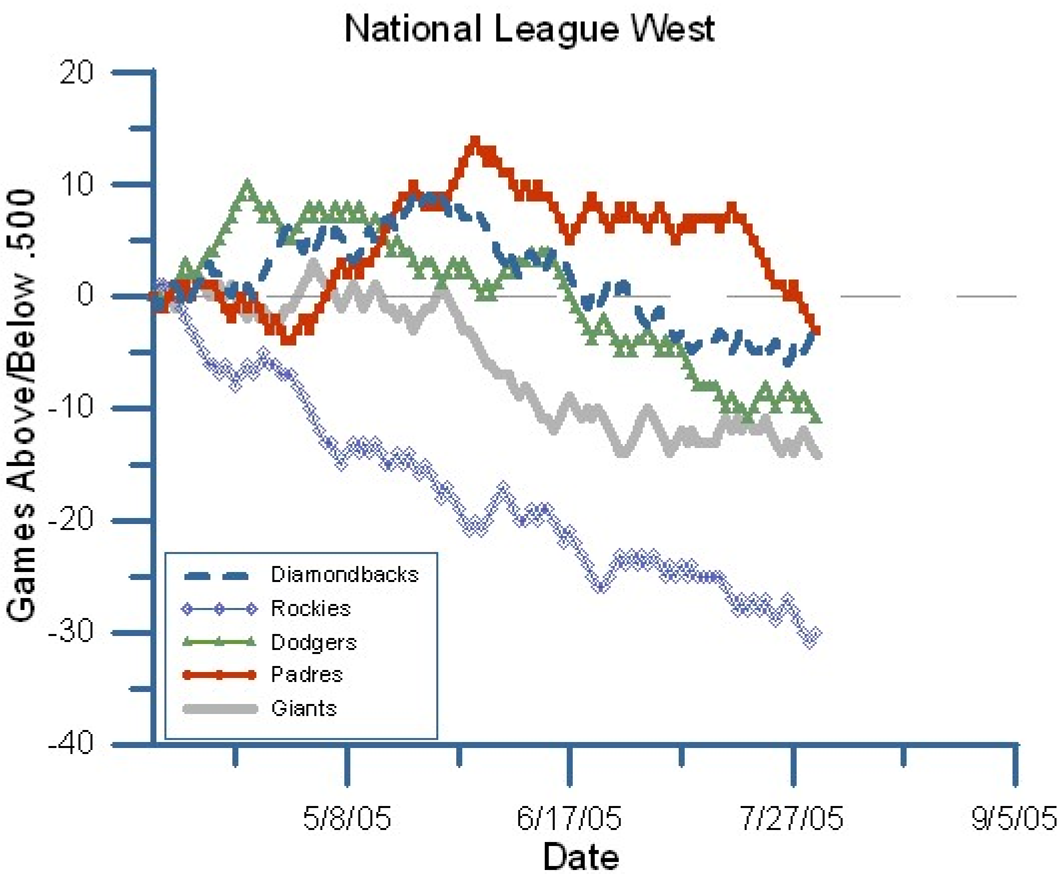




Discuss.

Les Pathetique

When we analyzed Berlioz's *Symphonie Fantastique* in my Symphony class in college, the professor insisted that we pronounce it correctly, instead of "Fantastic Symphony" or something like that. A little later, when we analyzed Tchaikovsky's *Pathetique* Symphony, he said "now you know why I insisted on correct pronunciation." I was reminded of this story while scanning the National League West:



According to Frank Vaccaro, of the SABR-L mailing list, the NL West is close to setting a record for the worst division ever. Here is Frank's list of divisions in which all teams were over or under .500, late in the season. The percentages on the left indicate the percent of the schedule that had been played at that particular point in the season.

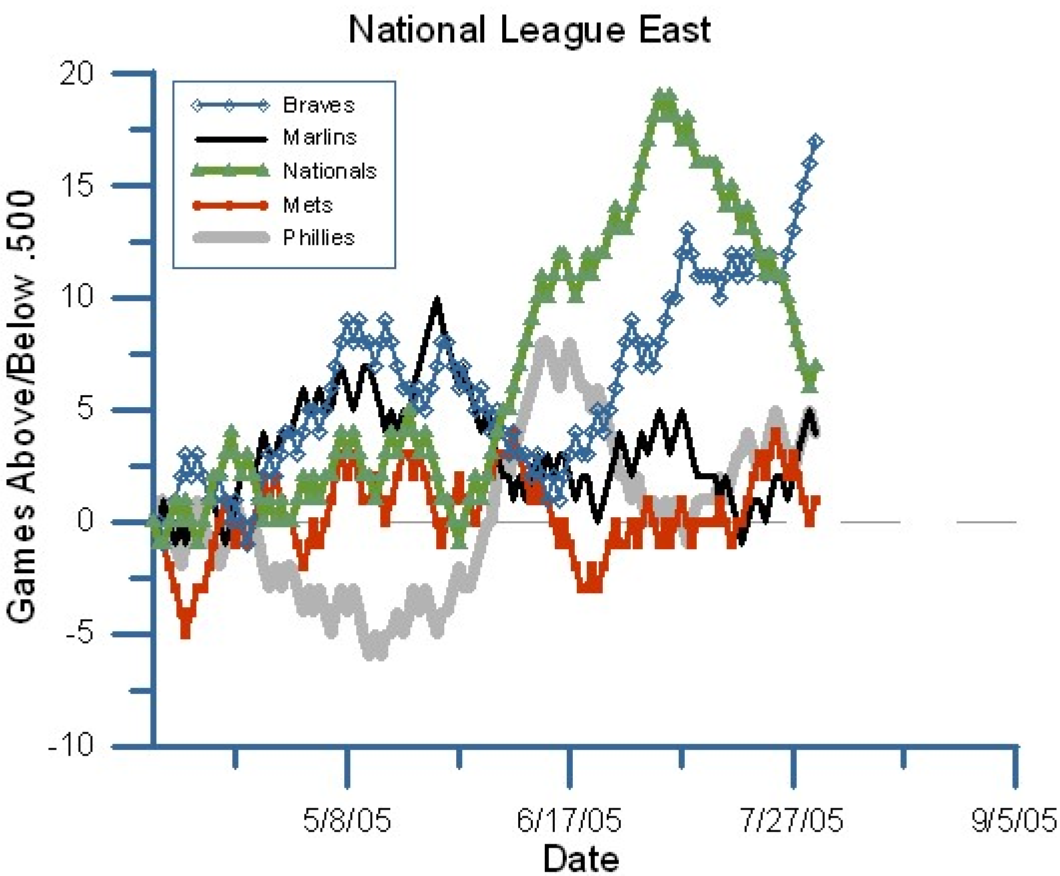
```
ENTIRE DIVISION OVER .500
Date, last place team and record shown, AM standings.

93.2%  1991 9/25  CAL         76-75  AL West
77.1%  1986 8/28  MIL         63-62  AL East
```

```
63.6%  2005 7/30  NYN         52-51  NL East
53.7%  1974 7/14  NYA, DET    44-43  AL East
52.0%  1981 9/7   TOR         13-12  AL East (2nd half)

46.2%  1996 6/23  SD          38-37  NL West
38.1%  1995 6/24  SF          28-27  NL West
36.4%  2000 6/10  TEX         30-29  AL West
30.2%  2001 5/27  COL, SF     25-24  NL West
27.7%  2004 5/31  PIT         23-22  NL Central

ENTIRE DIVISION UNDER .500
Date, first place team and record shown, AM standings.

70.3%  1994 8/12  TEX         52-62  AL West (strike occurred)
66.0%  1994 8/4   LAN         53-54  NL West
64.1%  2005 7/31  SD          51-53  NL West
56.1%  1997 7/13  HOU         45-46  NL Central
51.9%  1981 9/7   KC          13-14  AL West (2nd half)

47.5%  1996 6/24  HOU         38-39  NL Central
26.5%  1989 5/24  CLE         21-22  AL East
19.1%  1973 5/17  MIL         15-16  AL East
 9.2%  1999 4/22  TEX          7-8   AL West
```

In a nutshell, the NL West is eleven days away from a new record of being truly *pathetique*.

Peaked

It doesn't make it any easier for Nationals fans, but we all knew this was going to happen, right?



Man, even **Hannibal** would have had a hard time climbing that mountainous peak of Nationals wins and losses. As already noted by THT analyst Dan Fox, **one-run games giveth, and they taketh away**.

Don't forget that all of these graphs are available and updated daily in **our THT Teams section**.

References & Resources
The Boston Braves data is courtesy of **Retrosheet**, the most authoritative source of baseball statistics anywhere!

Thanks to Frank Vaccaro for permission to post his analysis from the SABR-L mailing list.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

Comments are closed.

My WAR Graph

by Dave Studeman

August 12, 2009

Let's take a retrospective look at four intriguing third/first basemen from the 1960s through the 1980s. **Harmon Killebrew**, **Graig Nettles**, **Dick Allen** and **Darrell Evans** were very different types of players, but all four played at similar positions at a high level for a long time. Only one of them (Killebrew) is in the Hall of Fame, however. Should he be?

This isn't going to one of those classic Hall of Fame articles, not quite. I'm using the Hall of Fame argument to discuss the ways we present data, particularly visual displays of data. But this is an interesting group of players. Let's start by comparing their standing in a couple of Bill James stats. I'll use the Hall of Fame Monitor score (from **Baseball Reference**—100 is someone likely to be voted into the Hall, based on MVP voting, basic batting stats, etc. etc.) and Win Shares Above Bench (WSAB—my own slight modification to Bill's Win Shares system, a measure of the contribution by each player to his team's wins).

Player	HOF	WSAB
Killebrew	178	190
Allen	99	207
Nettles	63	131
Evans	42	164

The HOF Monitor score is supposed to "predict" whether a player qualifies for the Hall. As you can see, it does pretty well in this case; Killebrew is far ahead of everyone else.

WSAB is a measure of how many wins each player contributed to this team, relative to a bench player. It doesn't care about things like winning MVP or batting titles. Plus, it includes the impact of fielding. While Killebrew has the highest HOF score, Allen outranks him in WSAB. Nettles and Evans appear to be far behind in both categories.

But there are other advanced win-based stats. One is **WPA**, though it doesn't yet include fielding. Using **Jeff Sagarin's WPA calculations** and adding my own WPA above bench calculation (using a .350 winning percentage for the bench level), I get the following totals for each player:

	WPA	WPAB
Killebrew	61	80
Allen	50	65
Evans	40	60
Nettles	16	35

Killebrew is still at the head of the class but Evans has crept up on Allen. (Sidenote: Evans had an .854 OPS in high-leverage situations vs. .786 in low-leverage situations. Allen did best in medium-leverage situations.)

When comparing these figures, remember that Win Shares are equal to three times wins. Comparing the numbers for Killebrew, for instance, he's 190 Win Shares Above Bench, or about 63 wins above bench, but he's 80 wins above bench on offensive WPA alone. Part of the difference may be that Killebrew hit better in high leverage situations, and he also wasn't a very good fielder.

Ah, fielding. The critical difference between WPAB and WSAB should be related to each player's glovework, but the difference doesn't appear to play out here. Graig Nettles was undeniably a great fielder, but WSAB doesn't give him the boost over WPAB that you'd expect (at least, not relative to the others on our list). Not a surprise—Win Shares is known for its conservative approach to giving players credit for their fielding contributions.

Enter Wins Above Replacement (WAR). WAR is very similar to WSAB, though it differs in two key ways:
{exp:list_maker}The goal of WSAB and Win Shares is to measure each player's contributions to his team's win. The goal of WAR is to measure the value of each player in terms of wins. The difference is subtle but important.
WAR is denominated in wins instead of wins times three. {/exp:list_maker}Actually, the math behind the two systems is very different, but I don't want to get into that here (particularly since I'm no expert on the subject).

WAR has been made possible by the tremendous work by the folks at **Retrosheet** and their detailed play-by-play database. Tangotiger and a lot of other people have worked out a system that includes not only the impact of batting, but fielding, baserunning and whatnot. The big difference between WAR and other systems, however, is the emphasis it puts on fielding. Witness the WAR numbers of our four players (a big thanks to **Sean Smith** for the data):

	WAR
Nettles	61.6
Allen	61.2
Killebrew	61.1
Evans	57.3

It's a virtual dead heat, with Evans a bit behind Nettles, Allen and Killebrew.

Think about this for a second. We've moved from the Hall of Fame Monitor (a pretty good system that represents how Hall of Famers are selected today) to some better win-based systems to one that fully incorporates the hardest critical skill to measure, fielding. And we've totally upset the order of things.

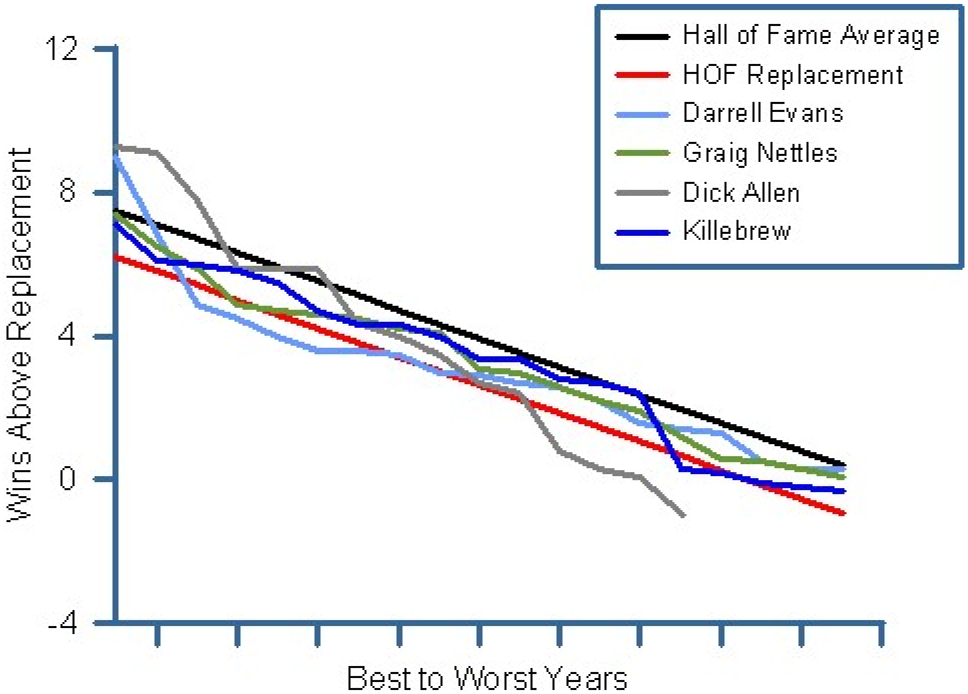
WAR does differ from the other systems in other respects, and those choices are certainly debatable. But it is fielding that really separates WAR from the other systems. It makes a huge difference. Graig Nettles is 140 runs above average in **Total Zone**, Sean's fielding evaluation system. That makes him one of the best fielding third basemen of the past 50 years. At the other extreme, Dick Allen was 97 runs worse than average in the field. That's a difference of almost 240 runs, or approximately 24 wins.

For the record, Evans was 39 runs better than average in the field and Killebrew was 63 runs below average.

We can all quibble about WAR and some of the details behind it. But I don't doubt that we're much closer to measuring a player's true value to his team than we have ever been before.

Whew. That was a long introduction. Let's take this examination of WAR a little further. Total WAR figures are good and all, but there's also the question of how much each player contributed in individual years. Arguably, huge, MVP-type years are worth more than the WAR scale would indicate. We can add some fancy math to calculate the difference, but this is one of those things that might be best served by a visual graphic. Enter the WAR graph.

I believe WAR graphs were first created by Sky Kalkman, who has made them a regular feature at **Beyond the Boxscore**, along with the other excellent BtB writers. This is what they look like:



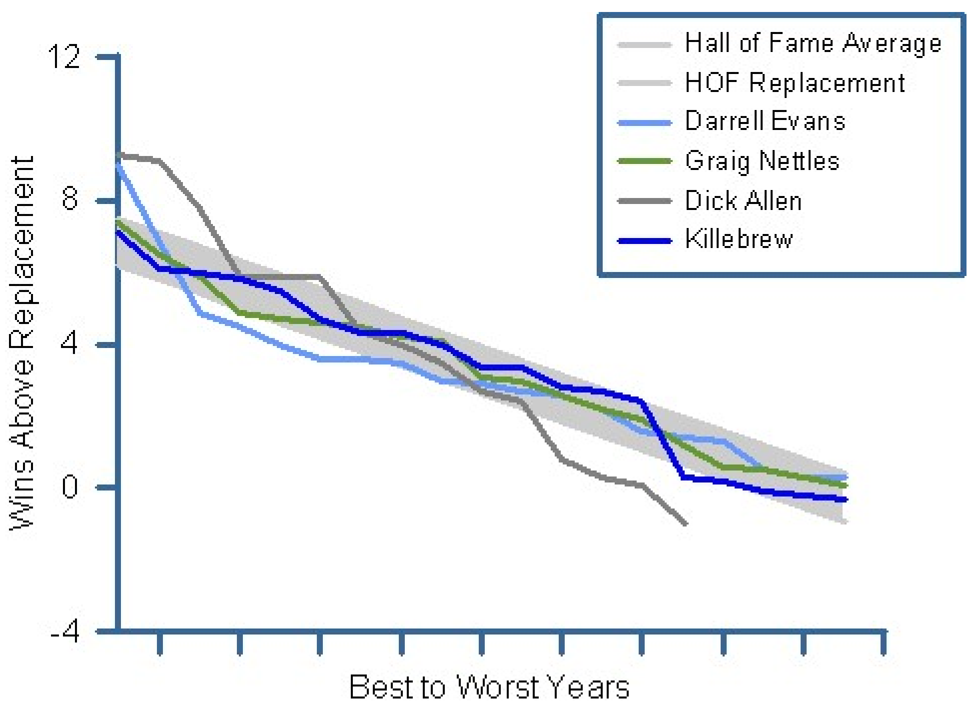
This is my own recreation of a WAR graph, not one that was posted at Beyond the Boxscore. The idea is that you line up each player's seasons ranked by WAR, from their best to their worst year. I think that the graphs originally just had a few players in the graph, but extra context was added by including lines representing an average Hall of Famer and a replacement-level Hall of Famer. (I just eyeballed those lines myself and added them to my graph.)

It's easy to see why these graphs are so popular. They give you more information beyond just listing each player's total career WAR. In particular, they allow you to see how consistent each player was, and whether he had some outstanding peak seasons. They pass the critical test of any graph: they communicate key information more easily than a table of the same data would.

Unfortunately, this graphic doesn't really help that much, at least in my opinion (of course, I have no one to blame but myself). The lines cross over so much that the entire thing looks like a "spaghetti chart" in which it's hard to untangle the facts. You know how they say that statistics lie? Well, graphics can lie even more easily unless the graphic designer is very careful about the layout of the graph and the way its constructed.

I'm going to try to improve my first attempt at a WAR graph. The most important step in untangling that mess of spaghetti is to figure out which strands are different than the others and can be treated with a different visual style. Said differently, can we throw away some of the strands?

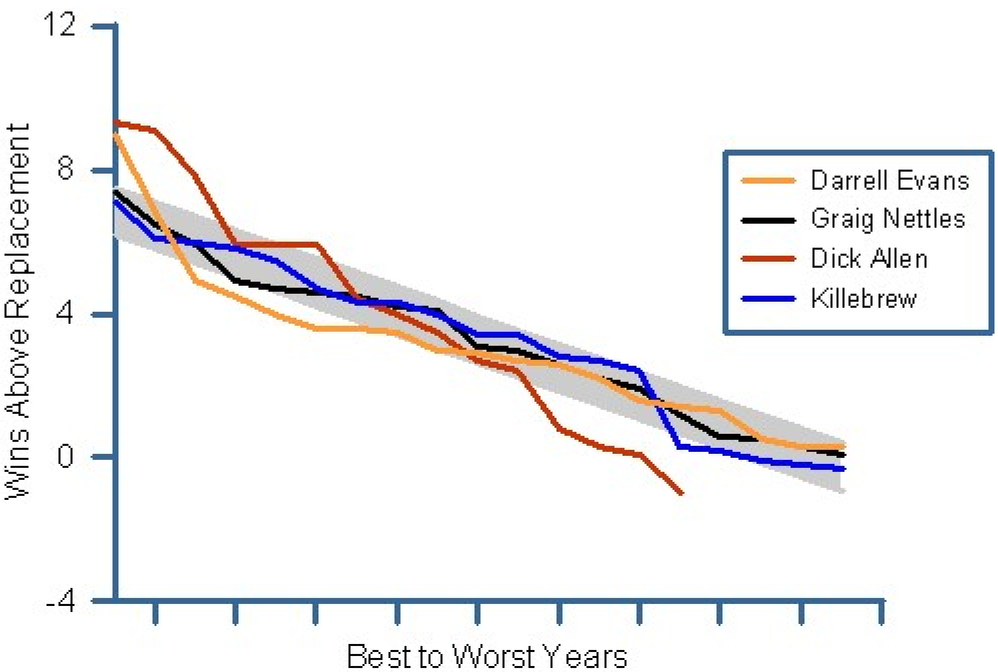
The obvious candidates are the Hall of Fame lines. In fact, the two Hall of Fame lines provide the really important context for the graph (more important than either the "X" of "Y" axis) and we can really make them stand out by graying in the area between them and getting rid of the lines altogether. Here's the result:



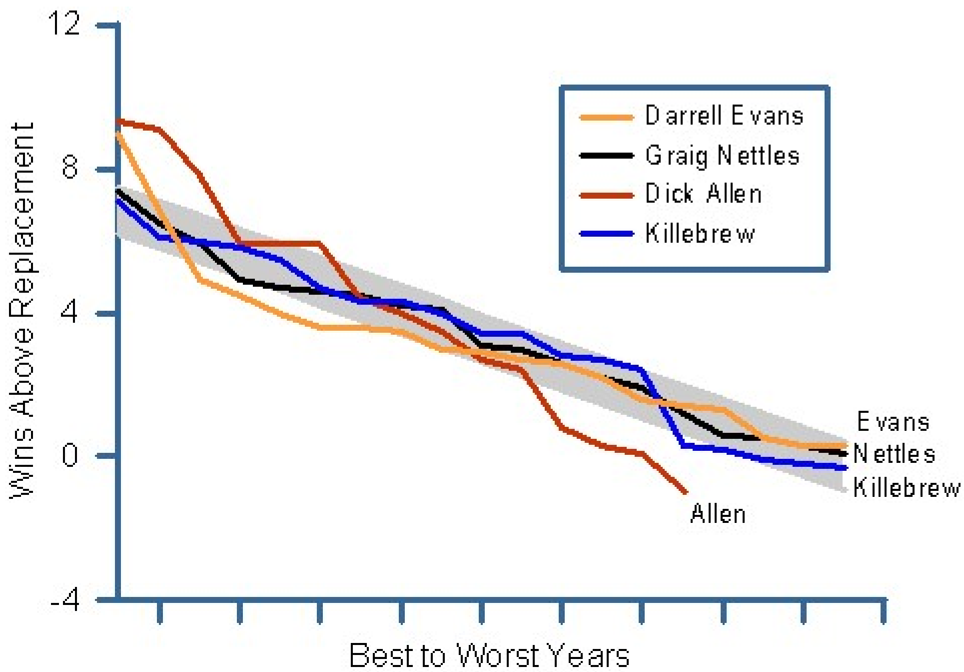
Okay, I think that's a big step in the right direction. However, it's still difficult to make out the strands and really understand what the graph is saying. That's because the differences between the lines aren't very stark, and the gray background has made it a little more difficult to pick out some of them.

There are several things you can do to untangle the lines of a spaghetti chart. One of the biggest mistakes I see in graphs is that people often only use color to distinguish between lines, and the colors are often hard to pick out (particularly for the 10 percent of men who are partially color blind, like me). I would encourage folks to change the line styles (dashes often work well) or add symbols to one or two of the lines. You don't want the graph to be too busy, but you want to make sure the lines are distinguishable.

I played around with my WAR graph and decided that symbols and such didn't work in this case because the lines are all so tight. But I created starker color differences between the lines and moved the most distinguishable one (the black line, Nettles, which is also the least critical because it follows a "normal" downward slope) behind the others. I think that makes these lines a little easier to pick out.

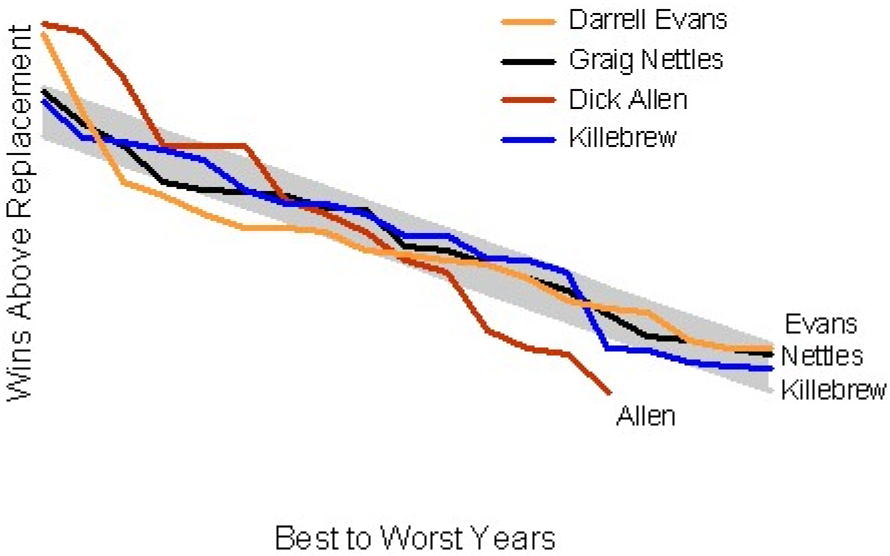


You'll notice that I also took the Hall of Fame references out of the legend cause they weren't really needed. Which leads me to another important point. Too often, I see people use legends that are far away from the actual visual data. That doesn't help the reader much, because it's hard to associate the two. So I was sure to keep the legend close to the actual data, and I also added names next to the end of each line to further help the reader label each line. Like so:



This is looking pretty good, I think. You can see that Allen had a couple of MVP years (1964 and 1972) and Evans also had one (1973, when he was second in the league in WAR but 18th in MVP voting). But Evans' third- through sixth-best seasons were below HOF replacement level. Allen had a short career, while Nettles and Killebrew were in that "Hall of Fame lower half" on a consistent basis.

That is one interesting graph. We've learned a lot from it. But I would like to do one more thing. The graph maven **Edward Tufte** talks about a graph's "data-ink ratio." The idea is that you don't want to waste ink in a graphic (gridlines and the like are usually useless) because it detracts from the data itself. We have an opportunity to really improve the data-ink ratio in the WAR graph, because the Hall of Fame band makes the "X" and "Y" axes pretty much irrelevant. In fact, let's get rid of them altogether, as well as the border around the legend.



That's my idea of a really good WAR graph.

References & Resources
Many of these breakouts were made possible by **Baseball Reference**. Many thanks to Retrosheet, Tangotiger, Sean Smith, Sky Kalkman and all the other BtB writers.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.

A Paean to APBA

by Dave Studeman

March 31, 2004

In my family, I'm known as the "other" baseball fan. My brother Woody, who's 10 years older than me, had already set the bar pretty high by the time I was even born. I'm not complaining, mind you. I'm bragging.

Woody was born in Cooperstown, NY at the end of the second World War (my Mom was raised in Cooperstown, and we spent our summers there as kids) and he's been thinking and breathing baseball ever since. In Sunday School, he invented a game with the other kids called "Bible Baseball," in which you had to open the Bible to a specific book within three tries. Each failed try was, of course, a strike. Three strikes and you were out.

In 1958, Woody began playing a game called **APBA**, or American Professional Baseball Association. In APBA, you play a simulated baseball game based on dice rolls and specific player, pitcher and situation cards. Needless to say, Woody took to APBA like a duck to water.

APBA cards are the essence of baseball; they are baseball logic made explicit, there for the eye to see and the mind to comprehend. No computer simulation game, with its hidden code and logic, can impart so much understanding.

When you play a game, you roll two die as your "chance" mechanism, then you look up the number on the player's card. Next, you interpret the number according to the pitcher and game board. It sounds complicated, but it's the complication of baseball. Going through the steps is an education. Of course, once you've played for awhile, the games go really fast.

In addition to the batting cards, APBA grades fielders and pitchers. A "BXYZ" is one heck of a good pitcher, and the actual interpretation of the batter's card depends on the pitcher he is facing and the quality of the fielders. Plus, there are game boards for each of the eight base situations (empty, runner on first, etc.) and the interpretation of the roll can differ by situation as well.

A little history: APBA was invented in 1951, about 10 years before Strat-O-Matic. In their book "**Curve Ball**," Jim Albert and Jay Bennett review the history of baseball simulation games. Comparing the two, they said:

> The Strat-O-Matic model, in terms of consistency, is an improvement over the APBA model… On the other hand, something is lost as well: the interaction between pitcher and batter is not taken into account. In APBA Baseball, the effect of the pitcher depends on how his characteristics (defined by the pitching grade) interact with those of the batter (defined by his batting card).

Once he started, Woody played APBA with a vengeance. One summer when he was a camp counselor, he decided to play APBA his entire day off. 24 straight hours. He managed to play 61 games.

Included in the APBA game set is a dice shaker, consisting of a small cardboard tube with a metal bottom. Rattling the dice in the shaker yields a distinct sound, one I remember falling asleep to many times as my brother played the game in his room.

It actually got really loud in our basement, where Woody and his friends played APBA every Friday night. There was no continuous ownership at that time. They just picked their favorite teams and played. In 1960, a few of them decided they wanted a league with continuous player ownership, trades and the like, and they established the North East League (NEL).

When they went off to college, they wanted to keep playing together, so they established rules for playing APBA by mail, creating the first play-by-mail baseball fantasy league. Keep in mind that this was many years before anyone had ever heard of rotisserie baseball. I mean, Kennedy was president.

It worked so well that the NEL is still going strong today, 43 years later. And Woody is its sole charter member. He has literally never stopped playing APBA.

This has made Woody a minor legend in APBA circles. He sometimes gets asked if he's the "real" Woody Studenmund when he runs into other APBA devotees. Over his 43 years in the NEL, he has won 3,852 games, at a .614 clip, including 14 NEL titles.

The NEL has been around so long that one of its current members, Mark Featherstone, is actually the son of an earlier member (who was in the league for 32 years). Mark has said, "It was a dream come true to join the league. But it's weird to trade with someone like Woody; who was always 'Mr. Studenmund' when I was a kid — it's like trading with a legend."

Amazingly, five of the current twelve members have been in the league for over thirty years. I asked Woody for the secrets of the NEL's success.

> "The NEL conventions have kept the league strong. The league members come from Canada, California, Michigan, Wisconsin, New York, New Jersey, Illinois, and Connecticut, we get together once a year for a live convention, and we've had a convention for something like 40 straight years. Some years, we have 100% attendance, and other years we go beyond that, with former and prospective managers chiming in.
>
> The friendships are amazing. Somewhere, and I'm not sure that I can put my finger on it, we stopped being angry competitors and began really caring about each other. Ken Meyer, for example, was in my wedding.

Woody likes to play in other APBA leagues as well. In one of those weird, small-world coincidences, I found out that **Rich Lederer** actually played APBA with Woody for five years in the 1970's. I recently asked Rich for some recollections of playing with Woody.

> "Playing Woody was like going up against the modern-day Yankees. I distinctly recall Woody drafting **Bobby Bonds**, **Cesar Cedeno**, and **Reggie Jackson** the first year — perhaps the three best power/speed combos in all of baseball back then. He also had **Johnny Bench** and **Bobby Grich** to boot and had this 19-year-old kid on his team by the name of **Robin Yount**.
>
> You had to be on your toes when talking trade with Woody. I had the utmost respect for him, and I learned a lot about not only APBA from him but how to judge players in

> general. I saw how he built his teams on power, speed, and walks. His player cards rarely had "7s" on them. That was the one "singles" number that even the best pitchers had trouble stopping. Instead, Woody emphasized players with 1s, 5s, 6s, 11s, and 14s. A "1" was always a HR and a "5" was more often than not a HR with runners on base. A "6" was a double. An "11" a single and stolen base (it acted almost as powerful as a double in APBA), and a "14" was a base on balls.

Along the same lines, I remember bragging to Woody about a great young shortstop the Mets were bringing up named **Teddy Martinez**. Martinez was a good fielder who had hit for average in the minors. Woody asked me, "What does he do?" I think my reply was a blank stare. Do? Why, he could field, and he hit for average. The reports were good. What else was there?

Woody said "Does he steal bases? Does he hit home runs?" I replied that no, he didn't, and Woody quickly lost interest. I didn't fully understand it at the time, but I had just received a vital lesson in sabermetrics.

I recently asked Woody for some of his favorite APBA memories. Here are a few:

> "Before the Internet, I used to contact players directly to find out information. When **Joe Morgan** hurt his knee early in his career, I wrote him to ask if he was going to be able to steal 20 bases a year (so he'd get an "11," a crucial number in APBA), and he wrote back and said no problem. I traded **Pete Rose** and **Willie McCovey** for Joe at the height of their careers, and I never regretted it.
>
> I also wrote **Billy Martin** when he was the manager of the Twins to encourage him to play **Harmon Killebrew** at third for one game (and he did, allowing me to play Killebrew at third for an entire season until the league changed the rules), and I called the Phillie front office when I was having trouble deciding whether to trade for **Mike Schmidt**. They said that the ball just jumped off his bat, so I traded **Vida Blue**, **Carl Yastrzemski**, **Graig Nettles**, and a ton of cash [think draft choices] for Mike, and I never regretted it, either.
>
> Turning over my roster in three and a half days. Over the years, I've often been in another league besides the NEL, and recently, my second team has been the Eagle Rock Rocks in the Mail3 league. When I first was granted the franchise, I didn't like any of the players, so I traded them all away in the space of three and a half days. One of the managers called that time "Woody's Reign of Terror."
>
> Mike Witt's perfect game was my biggest thrill. Back twenty years ago or so, I owned **Mike Witt**, and in his last start I'd calculated that he needed to pitch a shutout with

> zero walks to get a BZ pitching grade. I listened to the game on the radio and hung on each pitch as, amazingly, he pitched a perfect game!

For reasons I can't explain, I've never quite taken to fantasy baseball. Never played in a league and never wanted to. But I do understand and love APBA. In March of 1970, I replayed the 1969 Mets' season, just to enjoy it again. I find APBA cards as nuanced and informative as a box score. And sometimes, late at night, I think I hear the sound of dice rattling against metal, signalling that all is right with the world.

The 2004 NEL convention is being held this week. We'll let you know how it goes.

References & Resources
In addition to the official APBA site, there is:
– A page of **APBA downloads**
– A **fan's APBA FAQ**

The quote from Mark Featherstone is from the August, 1990 issue of "Fantasy Baseball."

"Curve Ball" is an excellent book about baseball statistics. I highly recommend it.

Dave Studeman was called a "national treasure" by Rob Neyer. Seriously. Follow his sporadic tweets **@dastudes**.