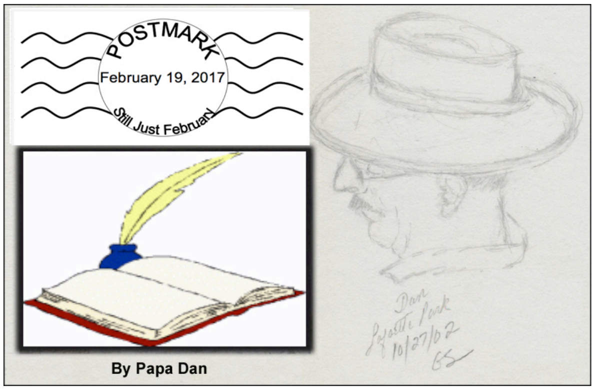
Getting Baseball Wrong: By the Numbers

By Dan Sapone



"Using wrong measurements has resulted in bad decisions on contracts playing time, trades, and draft picks." "It's led to picking the wrong players for MVP, Cy Young, and Rookie of the Year and even obvious stuff like the Hall of Fame." It drives conversations about teams and players right off the cliff."

— Keith Law, from "Smart Baseball"

Understanding Baseball Math

Baseball writer Keith Law has written an unusual book about the way we evaluate baseball players and teams — something we've all been doing since we were kids. He makes a case that we've been getting it wrong throughout the history of baseball. Let's consider his observations.

When I saw his book, "Smart Baseball," on the library's 'New Books' shelf, I thought it was about something else — you know, how to PLAY baseball smarter. I thought it might be useful for talking about baseball with my grandchildren. But it's about something else that all fans do: comparing players and teams during the season and, a tougher task, comparing today's players and teams with those of the past. We use statistics; and it affects a lot of baseball judgments.

First, how DO we evaluate and compare hitters?
Most of us have known these rules since we were kids.
1. We evaluate hitters by their batting averages, runs batted in, and slugging percentage.
• **Batting average (BA):** the number of base hits divided by the number of official times at bat — example: a batter who gets 30 base hits in 100 at-bats has a batting average of .300. In this calculation, all hits are treated as equal.
• **Runs batted in (RBIs):** the number of times a batter causes a runner to score because he gets a base hit or walks in a run — examples: a batter who gets a base hit driving in runners on second and third is credited with two RBIs; a batter who walks with the bases loaded is credited with one RBI. (BTW, a batter who hits a ball "booted" by a fielder causing a run to a score on the error does NOT receive an RBI.)
• **Slugging percentage (SLG):** the total bases a batter reaches divided by the number of official at-bats. Regardless of the number runs that are produced, a double is counted as two total bases; a homer is four. So, this statistic claims to measure the extent to which a hitter is a "power hitter." A batter who has 100 official at-bats and hits 20 singles, eight doubles, and two homers has a slugging percentage of .440.

These numbers have been used for making decisions on trading or acquiring players (or baseball cards), selecting a Most Valuable Player (or other batting awards), and setting salaries, literally for centuries. So, how well does that work?

Let's See

Every year the major leagues give their "**Batting Champion**" award to the player with the highest batting average. The assumption is that this is the single statistic that identifies the best hitter. In 2015, the Miami Dolphins' Dee Gordon led the National League with a batting average of .333. So, he was the 2015 NL Batting Champion. That was easy.

→ **BUT,** Bryce Harper of the Washington Nationals led the league in almost everything else that year: doubles, homers (42 to Gordon's 4), walks (124 to Gordon's 25), and runs scored. Harper beat Gordon in RBIs (99 to 46) and Slugging Percentage (.649 to .418). So, who was the best hitter? If the offensive goal of a baseball team is scoring runs to win games, which of them would you want on your team? Oh, by the way, Harper was also the National League's **MVP**, but, Gordon was the NL **Batting Champion**. So, tell me — what did that mean exactly?

What about RBIs?

Maybe the number of runs that score when a batter gets base hits is a better gauge of a hitter's ability. This statistic has long been valued in voting for MVP candidates, as it was (above) in 2015 in the National League. Despite the important role of the RBI stat in baseball reporting, many writers point out that it is one of the individual stats that depends almost entirely on the achievements of other players — specifically, how many players are on base when a batter gets a hit. Examples abound, but comparing two famous careers serves at least to consider the importance of the RBI statistic:

- As we all remember (with varying degrees of admiration for off-the-field reasons), Barry Bonds was the major leagues' all-time career leader in home runs with 762 compared with Henry Aaron's 755. To fill out the picture, judging them ONLY by their actual offensive statistics, Bonds hit those home runs in 300 few games than Aaron, got on base more often (.444 to .374), and hit for more power (i.e., a higher slugging percentage .607 to .555), was walked almost twice as many times (2558 to 1402), and was the National League MVP seven times to Aaron's one. SO, what about RBIs? With all those numbers what would you guess?

Aaron had 301 more career RBIs than Bonds. Why? During his most productive seasons, Bonds batted third in a Giants' lineup in which two of the three players who batted ahead of him were not so good at getting on base. This fact is what Keith Law refers to as one of the "stupid manager tricks" that make RBIs such a weak indicator of a hitter's ability. Turns out that Bonds has the distinction of hitting 450 **solo** home runs in his career. So, Bonds' lower RBI total has more to do with the lineup that batted ahead of him than anything about his own hitting prowess. So, what does a player's RBI total tell us about a player's hitting ability? Not so much.

2**. What about pitchers?**

We evaluate pitchers by the number of "Wins," Saves," and Earned-Run Average (ERA). How does that work out? All our lives, starting pitchers have been labeled by the number of "Wins" they are credited with. There are other pitching stats, but they all take a back seat to the "Won-Loss" record. We all want our pitching rotation to include "20-game-winners." This statistic is simply dumb. While **team** victories matter more than any other **team** statistic across a 162-game season, author Keith Law asserts "the idea of a single player earning full credit for a win or blame for a loss exposes deep ignorance of how the game actually plays out on the field."

I grew up in an era when a manager hoped, and a pitcher intended, that he would pitch a "complete game." Today — nope. Today, managers typically keep a starter on the mound for 100 pitches, certainly not more than 120. Example: in April of 2016, Dodgers' rookie Ross Stripling was pitching a no-hit, perfect game in the 8th inning against the Giants when, after his 100th pitch, he was yanked for a reliever. The Giants wound up winning 3-2 on Brandon Crawford's 10th-inning homer. Contrast that with a game I listened to on my transistor radio back in July of 1963 when 25-year-old Juan Marichal battled 42-year-old Warren Spahn on the mound into the 16th inning, both pitching a shutout, when Willie Mays broke up the party with a 16th-inning homer to win for the Giants 1-0. Even though Spahn had to pitch himself out of a bases-loaded situation in the 14th inning, both managers did the obvious thing — left their best pitcher in the game when they were pitching a shutout in a game they wanted to win. SO, what does it mean? In the first example, Ross Stripling pitched 7+ innings of perfect baseball and it didn't even show up in his Won-Loss record. In the second example, Warren Spahn finished the season with a 23-7 Won-Loss record; and this game, arguably one of the best pitching performances in history, was just one of his seven losses. What did that statistic mean?

Not much.

Conversely, a "Win" doesn't always mean that pitcher even pitched well. Example: in the 2000 season, Russ Ortiz pitched 6 2/3 innings and gave up **ten** runs; but he got the win because his team scored 16 runs that day. Truth is, a "Win" or a "Loss," as a pitching statistic, depends on the work of many other players — sometimes not because of, but **IN SPITE OF** the pitching performance. Then, when you factor in the role of relief pitchers who preserve a "Win" for a starter and the very real contributions of the defense and offense of a pitcher's team, the "Win-Loss" record is pretty useless in evaluating a pitcher's ability. AND that doesn't even consider the number of good pitchers who played long careers for bad teams.

So, let's agree that the "Won-Loss" record is, at best, misleading.

So, what about "Saves?" Rule 10.20 determines that a relief pitcher earns a save when he:

- finishes a game his team wins and he is not the winning pitcher; and any one of these apply:
  — enters the game with a lead of no more than three runs and pitches at least one inning
  — enters the game with the potential tying run on base or at bat or on deck
  — pitches effectively for three innings.

If he meets these criteria, he earns a "Save" — whether he pitches well or poorly; OR he can pitch very well and NOT earn a Save under the wrong conditions. Keith Law tells us: in 2015, there were 114 appearances when a relief pitcher pitched at least three innings and gave up zero runs; but only NINE of those appearances earned the pitcher a "Save." Compare that to the thirteen saves that year earned by pitchers who allowed two runs in one inning of work, but protected a lead and finished the game. Pitch four scoreless innings in a loss — no save. Give up two runs in the last inning — save. SO, what does the "Save" statistic mean? It's all about factors that have little to do with the performance of the pitcher. BUT, for relief pitchers, it's a key contributor to salaries, trades, and recognition.

What about "Earned Run Average (ERA)?"
This stat is a staple on baseball cards and on-screen graphics along with the Win-Loss record. It is the number of earned runs allowed per nine innings pitched. Earned runs are, of course, those that result from base hits and not caused by errors or passed balls. But, it does not account for the ability of outfielders to hold a hitter to a single, throw out runners on the bases, or the subjective work of official scorers to distinguish hits from errors. So, aside from strikeouts and walks, the act of getting a batter out or preventing runs from scoring is shared by others.

While ERA has limited value for evaluating starting pitchers, it is even more misleading for relievers. When a starting pitcher gives up a single, a double, and an intentional walk, he is responsible for those runners. When the relief pitcher replaces him in the 7th inning with the bases loaded and gives up two more singles before retiring the side, all three of the resulting runs are charged to the starting pitcher, none to the reliever. While getting outs is the primary task of any pitcher, it is too easy for a relief pitcher to come in during a tough situation, pitch badly, and leave with his ERA intact. The resulting stat is not an accurate reflection of who did what or the effect their performances had on the outcome of the game. We could cite many examples of the assignment of earned runs in a variety of situations and none would be entirely satisfying. In general, ERA is another case of baseball math that depends more on the ***sequence*** of events and the contribution of others than the quality of individual performances.

So, what's better?
In recent years, something called "Sabermetrics" has begun to creep into the mainstream of baseball writing and even management decision-making (remember the movie "Moneyball?"). Serious statisticians are now employed by more MLB teams every year and they are looking at some different numbers. The book "Smart Baseball" asserts that "On-Base Percentage (OBP) is the most complete of all basic hitting stats, because it includes everything a hitter does" and it excludes many factors that are not specific to the hitter's individual performance.

OBP = (Hits + walks + times hit by pitch
<u>divided by</u>
(At bats + walks + times hit by pitch + sacrifice flies)

The logic is simple: take the number of times a hitter gets on base and divide it by the total number of times he came up to the plate. Using the 2015 season of Bryce Harper (mentioned above) as an example, this method would have given him an OBP of .460 (46%). If that statistic had been used in 2015, Harper would have easily been the Batting Champion instead of Dee Gordon who simply had a higher batting average by .003.

Again, statistical examples are numerous — let's return to our previous example:
In 1987, the National League MVP was Andre Dawson. Dawson hit 49 home runs for the Cubs along with a batting average of .287 and 137 RBIs, and a slugging percentage of .568. OK.

However, Keith Law asserts that the real MVP that year should easily have been Tony Gwynn. While Dawson had more home runs and RBIs, Gwynn dramatically overshadowed Dawson in singles, doubles, triples, walks and — here's the key: **he had an On-Base-Percentage (OBP) of .447 compared to Dawson's .328.** Another way to look at it, given that they had a nearly identical number of plate appearances, Dawson made 70 more outs than Gwynn that season.

SO, by their own individual contributions, which of them increased their team's ability to score more runs? If that is the most important objective, it seems that our traditional baseball math may have chosen the wrong MVP in 1987. Using OBP could have suggested another choice.

One small refinement has emerged — the "triple slash" stat: batting average/on-base percentage/slugging percentage. In that case, here is the snapshot of those 1987 performances.

	BA	OBP	SLG
Dawson:	.287	.328	.568
Gwynn:	.370	.447	.511

What's the effect on the recognition of pitchers?
The most prestigious single-season recognition for pitchers is the Cy Young award. As a quiz, let's take a look at the details of the Cy Young decision in a year you might not remember, 1990, and see if we can decipher how that decision was made. Here were the top candidates:

	W	L	ERA	IP	Runs
Bob Welch (Oak):	27	6	2.95	238	90 (78 earned)
Dave Stewart (Oak):	22	11	2.56	267	84 (76 earned)
Roger Clemens (Bos)	21	6	1.93	228	59 (49 earned)

Do you remember who won the award? Without looking, who would get your vote? (Hint question: does one statistic dominate Cy Young award voting?)

Bob Welch won the award. Why? Of the three contenders, Welch gave up more runs, had the worst ERA, and . . . well, he had the best Won-Loss record — the one stat that depends the most on the work of others. Was he the best pitcher that year? Heck, Keith Law observes that he wasn't even the best pitcher on his own team! That shiny "Win" total is clearly the only reason.

The logic is inescapable, but not very useful — you have to look at a number of metrics to make such a judgment about pitchers, not just one. The goal be would to select players that can add as many wins to the team as possible, within a budget.

For those who want to dig deeper into the problem of judging pitchers — and for those who enjoy long descriptions of statistical tradeoffs — ***and you know who you are*** — I suggest you read Keith Law's book (try chapter 14). He actually has an approach to this problem — he calls it "Wins Above Replacement (WAR)," which focusses on an elusive metric: "Runs Prevented." For me, I skipped to the last chapter where he identifies the two main issues:
1. Baseball provides a lot of data points that are fundamentally interdependent — that is, so few baseball outcomes can be traced to the performance of a single player. It's a team game (Duh).
2. So much of what drives the way baseball is played and viewed is more cultural than statistical. Players, managers, team owners, and fans do what they do, and like what they like about the game, for reasons they grew up with since the first time each of us was captivated by "baseball fever" when we were kids.

So, when we try to dissect the statistics of the game to understand it better, we find out that baseball is more than it seems . . . and less.

Go Giants!