
By the Numbers

Volume 18, Number 1

The Newsletter of the SABR Statistical Analysis Committee

February, 2008

Review

Academic Research: The Effect of Steroids on Home Run Power

Charlie Pavitt

How much more power would a typical slugger gain from the use of performance-enhancing substances? The author reviews a recent academic study that presents estimates.

R. G. Tobin, On the potential of a chemical Bonds: Possible effects of steroids on home run production in baseball, American Journal of Physics, January 2008, Vol. 76 No. 1, pp. 15-20

This piece is really beyond my competence to do any more than summarize, but it certainly is timely, and I thought a description would be of interest. Tobin's interest is in using available data and models to estimate the increase in home runs per batted ball that steroid use might provide. After reviewing past physiological work on the impact of steroids on weightlifters, he decided to assume an

increase in muscle mass of ten percent from its use, leading to an analogous increase in kinetic energy of the bat swing and a five percent jump in the speed of the bat (Energy is

proportional to speed squared, and 1.05 squared is approximately equal to 1.10). Using an equation for bat/ball collision from past work results in a four percent increase in the speed of the ball as it leaves the bat.

Next on the agenda is a model for the trajectory of the ball. There is apparently disagreement from past studies on the impact of air resistance, with quite different models following

from different assumption about it. Tobin examined the implications of several, with the stipulation that a batted ball would be considered a home run if it had a height of at least nine feet at a distance of 380 feet from its starting point.

Computations based on these models results in an increase from about 10 percent of batted balls qualifying as homers, which is the figure one would expect from a prolific power hitter, to about 15 percent with the most conservative of the models and 20 percent for the most liberal. These estimates imply an increase in homer production of 50 to 100 percent.

Tobin then takes on the impact on pitching, with a ten percent increase in muscle mass leading to a five percent rise in pitching speed, which is close to five miles an hour for a power pitcher. Tobin cites some work, including a piece from John Dewan's

ActaSports (www.actasports.com/sow/php?id=101), implying that such a rise would lead to a half run a game improvement in earned run average. I looked at that piece and, given what little information it has, that seems a fair assessment.

Charlie Pavitt, chazzq@udel.edu ♦

In this issue

Academic Research: The Effect of Steroids on Home Run Power	Charlie Pavitt	1
Low Risk – Any Reward?.....	Eric J. Seidman	3
Career W-L Records for Hitters.....	Tom Hanrahan.....	7

Informal Peer Review

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

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

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

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

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

Low Risk – Any Reward?

Eric J. Seidman

While it makes big headlines when a team signs a free-agent superstar, there are at least as many cases of a team signing an average – or below-average – player, in hopes that his performance will exceed his modest salary requirements. Here, the author investigates "low risk" pitcher signings of this type, to see if and when they work out.

Without fail, each offseason brings with it a multitude of personnel concerns for each team. Players lost to free agency need to be replaced; those no longer effective require upgrades; and it seems that every team has the worst bullpen in the league. While the big name, big money acquisitions hog the headlines, the majority of transactions involve less money given to average, slightly above-average, or risky players. These moves have garnered the nickname "low risk, high reward." Since such little money is being committed it seems to be a worthwhile investment to take a flyer on a formerly successful veteran; he may be able to regain past form. If not, the team will not suffer much because they were not terribly committal in the first place. With these signings becoming more prominent each year I decided to investigate, with regards to pitchers, whether or not teams are actually rewarded highly for their low risk.

Defining Low Risk, High Reward

The first aspect of these signings refers to contract length; low risk implies a lack of commitment to the duration of a contract. When discussing the contract duration I am going to consider the following situations to be of low risk:

- Minor League deal
- Waiver claim
- 1-yr deal
- 1-yr deal w/option

The second part of low risk signings involves offering the pitcher a lesser contract in exchange for said pitcher getting another shot at playing in the major leagues. When determining monetary criteria the major factor to take into account is the differential in team payrolls. It would not be fair to set a maximum dollar value at four million dollars because that figure would account for a different percentage of payrolls for certain teams. Four million dollars would equate to 3.96 percent of the 2007 Chicago White Sox payroll while simultaneously accounting for 10.71 percent of the 2004 Washington Nationals payroll. Clearly the figure was more significant to the Nationals because it made up a higher percentage of their payroll. Then you've got instances of 2008 Alex Rodriguez making as much as the 2007 Florida Marlins!

Therefore, using percentages of team salary makes more sense as opposed to raw figures. In order to qualify as a Low Risk, High Reward pitcher signing one must meet the aforementioned duration criteria as well as account for no more than 5.25 percent of the team payroll. If there are 25 players on a team they will average 4.0 percent per player; I allowed an extra 1.25 percent to give some leeway to small market teams in rebuilding phases. For consistency's sake, though, I evaluated all of these signings on a case by case basis in order to determine if any truly should not merit inclusion. For the most part the moves I logged consisted of salaries below 4.0 percent.

Logging Transactions

Using the ESPN.com transaction archive I logged all of the moves from October 2002 through August 2007 that met the contract duration criterion. Then, the USA Today Salary Database helped determine individual and team salaries. After entering all of the data I removed any signing in excess of the 5.25 percent. Overall, this left me with 352 pitcher signings to examine.

It is interesting to see the frequency, or lack thereof, per team. For instance, the Cleveland Indians made 22 such moves in this span whereas the San Francisco Giants made just two: Al Levine in 2005 and Russ Ortiz in 2007. I expected to have five pages solely consisting of Athletics transactions but found only eight qualifying moves. It then dawned on me that Billy Beane has an undying love of young arms and his low risk signings were more on the offensive front.

In terms of individual players, a fair number had been low risk signings as many as three times. However, only three pitchers achieved the feat four or more times: Terry Mulholland (5), Pedro Astacio (4), and James Baldwin (4). For the record, of the thirteen total seasons between these three pitchers, only Astacio's 2005 with the Padres and Mulholland's 2004 with the Twins actually produced significantly positive results.

Table 1 shows the number of low risk pitcher signings per team.

Determining the Amount of Reward

With everything logged I decided to use Keith Woolner's VORP (Value Over Replacement Player) statistic to gauge the actual reward levels of these signings. VORP made the most sense to me since the statistic acknowledges the contributions made by league average players; though these players are not superstars it would not be fair to deem them ineffectual in a zero-sum game like baseball. Other statistics will use average players as the 0.0 baseline but VORP uses a below average player.

In its simplest definition VORP measures the amount of runs contributed above what a replacement-level player would produce in the same percentage of team plate appearances. With regards to pitchers it refers to the amount of runs saved above what a slightly below average pitcher would give up if given the same amount of opportunities. It does not account for defense in the way that Win Shares does but, because we are measuring pitchers, the amount of runs saved vastly outweighs this; I am not sure any general manager has ever signed a pitcher solely based on his defensive ability.

In terms of logging statistics, only the VORP total(s) for the duration of the low risk contract qualified. For instance, if a player signed a low risk deal in 2003 and then went on to have productive seasons with the same team in 2004 and 2005, only the 2003 VORP was recorded. The 2004 and 2005 seasons were under different contracts unlikely to meet the 5.25 percent maximum. Additionally, due to the productivity of the initial low risk season, the risk is no longer existent; the team understands what type of production the pitcher could provide. Signing him to a deal still 5.25 percent or less of the team's salary would not necessarily qualify as low-risk but rather as a reasonable upgrade over the low risk contract.

Table 1 – "Low Risk" Pitcher Signings, 2002-2007

Arizona	12	Milwaukee	13
Atlanta	10	Minnesota	8
Baltimore	9	NY Mets	20
Boston	14	NY Yankees	10
Chicago (AL)	6	Oakland	8
Chicago (NL)	8	Philadelphia	9
Cincinnati	20	Pittsburgh	11
Cleveland	22	San Diego	13
Colorado	19	San Francisco	2
Detroit	8	Seattle	9
Florida	10	St. Louis	19
Houston	17	Tampa Bay	15
Kansas City	15	Texas	9
LA Angels	5	Toronto	12
LA Dodgers	10	Washington	9

Analyzing VORP

In order to determine if these signings worked out the question of what constitutes reward must be answered. Clearly, anything 0.0 or below would be detrimental as the pitcher would have produced a slightly below average season at best; 0.0 or below would imply no reward all the way to negative effects. Additionally, players never called up to the big leagues after signing a low risk deal provided no reward as they were never given a shot. While this would not necessarily be negative in the vain of Jose Lima's 2005 season it would still suggest no reward was earned. The question then becomes: How do we analyze positive VORPs?

After some careful thought it was determined that, while anything above 0.0 is technically positive, there are different levels of positive rewards. A pitcher could provide low reward, medium reward, or high reward. Here is the reward criteria in terms of VORP totals:

Negative Reward: VORP < 0.0
 No Reward: VORP = 0.0 or N/A
 Low Reward: VORP = 0.1 to 9.99
 Medium Reward: VORP = 10.0 to 19.99
 High Reward: VORP = 20.0 +

Low Risk Breakdown

Of the 352 low risk pitcher signings in this five-year span:

16 were High Reward
47 were Medium Reward
101 were Low Reward
96 were No Reward (0.0 VORP or never called up)
92 were Negative Reward

By combining the five subjects into two—medium reward or higher and then everything else—we are left with 63 significant rewards and 289 instances of little, no, or negative rewards. Essentially, from September 2002 to August 2007, these low risk pitcher signings have truly worked out approximately one-sixth of the time (17.8 %).

The highest reward belonged to Chris Carpenter, who recovered from Tommy John surgery and had a then-career year in 2004 with the Cardinals. Jaret Wright had the second best reward during his 2004 season with the Atlanta Braves; incidentally, Wright also had the second worst VORP of low risk pitcher signings during his 2003 season with the Padres. Wright's 2003 VORP was -15.7, a distant second-to-last from Jose Lima's -31.6 in 2005.

VORP to WAR

Using the rule of thumb that 10 VORP runs equates to one win above replacement (WAR) allows us to quantify the results in a form more suitable in determining team contribution. Carpenter's 40.5 VORP equates to 4.05 wins; his production in saving runs relative to the amount a replacement level pitcher would surrender resulted in a contribution of about four wins. The numbers essentially stay the same as they are merely being scaled down but converting saved runs to wins helps in determining whether or not these moves are worth the risk.

Here are the low-risk pitchers accounting for two or more wins above replacement:

Chris Carpenter, 2004	4.05
Jaret Wright, 2004	3.99
Jeremy Guthrie, 2007	3.82
Paul Byrd, 2005	3.54
Takashi Saito, 2006	3.36
Jeff Suppan, 2003	3.25
David Bush, 2006	3.09
David Riske, 2007	2.77
Kenny Rogers, 2003	2.54
Russ Springer, 2007	2.51
John Thomson, 2003	2.51
Todd Jones, 2004	2.49
Tom Gordon, 2003	2.11
Darren Oliver, 2006	2.10
Mike Timlin, 2003	2.04
Steve Trachsel, 2007	2.03

Worth the Low Risk?

Now that the results are there it becomes necessary to utilize them in order to determine whether or not this is a sound strategy. By simply looking at the breakdown of reward-types shown earlier it can be deduced that, in this span of five years, just a small percentage of low-risk pitcher signings have provided a significant reward. This does not necessarily undermine the strategy, however, as it is important to analyze the results with the mindset of a prospective general manager unaware of whether or not his low risk pitcher is going to work out or not.

In determining whether or not the strategy is worth continuing the ideas of betting and probability come into play. The Cardinals gambled on Chris Carpenter and their strategy paid off immensely; however, the success of the strategy is dependent on how often it pays off overall, not how well certain select moves pay off. If you make five low risk pitcher signings and only one of them works out, regardless of how much that pitcher produced, the strategy does not look very sound; 80 percent of those signings were not worth it.

The key here is finding the sum total of salaries and wins above replacement amongst these pitchers and then calculating how much each win costs. With this group of 352, the salaries add up to 214.2 MM and they combine for 110.4 WAR. This results in a salary of approximately 1.94 MM per win. Generally speaking, the more reliable free agents will cost somewhere between 4MM – 7MM per win. Signing low risk pitchers appears to be a sound strategy due to the fact that a general manager is paying a significantly lesser amount of money per win. If the low risk pitchers had cost somewhere near the aforementioned range, if not more, then the strategy would not make sense; if you are going to end up paying a similar amount per win then it is a much safer investment to sign a pitcher who brings with him less of a question mark.

When it comes to these signings, though, it is important to remember that for every Chris Carpenter there are four Bruce Chens; it might make sense to take a flyer on a pitcher to fill out a rotation but building the majority of a rotation with these pitchers would not be a good usage of money.

Eric J. Seidman, Seidburns850@aol.com ♦

Submissions

Phil Birnbaum, Editor

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

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

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

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

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

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

W-L Records for Hitters

Tom Hanrahan

A pitcher's career is well-summarized by a W-L record, which is instantly understood by fans. But for hitters, there's no simple set of numbers that's as easily comprehensible. Here, the author tries his hand at creating a quick-and-dirty way to quickly compute the equivalent of a W-L record for batters.

This is not a knock-off of Win Shares, or some precocious attempt to 'improve' them. There is no math-heavy analysis herein that attempts to go where no analyst has gone before. This instead is an attempt at creating a simple yet clear way to sum up a batter's career performance; and to do so in a way that any casual baseball fan could recognize the statistic and say "he was (good, great, lousy, durable, superb-but-brief; pick your adjective)". And while we're at it, if such a statistic can be created, let's find one such that the achievements of both great position players AND great pitchers of baseball history can be expressed in a common language.

Stated another way, my primary goal is not accuracy; it is simplicity. There are more accurate (and complicated!) metrics available than the ones I will use here. This is about taking hitter records and, without making any adjustments for quality of play and live vs dead ball, etc., creating for batters a simple record like pitchers have – simple, yet just as accurate a summary of accomplishments.

The problem: hitters' records are very complex. We keep track, and indeed have kept track for decades, of batting average, hits, home runs, total bases, OBP, SLG, walks, runs, RBI, steals, and sundry combinations of these. There is no one agreed-upon metric that neatly sums up a batter's accomplishments; which makes it rather complicated to compare them. Now, of course there *are* good ways to compare them, and these statistically-oriented comparative methods have indeed proliferated as we all search for the Grand Unified Statistic (GUS). Again, I'm not attempting to publish a better GUS than TPR or VORP or WARP or WS. Each of these are very complex, which is both good (accuracy) and bad (your typical sportswriter has a cow every time a new one comes out). I'm going to put forth here a way to make a hitter's career stats make sense in one statistic that is very easy to calculate.

Pitching

First, consider pitchers' records (warning: long diversion). What stat best sums up a pitcher's career accomplishments? That's obvious, isn't it? We analysts whine a lot about measuring pitchers by wins and losses, but over a long period, a pitcher's W-L record is generally a very good indicator of what kind of career he had. Oh yes, it was easier to win 25 games a year in 1905 than in 2005, and certainly it's easier to win games with a good team behind you... but when it's all said and done, career W-L isn't half bad. It sure beats career ERA, which a) says nothing about career length, and b) has many more distortions associated with it; mostly timeframe and park. Not many people would claim that Ed Walsh, Addie Joss, Jack Pfeister, Joe Wood, and Jim Devlin were 5 of the greatest pitchers who ever lived; and yet they own the best 5 career earned run averages (try *that* as a trivia question!). On the other hand, if you wished to rank pitchers by most Wins, you get

1. Cy Young
2. Walter Johnson
- 3T Pete Alexander
- 3T Christy Mathewson
5. Pud Galvin (years pitched: 1875-1892)
6. Warren Spahn
7. Kid Nichols
8. Roger Clemens (active?)
9. Greg Maddux (active)

Not a bad list, but of course it misses lots of great pitchers who didn't get in a very full career.

Now, we could use winning percentage (with some minimum number of games pitched; baseball-reference.com uses 100 decisions); but is the following leaderboard of all-time best winning percentage very descriptive of the best pitchers?

1. Al Spalding (National Association, 1871-75)
2. Spud Chandler
3. Pedro Martinez (active)
- 4T Whitey Ford
- 4T Dan Foutz
6. Bob Caruthers
7. Don Gullett

Wow, that doesn't look all that helpful.

Maybe if we constructed a "Wins minus Losses" metric; that would be, basically, wins above average. The all-time leaderboard looks like this (now I'll drop Mr. Spalding, for those of you who justifiably question the legitimacy of the pre-1876 NA league):

1. Cy Young
2. Christy Mathewson
3. Roger Clemens
4. Pete Alexander
5. Lefty Grove

All right! I mean, there are still large imperfections; it's too heavy with pre-integration hurlers, and the inclusion of Matty and exclusion of the Big Train is obviously related to their teammates; but doesn't this describe five of the best pitching careers pretty well? Not bad for boiling it all down to one number. Cy Young, 195 games above average. Beat that, he says from the grave.

Where were we? Oh yes, hitters. If only hitters had one stat which describes their careers as well. And, what if it was in the same denomination as the pitcher's stat; wins and losses? Wouldn't that be cool? Well, once again, smart minds have created their versions of the GUS, often in terms of wins and losses; but these are all not exactly using numbers you can easily create when watching a game or even a whole season's worth of box scores. I am not, repeat *not*, trying to redo, undo, or any other 'do' on some fancy schmancy metric (nor do I mean to poo-poo them; Win Shares and WARP, among others, have contributed to our understanding!). But I *am* saying that it is easy, simple, a piece of cake, to *create a "won - loss" record for hitters* that is about as true, maybe even more so, to their value as pitchers' W-L marks are to them. Okay, buckle up, here we go.

Batter Wins

What is a "win" for hitters? Not for players in general; this is not measuring the value of defensive position or ability. Batters win games for their teams by generating runs. Hits, walks, power, speed, delivering in the clutch, all contribute to the overall goal of runs. Runs won games in 1905 when the ball was dead, in 1950 when stolen bases were a forgotten weapon, and in 1985 when Wade (on-base machine) Boggs coexisted with Vince (zoom!) Coleman, and Steve (bye bye) Balboni. Over the course of time, teams have averaged scoring about 4½ runs per game. This isn't always true – we had the high mound of the mid 60s and the lively era of the 30s – but if you look at a typical 20-year career of a great hitter, it is rare for a man to spend his whole career in an extreme 'live' or 'dead' era (Todd Helton in Coors is attempting to become the ultimate exception). I submit that runs scored plus RBI is a fair and very simple measure of career "wins" for hitters. Table 1 shows the hitters with the most career R+RBI. All tables show data through the 2007 season only.

Table 1 is the equivalent of the pitchers with the most wins. And, unlike the pitchers' list, there is good representation across almost every era; which is an indication that this measure may be more "fair" than using wins to rank pitchers. The next step will be converting R+RBI into "wins"; this will wait until later.

Table 1 – Career runs plus runs batted in

Aaron	4471	Yastrzemski	3660
Ruth	4391	F Robinson	3641
Bonds	4223	T Williams	3637
Cobb	4183	Murray	3544
Anson	4072	Winfield	3502
Mays	3965	Palmeiro	3498
Musial	3900	Rose	3477
Gehrig	3883	Wagner	3468
Ott	3719	Speaker	3411
Foxx	3673	Henderson	3410

The surprises in Table 1 might be

-- Cap Anson: for those unfamiliar with 19th century players, Anson played for 27 (!) seasons, was a top player for much of that time, and an above-league-average hitter until his very last year.

-- Mel Ott: not often thought of as one of the best hitters ever, the simple facts are that he was an all-star virtually his whole career, scored 1859 runs, and drove in more than he scored. Yes, he did benefit from the high-offense 1930s.

Batter Losses

What is batter “losses”? Teams lose games when they use up their allotted number of outs, 27, before scoring enough runs. Therefore, outs are the easy, simple equivalent of losses. Batting outs (at bats minus hits) are the primary source of all outs, but to be completely inclusive we would also have to count caught stealing, sac bunts, sac flies, GDP, and baserunning outs. There are sources for most but not all of these. For example, Baseball-reference.com has a list in the leaderboard section of career outs made (the top 1000 batters), which includes things like CS and GDP. However, I am trying to make this simple both to calculate and to find; since most everyone can look up a source that lists at-bats and hits, I will use “at bats minus hits” as a simple proxy for all outs. Across MLB history, about 25½ to 26 outs in a game are batting outs; the other 1-1½ come from the other sources. Table 2 shows the hitters who made the most outs (AB – H) in MLB history.

Rose	9797	Yount	7866
Aaron	8593	Biggio	7816
Yastrzemski	8569	B Robinson	7806
Ripken	8367	Mays	7598
Murray	8081	Aparicio	7553
Henderson	7906	Molitor	7516
Winfield	7893		

Table 2 is a list of men who played forever, while not posting superlative on-base skills (note that Cobb and Bonds are missing). Now we begin to see more “glove” men; guys like Luis Aparicio and Brooks Robinson played until they were very old, despite not being great hitters, because of their high defensive value.

W-L

So, how do we turn R+RBI and batter outs into wins and losses?

One simple method would be to assign one loss to every 25.5 or 26 outs (one game), and one win to every 9 R+RBI (league average runs per game). This overstates batters’ value, however; you can still win a game after making 25 to 27 outs, and 4½ runs hardly guarantee a win.

Another try: Pitchers average one decision (W or L) per 9 innings (27 outs), so really the total of W+L should relate to the amount of playing time. If “games” for hitters = outs divided by 27, wins then could be (R+RBI)/18, since on average it takes an extra 9 runs to create one more win (and there is an RBI for most runs scored); then, losses = games minus wins. This might be more representative of the actual numbers of games to be assigned to each batter. However, if done this way, the W and L for long-career pitchers will completely dwarf those of hitters. Is this “correct”? Not really. Pitchers’ records have fielders and hitters abilities included in them, which inflates their observed (W-L) value; no one really believes that Walter Johnson literally won over 400 games by himself; while virtually all hitters have some defensive value that is not included in their statistics accumulated while at the plate.

I would like to find a method by which we can compare hitters and pitchers, so their records “look” similar by inspection. So, I tweaked the coefficients to turn R+RBI into wins and outs into losses until I was satisfied with the results. The only qualification was that an average hitter ought to have an equal number of wins and losses. Therefore, the ratio of coefficients for (R+RBI) and outs must stay very close to 9:26.

After lots of fiddling, I used Wins = (R+RBI)/12 and Losses = outs/34. By this method, there have been 12 “300 win” hitters since 1900, and Pete Rose leads in losses with 288. In comparison, 16 pitchers have won 300 games since 1900, while Nolan Ryan has the most losses with 292.

Are the divisor coefficients 12 and 34 arbitrary? Yes they are. If I used 9 and 26, Hank Aaron’s lifetime W-L record would be 497 and 331. If I used 18 and 51, his record would be 248 and 168. The comparisons come out better with the coefficients I used; that’s my only justification.

Table 3 shows this method’s W-L record for the 22 hitters with the most wins, next to the pitchers ranked by most wins. At the bottom of the table are a few batters and pitchers with the most losses, who are not among the wins leaders. In this table, the top pitchers have more wins than the top batters, but most of this is the distortion among the pre-1920 players. Since the live ball era began, the totals among these groups are fairly equal.

Is there much doubt that The Hammer, Henry Aaron, has a breadth of offensive accomplishment that is virtually unparalleled? Who else would you choose as the batter with the most ‘wins’?

Earlier, I showed a listing that ranked the top pitchers by most wins minus losses. Table 4 (see next page) is a complete ranking of batters next to pitchers by this metric, again with a few other relevant players at the bottom.

Any surprise that the Babe is the leader in this table, with 203 more wins than losses? I didn’t think so. And this doesn’t even count his pitching record!

I emphasize again, these metrics have definite limitations. Lou Gehrig, for example, played when runs were plentiful, greatly boosting his W-L record. Willie Mays had lots of other value, being a great centerfielder. Bert Blyleven had lots of other value, being obscured here by poor teammates. My gut feel is that if you ignore the obvious missing value of defense, that the hitters’ W-L records are closer to their “true” value than the pitchers’ records are. So I say, limitations, schmitations. With reasonable accuracy, we have translated the panoply of achievements of the greatest hitters in MLB history into a statistic that is generally used (and useful!) for pitchers.

Matching the Bats With the Arms

If hitters had W-L records like pitchers do, we could compare them to each other. Let’s do it! Some of this is objective, and some is not. I will try to put similar W-L records together, but also try to match career shapes, contemporaries, perception by peers/fans/media, men who were over (or under) –rated and the reason why, and personality or circumstantial factors. So, here we go:

Hank Aaron 373-253
Denton “Cy” Young 511-316 (or Warren Spahn, 363-245)

Most batting wins, and most pitching wins. Aaron played in the tougher National League post-integration, and in a fairly low-scoring era; these numbers actually *understate* his accomplishments. While Cy Young may not have reached as great a peak of performance as other pitchers, he was as utterly consistent as Aaron, and his career accomplishments are simply astonishing; take away 99 wins and he’s still the all-time leader. Spahn, while not really Aaron’s equal, might have been considered the greatest hurler of the last 80 years if we could pretend that he went 4-F during WWII, and then blew away the wartime competition, boosting his record to something like 420-270. I also like that he and Aaron were teammates.

Barry Bonds 352-203
Roger Clemens 354-184

Two men who have solid arguments to be the greatest ever. Seven MVP awards, versus seven Cy Young awards. Both of course are also linked as accused of steroid use. They also have taken heat for perceived poor post-season performance, although the arguments in each case are overdrawn.

Aaron	373-253	Young	511-316
Ruth	366-163	Johnson	417-279
Bonds	352-203	Mathewson	373-188
Cobb	349-213	Alexander	373-208
Anson	339-202	P Galvin	364-310
Mays	330-223	Spahn	363-245
Musial	325-216	Nichols	361-208
Gehrig	324-155	Clemens	354-184
Ott	310-194	Maddux	347-214
Foxx	306-161	T Keefe	342-225
Yastrzemski	305-252	Carlton	329-244
F Robinson	303-208	J Clarkson	328-178
T Williams	303-149	E Plank	326-194
Murray	295-238	Ryan	324-292
Winfield	292-232	Sutton	324-256
Palmeiro	292-219	P Niekro	318-274
Rose	290-288	G Perry	314-265
Wagner	289-206	Seaver	311-205
Speaker	284-187	C Radbourn	309-195
Henderson	284-233	M Welch	307-210
Ripken	279-246	Glavine	303-199
Simmons	278-172	Grove	300-140
...		Wynn	300-244
Yount	256-231	...	
B Robinson	217-230	J Powell	245-254
Biggio	254-230	E Rixey	266-251
Aparicio	177-222	Blyleven	287-250

Ty Cobb 349-213
Willie Mays 330-223
Walter Johnson 417-279

Two of these were among the original 5 Hall of Famers. All three of them are much better than their W-L record: the Big Train's teammates didn't give him much help; Cobb began play when teams scored only 3½ runs per game, and he still scored more runs than any player until 2001; and Mays played in the low-scoring tough 60s, while of course playing an amazing center field. I almost placed Willie in the group above, with his godson from the same team. These two groups are very similar. I'd like to match Cobb with someone of similar persona; if such a person existed...

Babe Ruth 366-163
Nobody

Seriously. Maybe if you add in Lefty Grove's minor league numbers, and also pretend he was an all-star right fielder when not pitching?

Cap Anson 339-202
Kid Nichols 361-208

Anson played in weak leagues; however, early in his career the whole "season" was often only 65 games!

Stan Musial 325-216
Grover "Pete" Alexander 373-208

Both men are often forgotten today in discussions of the "greatest ever". When the all-century team was named in 1999, Pete was left off, and Stan the Man was only added by a special panel when the fans snubbed him. Musial was overshadowed in his day by the Splendid Splinter -- and Alexander by the Big Train -- but both of these guys were absolutely superb.

Lou Gehrig 324-155
Christy Mathewson 373-188

1. Heroes. Great men of character.
2. World Series stars.
3. Died young.
4. And, though it's heresy to say it, overrated by the general public. Lou's superhuman RBI totals were partly a byproduct of team and era, and Matty pitched for great teams in the deadball days.

Mel Ott 310-194
Eddie Plank 324-194

Ott played when runs were scored freely, and Plank pitched for fine teams in the early 1900s. Consensus opinion is that neither was quite as good as the numbers show.

Table 4 – Most 'wins minus losses' by hitters, and most pitcher wins minus losses (AA, 1871-75, not included)

Ruth	366-163	Young	511-316
Gehrig	324-155	Mathewson	373-188
T Williams	303-149	Clemens	354-184
Bonds	352-203	Alexander	373-208
Foxx	306-161	Grove	300-141
Anson	339-202	Nichols	361-208
Cobb	349-213	Clarkson	328-178
Ott	310-194	W Johnson	417-279
Aaron	373-253	R Johnson	284-150
Delahanty	255-144	Maddux	347-214
Hornsby	264-154	E Plank	326-194
Musial	325-216	Ford	236-130
DiMaggio	244-136	Spahn	363-245
Mays	330-223	T Keefe	342-225
Simmons	278-172	P Martinez	209-93
Brothers	235-130	C Radbourn	309-195
M Ramirez	246-143		
Mantle	266-167		
Duffy	238-140		
...		...	
Schmidt	258-180	Koufax	165-87
O Smith	166-204	B Gibson	251-174
Santo	206-175	Glavine	303-199
B Hamilton	202-120		

Jimmy Foxx **306-161**
Randy Johnson **284-150**

The Big Unit has great W-L stats, but he has been lost a little among his superb peers. Ditto Foxx, who was unfortunate to play 1B at the same time as Gehrig.

Carl Yastrzemski **305-252**
Pete Rose **290-288**
Cal Ripken **279-246**
Nolan Ryan **324-292**

Contemporaries, extremely long careers, amazing plusses but some minuses thrown in. All of them were practically icons. Rose has the most wins of anyone who played a significant amount of time in the infield, while the Iron Man is not far behind.

Ted Williams **303-149**
Lefty Grove **300-141**

Hotheads. A few missed years for each; two wars for Ted, stuck in the minors for Robert Moses Grove. Both could be called the best ever if you asked "who do I want to bat (pitch) when I really need a hit (out)?", and their W-L percentages bear this out. I love how close their W-L records are.

Joe DiMaggio **244-136**
Sandy Koufax **165-87**
Pedro Martinez **209-93** so far

Each of these studs are considered better than their W-L record, with good reason; the pitchers had dominating ERAs and multiple Cy Youngs. Joltin' Joe has deserved hardware as well. All were the best at what they did in their primes. Two had very short careers, while Pedro has not yet had a full one.

Honus Wagner **289-206**
Nobody

Wagner, stunningly forgotten and underrated until Craig Wright and Bill James in the 1980s awakened many to his great play, was a Gold Glove caliber shortstop who would often score and/or drive in 100 runs in a year, while the league ERA was about 2.50. Maybe Tom Seaver would be his equal if he was a star pinch hitter on the side. Walter Johnson might be his closest comparison, but the W-L records themselves do not show it.

Frank Robinson **303-208**
Greg Maddux **347-214** and still going (or **Warren Spahn, 363-245**)

Consistent, solid, durable, great players.

Eddie Murray **295-238**
Dave Winfield **292-232**
Gaylord Perry **314-265**

One notch below the tandem above them.

Rafael Palmeiro	292-219
Don Sutton	324-256

Played forever, both in favorable environments, all the while being very good but never “great”. Gaylord is also a good match for Raffy, in the sense of matching alleged cheaters.

Tris Speaker	284-197
Mickey Mantle	266-167
Rogers Hornsby	264-154
Mike Schmidt	258-180
Tom Seaver	311-205

All truly great, but underrated by their W-L marks: Seaver’s lack of good teammates, Speaker and Mantle for their fine CF play, Schmidt for his gold glove at 3B, and Hornsby because he played 2B (note that Hornsby had the highest rank in table 4 of any middle infielder).

Al Simmons	278-172
Bob Gibson	251-174

Gibson is much more famous, partly because of his World Series heroics, as well as his 1968 gem of a season, but bucketfoot Al put up some really fine numbers. Neither lasted long enough to garner coveted career marks.

Ron Santo	206-175
Bert Blyleven	287-250

The truly underrated club. Both ignored by the Hall of Fame so far, Santo for lacking the One Big Number to hang his hat on, Bert for having poor taste in teammates. Craig Biggio (254-230) might fit in here also.

Ed Delahanty	255-144
Billy Hamilton	202-120
Whitey Ford	236-106

Two 19th century hitters (and teammates) with stats puffed up by the high-tech scoring environment. Matched up with a fine hurler who had the good fortune to be on the 1950s Yankees.

Dan Brouthers	235-130
John Clarkson	328-178
Tim Keefe	342-225

Who?? 1880s great slugger and 1880s great hurlers. All were dominant in their day.

Gary Sheffield	260-179
Sammy Sosa	262-188
Jeff Bagwell	254-161
Frank Thomas	262-163
Tom Glavine	303-199

Contemporary monster bats, born within a few years of each other; each with nice numbers, aided by the modern lively game. Glavine benefited from having fine teammates. All of these, if they retired today, would have numbers to get into the Hall of Fame, but not without some detractors.

So there you have it, my friends. It all comes down to this:

- Runs scored, RBI, and outs made are a very good equivalent for pitcher's W and L.
- It is easy to turn R, RBI, and outs into equivalent W and L for batters. And it makes sense.

And when we do that, there are some obvious pairings between great hitters and great pitchers. Some of my favorites are:

Lou Gehrig = Christy Mathewson
Barry Bonds = Roger Clemens
Ted Williams = Lefty Grove
Ty Cobb = Walter Johnson
Stan Musial = Pete Alexander
Hank Aaron = Cy Young
Pete Rose = Nolan Ryan
Ron Santo = Bert Blyleven
Frank Robinson = Greg Maddux
Tris Speaker = Tom Seaver.

Tom Hanrahan, Han60Man@aol.com ♦

Get Your Own Copy

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

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

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

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