Comparing Players Using Player Won-Lost Records

Comparing Players Using Player Won-Lost Records

The core calculation of Baseball Player won-lost records is pWins and pLosses. These are calculated play by play such that the total player decisions by a team are exactly three per game, 2 pWins and 1 pLoss for a winning team, 1 pWin and 2 pLosses for a losing team (teams earn 1.5 pWins and 1.5 pLosses in a tie game). Because the number of pWins and pLosses are known for certain at the team level, by construction, pWins and pLosses are a fairly objective calculation. The exact pWins that I calculate for a specific player may not be precisely correct, but any mis-allocation of pWins (and/or pLosses) will be confined to teammates within a specific game.

Given a set of pWins and pLosses, I then calculate a set of context-neutral Player won-lost records, eWins and eLosses. Oversimplifying, eWins for a given event are calculated by taking the average value of the pWins associated with the event over the course of the season. The precise calculation of pWins and eWins are described in more detail in an essay on my website and in my first book (*Player Won-Lost Records in Baseball: Measuring Performance in Context*, McFarland, 2017).

Baseball Player won-lost records, pWins and pLosses are, in my humble opinion, a perfect measure of player value. When context and the effects of teammates are controlled for, eWins and eLosses can also, in my opinion, serve as an excellent starting point for measuring player talent. But in order to make comparisons between players, it is necessary to make one additional calculation: positional averages.

- **Positional Averages**

In constructing Player won-lost records, all events are measured against expected, or average, results across the event. Because of this, fielding won-lost records are constructed such that aggregate winning percentages are 0.500 for all fielding positions. Hence, one can say that a shortstop with a defensive winning percentage of 0.475 was a below-average defensive shortstop and a first baseman with a defensive winning percentage of 0.510 was an above-average defensive first baseman, but there is no basis for determining which of these two players was a better fielder – the below-average fielder at the more difficult position or the above-average fielder at the easier position.

From an offensive perspective, batting won-lost records are constructed by comparing across all batters, not simply batters who share the same fielding position. In the National League, this means that offensive comparisons include pitcher hitting, so that, on average, non-pitcher hitters will be slightly above average in the National League, while, of course, because of the DH rule, the average non-pitcher hitter will define the average in the American League.

These are, in fact, two sides of the same coin. There is a clear negative correlation between the average offensive production at a defensive position and the importance and/or difficulty associated with playing that position. That is, players at the toughest defensive positions tend to be weaker hitters than players at easier defensive positions.

Bill James used this observation to define what he called the Defensive Spectrum:

1B – LF – RF – 3B – CF – 2B – SS – C

Positions get more difficult/valuable defensively moving left to right (e.g., shortstop is a more defensive position than second base) while offensive production increases moving right to left (e.g., on average, first basemen tend to out-hit left fielders).

When comparing, for example, a left fielder to a shortstop, one must somehow balance the fact that left fielders are expected to hit better than shortstops against the fact that shortstops are, on average, better defensive players than left fielders.

There are three ways to do this:

(1) One can adjust offensive Player Won-Lost records based on the defensive position of the player,

(2) One can adjust defensive Player Won-Lost records based on the defensive position of the player, or

(3) One can adjust the baseline against which players are measured.

The problem with both (1) and (2) is that either of these approaches would be difficult, if not impossible, to incorporate within a system that tied directly to actual team wins and losses as is my goal here.

By process of elimination, then, I believe that the best choice is (3), measuring players against different baselines based on the position(s) which they played, what I call "positional averages". I described the calculation of positional averages in great detail in Chapter 3 of my first book, *Player Won-Lost Records in Baseball: Measuring Performance in Context* (McFarland, 2017).

In discussions of Player won-lost records with other people, particularly in discussions at the Hall of Merit at Baseballthinkfactory.org, there are significant differences of opinion as to the proper way to calculate and use positional averages. My hope is for Player won-lost records to be as widely accepted as possible. While I think it is important to take player position into account in evaluating players, the specific calculation of positional averages is not central to the calculation of Player won-lost records. It is also very important to me that Player won-lost records be understood as a set of numbers which can be used, interpreted, and analyzed differently by different people as those people see fit; not a single inscrutable number.

To facilitate this, I have modified the Player won-lost records on my website, so that positional averages can be selected by the user. This section of this essay begins by explaining the mechanics of how one can choose one's positional averages and then looks in more detail at the specific options, what they mean, and how one might think about how to choose one's positional averages.

- Selecting Positional Averages

On any page for which positional averages will be displayed or used in calculation – player pages, team pages, leader boards, etc. – the following row(s) will appear near the top of the page:

Positional Average (Weighted Avg of Options): 0.500 _ (one-year) _ (nine-year) _ (Long-Run) _

Positional Average for Pitchers (Wgted Avg of Options): 0.500 _ vs. Different Values for Starting vs. Relief Pitchers: (Empirical) _ (Based on Pitchers who Do Both) _

Go

Starting with the second row first, there are three ways to calculate positional averages for pitchers. First, one can treat all pitchers the same and use a constant positional average of 0.500. Alternately, one can assign different positional averages to starting pitchers versus relief pitchers. This can be done in two possible ways. The first is what I call "empirical" positional averages. That is, calculate the overall winning percentage of all starting pitchers and use that as the positional average for starting pitchers and do the same for relief pitchers. Alternately, one can narrow one's focus to only those pitchers who pitched as both starting pitchers and relief pitchers within the season(s) of interest.

One can enter any numbers one desires in these three boxes and the positional average for pitchers will be calculated as a weighted average of these three options using the weights chosen. For example, entering the numbers 1, 2, and 3, for "0.500", "Empirical", and "Based on Pitchers who Do Both", respectively, would produce weights of 1/6 for 0.500 (1 divided by 1+2+3), 2/6 (one-third) for "Empirical", and 3/6 (i.e., 50%) for the final option.

The row for "Positional Average for Pitchers" may not appear on a web page if the player(s) for whom positional averages are being counted never pitched.

The first row applies to all positional averages, including pitchers. For non-pitchers, positional averages are calculated empirically, based purely on the players who played the position of interest. That is, the basis for calculating positional averages for second basemen is to focus exclusively on second basemen. The positional average for fielding decisions is 0.500 at all positions in all seasons by definition. The choices here are applied only to offensive player decisions.

The first option is to simply use a positional average of 0.500 for all players, regardless of position. This is the one option here which I would be inclined to say is "wrong" although I think a case could be made for including 0.500 within one's weighted average as a way of "regressing" positional averages if desired.

The other three options all involve calculating an empirical positional average for offensive performance by position over the time period specified. That is, "one-year" calculates positional average a single year: the positional average for second basemen in 1970 is based entirely on the offensive performance of second basemen in 1970 (relative to the overall offensive performance in 1970). The "nine-year" positional average is calculated over a nine-year period centered on the season of interest: i.e., the "nine-year" positional average for second basemen in 1970 is calculated based on the offensive performance of second basemen from 1966 through 1974 (relative to overall offensive performance in these seasons). The "long-run" positional average is calculated across all seasons for which I have calculated Player won-lost records. At the time of my writing this, this is the 98 seasons from 1921 through 2018.

As with the pitcher numbers, enter any numbers desired in the four boxes here and these will be converted to weights which will be used to construct weighted positional averages. Some examples of possible weighting options and their effect on player value are discussed later in this essay.

- Positional Averages over Time

The next table summarizes positional averages by position across all seasons for which I have calculated Player won-lost records (1921 – 2018). The numbers here are (context-neutral, teammate-adjusted) offensive player winning percentages in non-DH leagues. The next-to-last row shows what I call a "DH-league adjustment", the difference between offensive winning percentages for position players in the NL versus the AL since 1973. The last row, then, shows the positional average for designated hitters (for those league-seasons for which the DH rule was in effect).

This table excludes fielding won-lost records, for which the average is exactly 0.500 at every position in every season by construction. This table also excludes the defensive contributions of pitchers which will be discussed later.

The first column shows the long-run positional averages across all seasons for which I have calculated Player won-lost records. The second column shows positional averages for the most recent season for which I have calculated Player won-lost records. The final two columns then show the minimum and maximum single-season positional averages by season. Obviously, the minimums (and maximums) at different positions occurred in different seasons.

	1921 - 2018	2018	Minimum	Maximum
Pitcher (Offense)	0.353	0.276	0.276	0.408
Catcher	0.491	0.483	0.475	0.508
First Base	0.531	0.522	0.504	0.547
Second Base	0.494	0.502	0.470	0.516
Third Base	0.509	0.518	0.480	0.525
Shortstop	0.484	0.507	0.456	0.507
Left Field	0.525	0.521	0.507	0.545
Center Field	0.516	0.508	0.500	0.530
Right Field	0.526	0.521	0.511	0.543
Pinch Hitter	0.478	0.476	0.459	0.502
Pinch Runner	0.500	0.496	0.411	0.578
DH-Adjustment	0.009	0.010	0.008	0.010
Designated Hitter	0.520	0.524	0.505	0.529

The long-run averages are generally consistent with the defensive spectrum as I outlined it earlier. Focusing on the eight non-pitching, fielding positions, the highest long-run positional average is at first base (0.531) followed relatively closely by the two corner outfield positions. The lowest positional averages are second base (0.494), catcher (0.491), and shortstop (0.484).

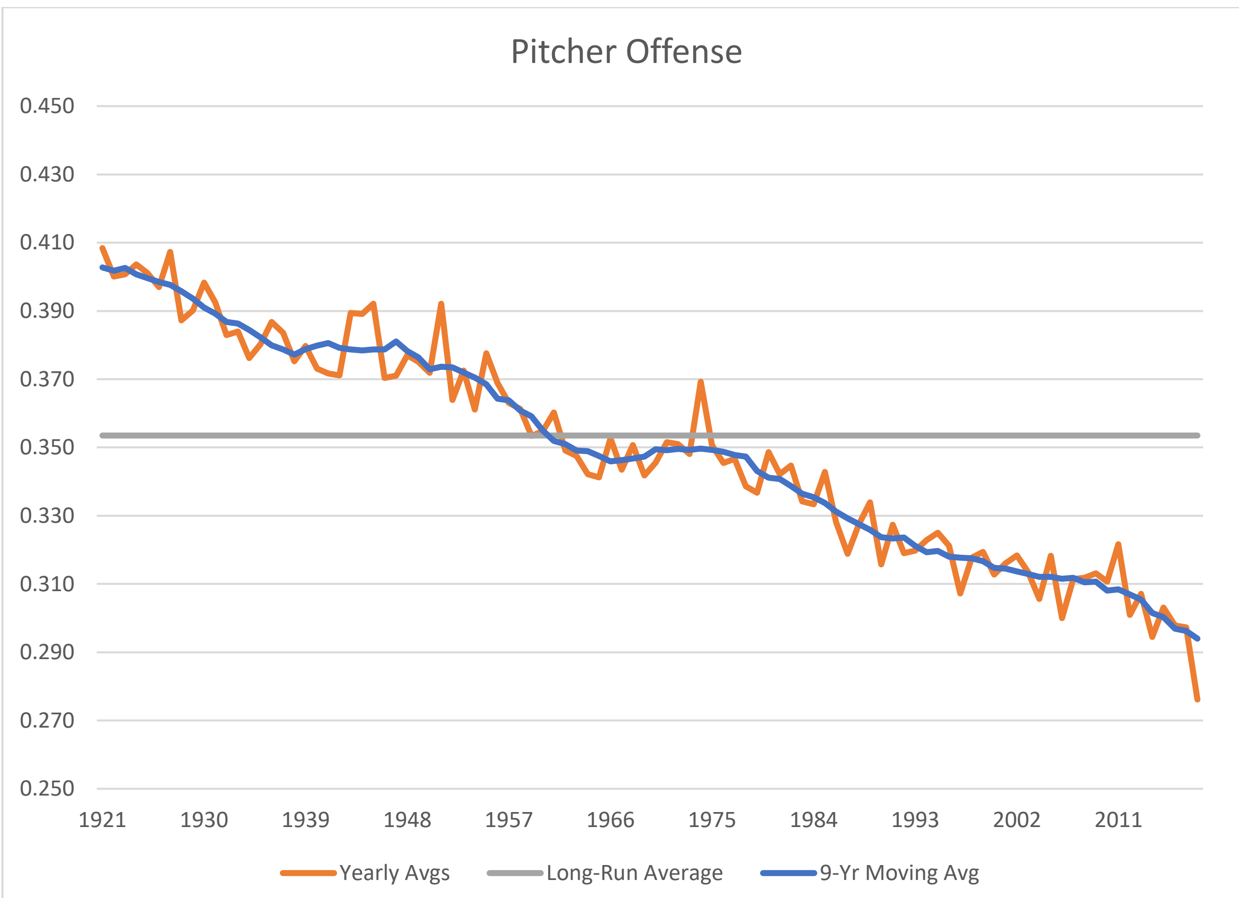
In 2018, however, the positional average at second base (0.502) was slightly lower than at shortstop (0.507). It is also interesting that, except for catcher (0.483 in 2018, 0.491 long-run), the 2018 positional averages tended to be more clustered, with first base and all three outfield positions lower than their long-run averages (but all still above 0.500) but the positional averages for second base, third base, and shortstop all higher than their long-run counterparts.

The next several pages take a closer look at the numbers by position. Rather than overwhelm you with a set of tables filled with eye-glazing numbers, let's break out some graphs.

- Pitcher Offense

The most obvious example of the need to allow positional averages to change over time is pitcher batting. In 1921, major-league pitchers batted a combined .212/.250/.274 which translated into a Player won-lost winning percentage of 0.406. In 2018, major-league pitchers batted a combined .115/.144/.149, which translated into a Player won-lost winning percentage of 0.275. I have calculated Player won-lost records for 98 seasons, 1921 through 2018. The highest positional average for pitcher offense was in the first of these seasons, 1921. The lowest positional average for pitcher offense was in the last of these seasons, 2018.

Here is a graph of the positional average for pitcher offense for all 98 seasons for which I have calculated Player won-lost records. The solid gray horizontal line is the average across all 98 seasons. The orange line is the year-by-year positional average. The blue line is a 9-year moving positional average for the season, the four seasons before, and the four seasons after.



It’s not a perfect downward trend, but it’s pretty close. Major-league pitchers have gotten uniformly worse at hitting for (at least) the last 100 years.

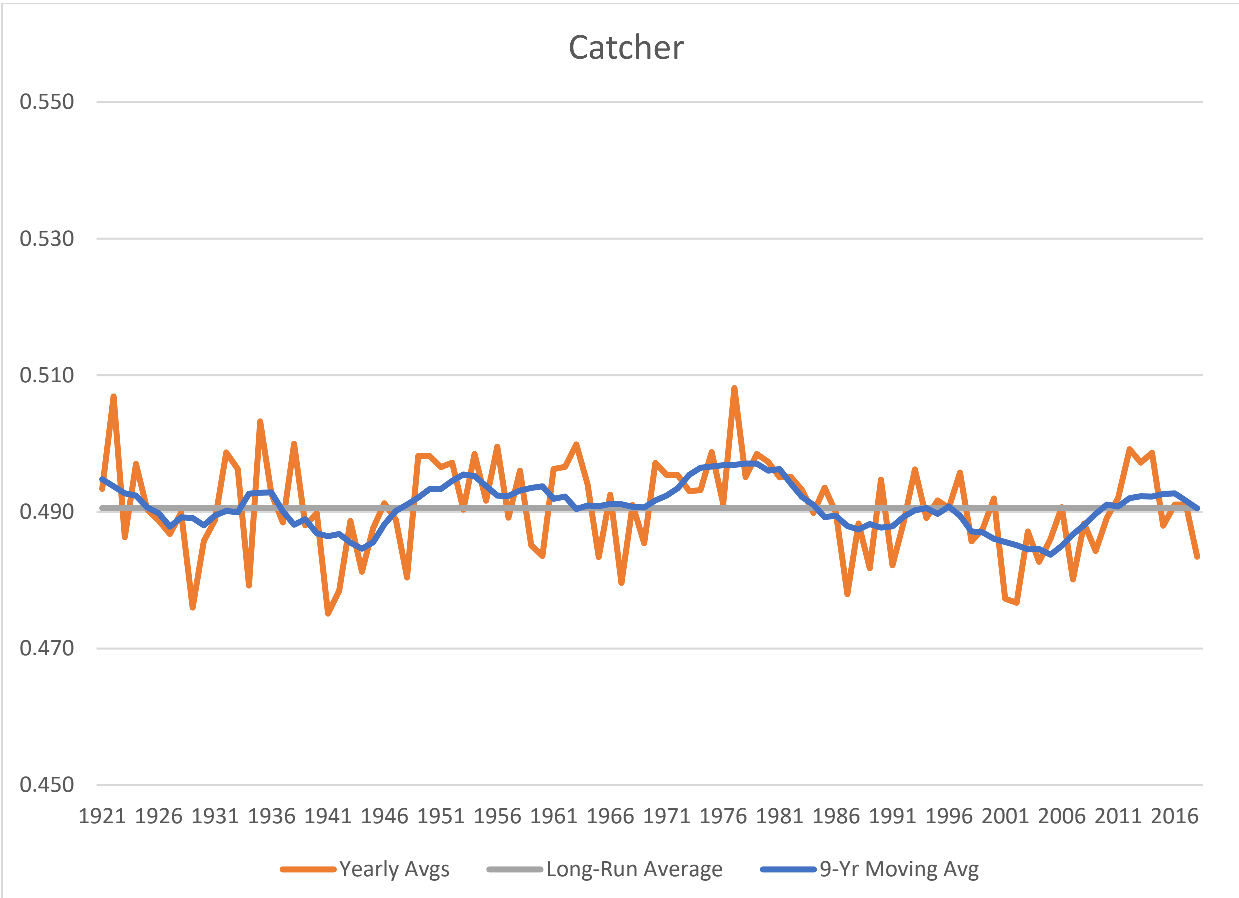
There was a bit of a spike during World War II (the 3-year plateau just under 0.390 is 1943-45) and there have been a few one-year spikes that I can’t really explain – 1951, 1974, 2011.

I have seen a couple of explanations for this trend. Bill James (and I think others) has speculated that pitcher hitting is a measure of the quality of play in major-league baseball. The premise is that, because pitchers are not selected for their hitting, the average quality of pitcher hitting is likely to remain constant over time. The fact that pitcher hitting has declined relative to overall offense can then be taken as evidence that the quality of major-league baseball has improved dramatically over the past 98 years. The spike in the above graph during World War II, when the quality of play in major-league baseball declined for a few years, is consistent with this hypothesis.

The other argument which I have seen posited is that the introduction of the DH rule – and its expansion to lower levels of baseball – has led to major-league pitchers being less experienced at hitting. There is a bit of a counter-factual underlying this theory. That is, it’s hard to say how much the trend since the introduction of the DH rule differs from what that trend might have looked like in the absence of such a rule. Just eyeballing the data, the decline in pitcher offense from, say, 1973 (0.347) to 2017 (0.296), a decline of approximately 0.0011 per year, is not meaningfully different from (and, in fact, slightly smaller than) the decline from, say, 1921 (0.406) to 1959 (0.352), which works out to 0.0014 per year. On the other hand, one could look at the data here and hypothesize that the negative trend in pitcher offense had actually stopped by 1959 or so. The positional average for pitcher offense was 0.352 in 1959 and 0.347 in 1973, which is really not very different. So, it could be that the decline through the 1950s is the result of improving quality of major-league baseball while the decline since the mid-1970s is the result of the designated hitter rule.

Regardless, this is the clearest example of why one might not want to use a simple 98-year average to calculate positional averages for pitcher offense. The implication would be that the vast majority of pitchers in the 1920s and 1930s were above average while virtually all pitchers over the past two to three decades have been below average. That would seem to distort what I think most people are trying to accomplish in measuring “average”.

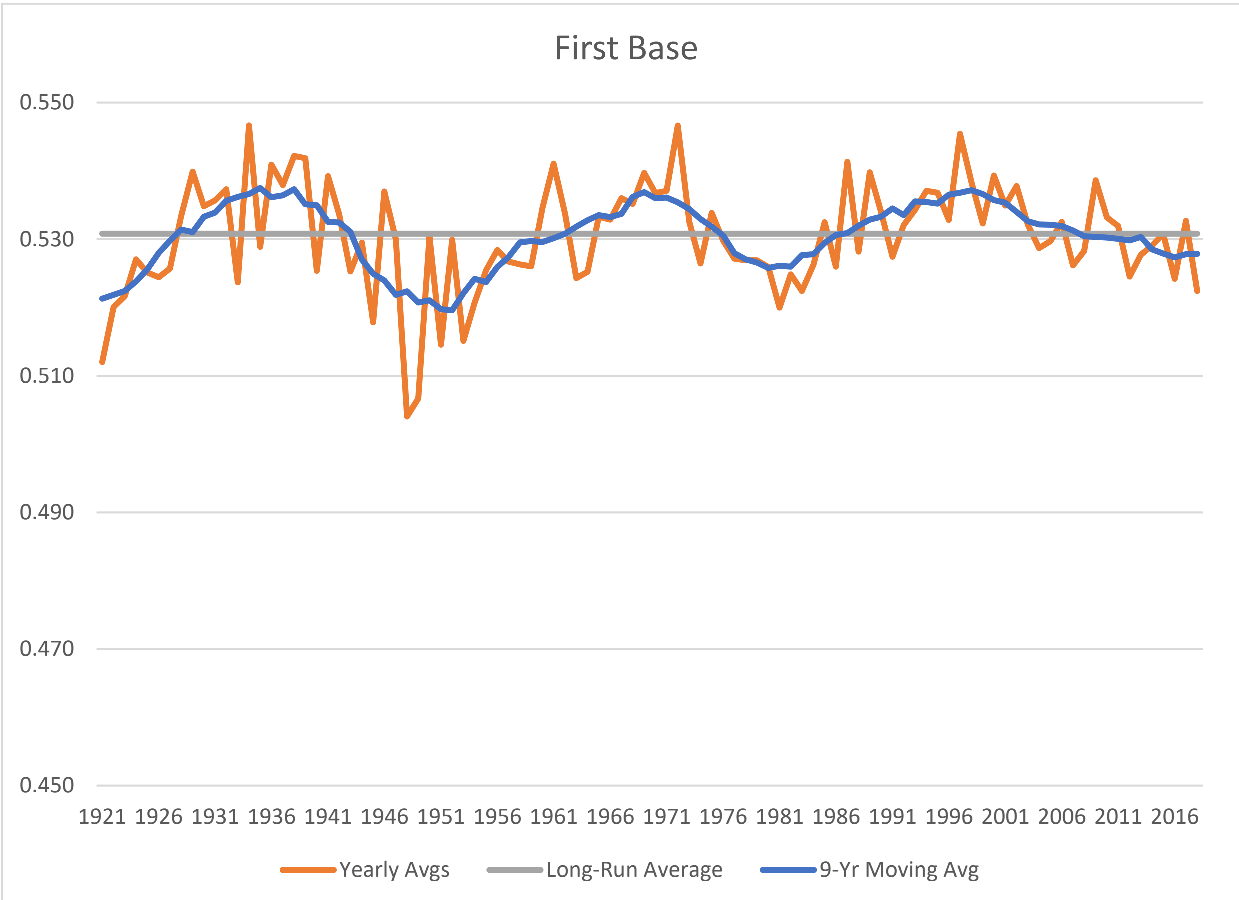
- Catcher



One quick note of explanation. The numbers here are for non-DH leagues. This is necessary because there were no DH leagues prior to 1973. But because of this, the combined positional average for non-pitchers is over 0.500. If you want to re-center things to 0.500, subtract approximately 0.009 (i.e., just less than one percent). So, in this case, the average positional average for catchers (the gray line) is 0.490, which translates to 0.481 in a DH league.

Certainly, there is nothing like the trend we saw in pitcher offense. In general, I don't really see a trend at all. Catcher positional averages tended to be above the historical average from about 1949 (0.497) through 1985 (0.493) although there were a number of down years within that time period. The recent three-year spike is 2012 – 2014, when positional averages were 0.499, 0.497, and 0.498, respectively. The average since 1986 (33 seasons) has been 0.488, which is virtually identical to the 98-year average of 0.490.

- First Base



First base has tended to have the highest positional average of any position through history. I have put the positional graphs here over the same scale, 0.450 to 0.550, so that, hopefully, you can visualize the positional averages relative to each other. There are two exceptions to this, the first and last of these. The first, pitcher offense, is on a lower scale because positional averages for pitcher offense have all been below 0.450. The last, pinch runners, I'll explain when I get to it.

Anyway, a fair bit of variation, but no obvious trends. The results were much more erratic before the late 1950s. Positional averages were stronger for first basemen in the late 1920s and 1930s – the era of Lou Gehrig, Jimmie Foxx, Hank Greenberg, et al (average of 0.537 from 1928 – 1939). Positional averages tended to be below historical norms from 1948 through 1959 (average of 0.521 for these 12 seasons).

Using single-season positional averages, as I have done in the past, helps Gil Hodges look better in Player won-lost records, for example, as he gets credit for being the best first baseman of the 1950s.

There was something of a glut of quality first basemen from 1987 through 2002 (average of 0.536) with the positional average peaking at 0.545 in 1997.

Since 2002, the positional average for first basemen has trended downward, although "trended" is not exactly the right word. Basically, the positional average dropped from 0.538 in 2002 to 0.532 in 2003 and held fairly steady – with some year-to-year variation – through 2011 (0.531). It has then hovered around an average of 0.527 since 2012.

One thing I want to emphasize is that the numbers here are offense-only. The positional average for fielding is exactly 0.500 at every position in every season by construction. One possible explanation for some of the longer-term variation here (and at other positions) could be shifts in the fielding standards which are considered acceptable for a first baseman. One somewhat odd result I observed here is that the positional average for first basemen tends to be higher in more extreme run-scoring environments. The 1920s and 1930s, for example, had very high levels of run-scoring as did, of course, the mid-to-late 1990s and early 2000s, and both of these time periods saw above-average positional averages for first basemen. But, interestingly, there was also a spike in the positional average for first basemen from 1965 through 1972 (0.537) which was the time period with the lowest run scoring of the past 98 years.

Overall, from 1921 through 2018, the average runs scored per 27 outs was around 4.77 (that is a simple average of the 98 annual averages). One can calculate a measure of how "extreme" run scoring is in a season by taking the absolute value of the difference between run scoring (per 27 outs) in a particular season and 4.77. The correlation between that measure and first base positional average is 0.304 which is moderately strong.

What might this mean? Thinking about it, I have a hypothesis. In extreme run-scoring environments, it is most important for teams to have a slugger. In low-scoring environments, it's valuable to have a slugger because home runs become more valuable in an environment where it is difficult to string together hits. The surest way to score runs in a low-scoring environment is via the home run. In high-scoring environments, it's important to have as many good hitters as possible so that teams are, perhaps, more willing to find room for defensively-challenged sluggers. In both of these situations, the increase in the average offensive output of first basemen could logically be expected to coincide with a decrease in the average fielding ability of first basemen.

Seven-time Gold Glove winner Vic Power's last season was 1965; seven-time Gold Glove winner Bill White saw his playing time drop sharply in 1967 and 1968. In the high-run-scoring 1990s, Rafael Palmeiro infamously won a Gold Glove at first base in 1999 despite playing only 28 games at the position. Obviously, Palmeiro was a poor choice, but the fact that he won suggests that there may not have been a particularly good choice.

But there is no direct mechanism for recognizing that the average defensive first baseman in 1999 was below average. That very sentence makes no sense.

The lowest single-season positional average for first base was in 1948, 0.504, and first-base positional averages were generally below their long-run average from 1948 through 1959 with an average value of 0.521. This roughly corresponds to the career of Gil Hodges who, using single-season positional averages, had a career positional average of 0.518.

In contrast, the twenty years from 1992 through 2011 saw a positional average for first base of 0.534. The latter half of this time period corresponds to the career of Todd Helton who, using single-season positional averages, had a career positional average of 0.526, 0.008 higher than Gil Hodges's career positional average despite both players being almost exclusively first basemen who played their entire careers in non-DH leagues. Now, a difference of 0.008 doesn't sound like much, but Helton had just over 450 player decisions (pWins plus pLosses) in his career: 0.008 times 450 translates into approximately 3.6 fewer pWins over positional average (pWOPA) for Helton relative to Hodges because of positional average. I calculate Gil Hodges to have earned 18.9 pWOPA in his career versus only 6.8 pWOPA for

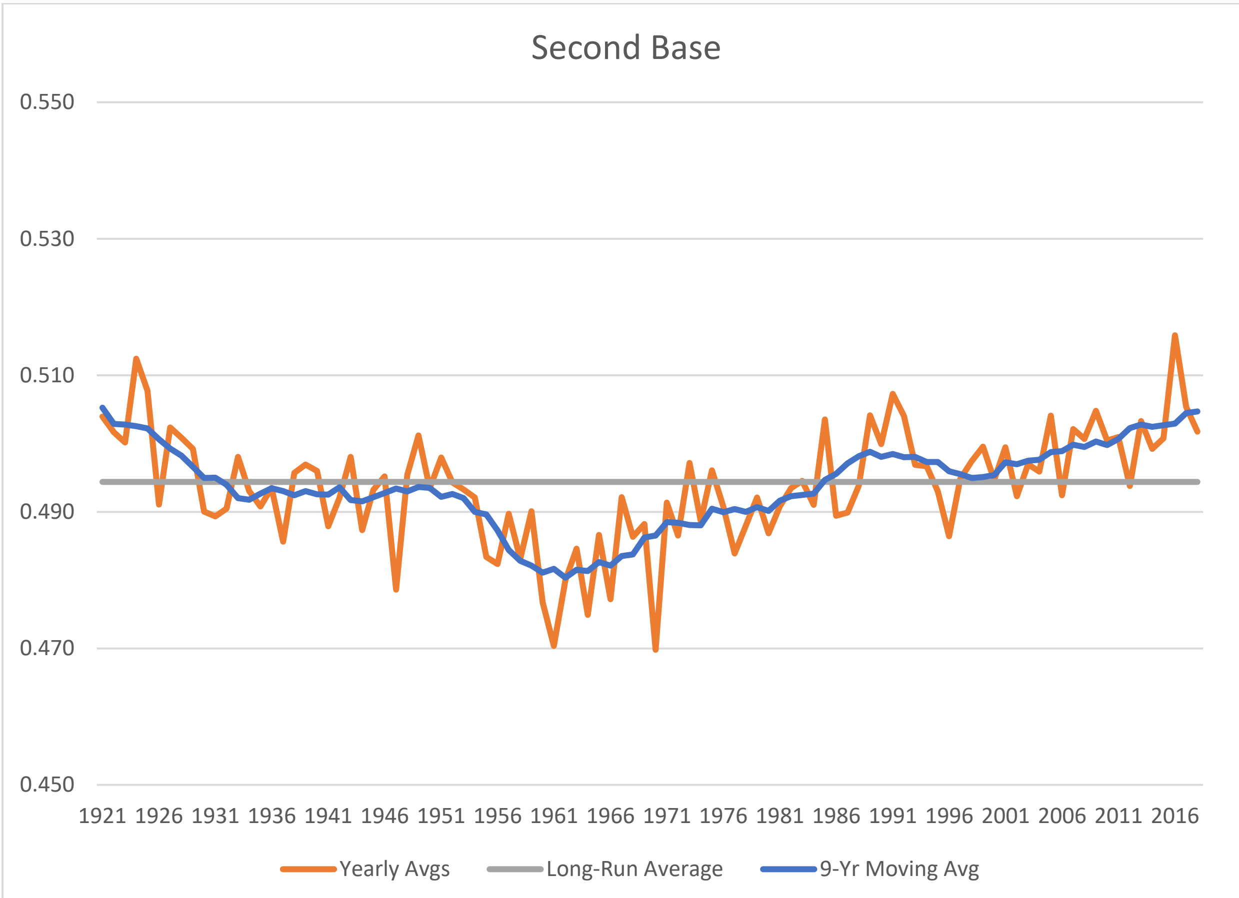
Todd Helton, a difference of about 12 pWins – of which perhaps 30% can be attributed to the difference in the two players' positional averages (the remainder is due to context – Helton beats Hodges in career eWOPA, 13.3 to 13.0).

Is that unfair to Todd Helton? Not necessarily.

Gil Hodges had a reputation as an excellent defensive first baseman (he won the first three Gold Gloves awarded at ages 33, 34, and 35). And Player won-lost records recognize him as an excellent defensive first baseman with a career fielding winning percentage of 0.539 with 4.3 more fielding wins than fielding losses. Todd Helton was also an excellent defensive first baseman (he also won three Gold Gloves, at ages 27, 28, and 30) with a career fielding winning percentage of 0.530 with 4.4 more fielding wins than fielding losses. Those seem directly comparable, but are they? Not necessarily. Gil Hodges' 0.539 winning percentage is measured against the first basemen of his time. But what if there were more good-fielding first basemen in Gil Hodges' time? With relatively few exceptions (e.g., Ted Kluszewski), the first basemen of Hodges' day were smaller, more athletic men than the first basemen of Todd Helton's day (e.g., Jason Giambi, Jim Thome, Ryan Howard). These seem like two sides of the same coin. Doesn't the fact that first basemen didn't hit as well in the 1950s as they did in the 2000s suggest that poor-fielding sluggers were more tolerated in the latter time period (e.g., Jason Giambi, Jim Thome, Ryan Howard) than in the former time period?

Vic Power had a 1,626-game major-league career, most of it playing first base (1,307 games) with a lifetime batting line of .284/.315/.411 and a career-high in home runs of 16. Player won-lost records (a) recognize that Vic Power was an excellent defensive first baseman (career winning percentage of 0.545), but (b) do not think that was enough to make up for his glove, even with the lower positional averages of the 1950s (Power had -7.3 career pWOPA). Power's bat brings down the positional average against which Hodges's batting is compared. But Power's glove brings **up** the positional average against which Hodges's fielding is compared. But because of the nature of the way Player won-lost records are calculated, the former is done (at least somewhat) explicitly, because we can observe how first basemen hit relative to other positions; while the latter is implicit, because the average fielding winning percentage will always be 0.500 at every position in every season. But allowing the former, the offensive positional average, to vary across seasons helps us to capture the implicit differences in the latter.

- Second Base

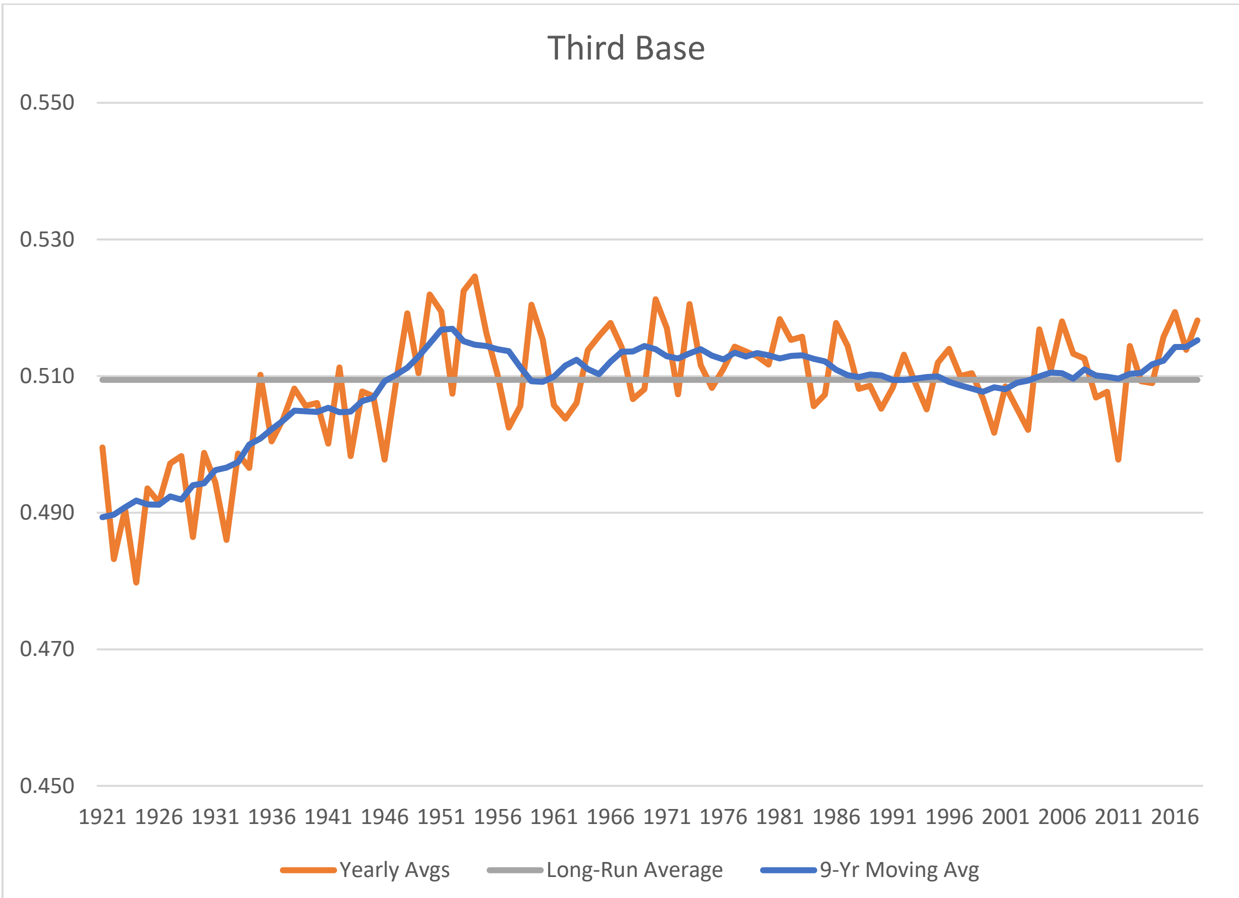


Positional averages for second base trended down fairly consistently over the first 40 years for which I have data. This is consistent with an observation by Bill James – which he incorporated into his calculation of Win Shares – that second base was a "hitter's position" into the 1920s and 1930s before switching places with third base on the defensive spectrum. In his Win Shares, James has second base establish its modern place in the defensive spectrum by 1946. The data here suggest that the transition of second base to a defense-first position may have lasted a decade or two longer than that. As with pitcher offense, this shift over time is something that I think one ought to incorporate into the positional averages used over time.

But interestingly, that only tells half the story for second base. After bottoming out between 1961 and 1970 (the two bottom-most points in the above graph), the positional average for second base increased from 0.469 in 1970 to 0.491 in 1971 and more or less stayed at this higher level into the late 1990s. And positional averages for second base have then trended upward since the late 1990s.

The positional average in 2016 was 0.516, which was the highest positional average for second base over the past 98 years, and the 2018 positional average, 0.502, was identical to the positional average in 1922 (it was 0.504 in 1921, 0.500 in 1923). To be honest, I don't know what to make of this. But like the trend of the first 40 years of the graph, it definitely seems like something that ought to be taken into consideration in calculating positional averages for second base.

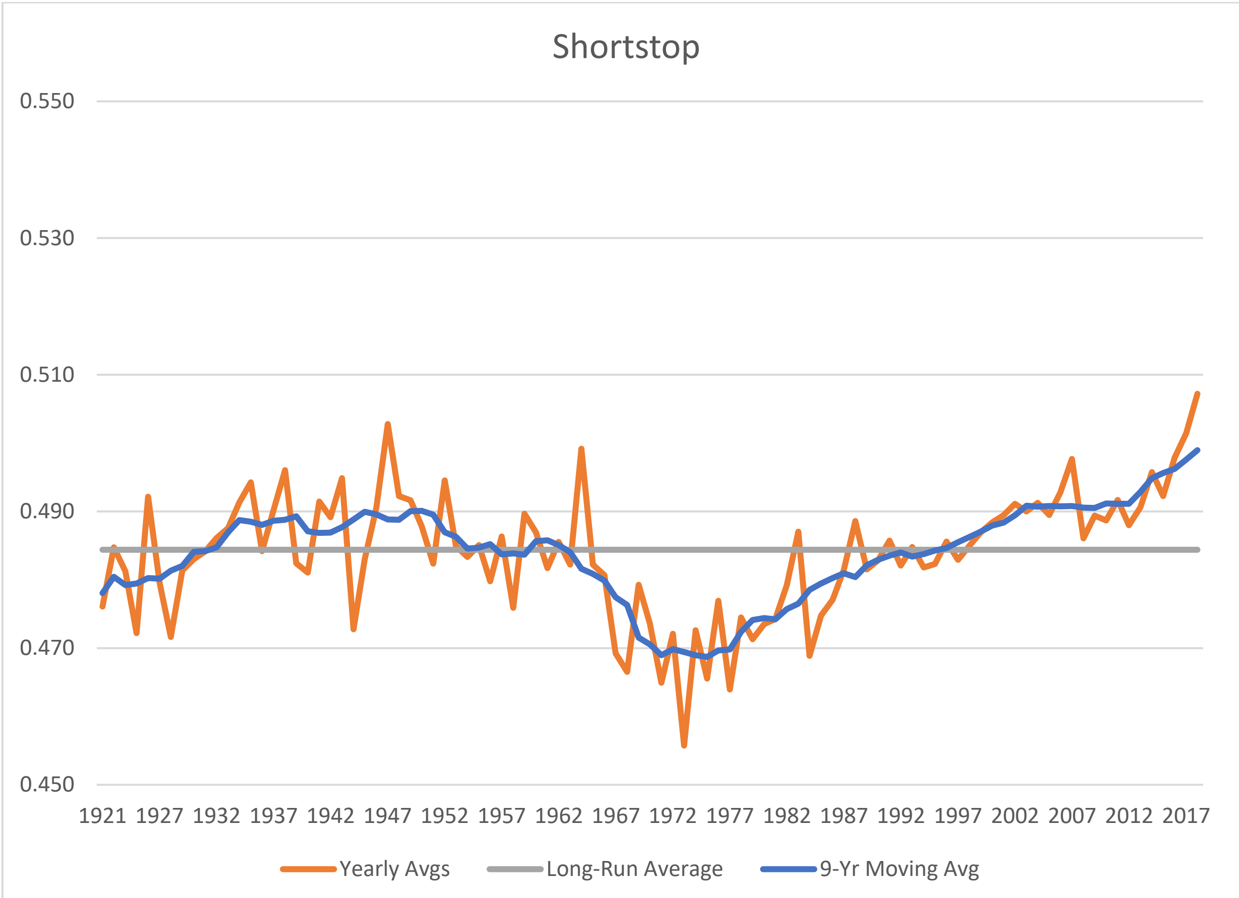
- Third Base



Positional averages at third base trended up from 1921 through 1948 or so. This is the mirror image of the results seen with second base which Bill James discussed in his *Win Shares* book. Since then, however, the results for third base have held fairly steady.

From 1948 through 2018, the positional average for third base has averaged 0.512. If you incorporate the DH-adjustment which I discussed earlier, this would work out to about 0.503 in a DH league. In other words, the average third baseman has tended to be approximately a league-average hitter over the past 70 years or so.

- Shortstop

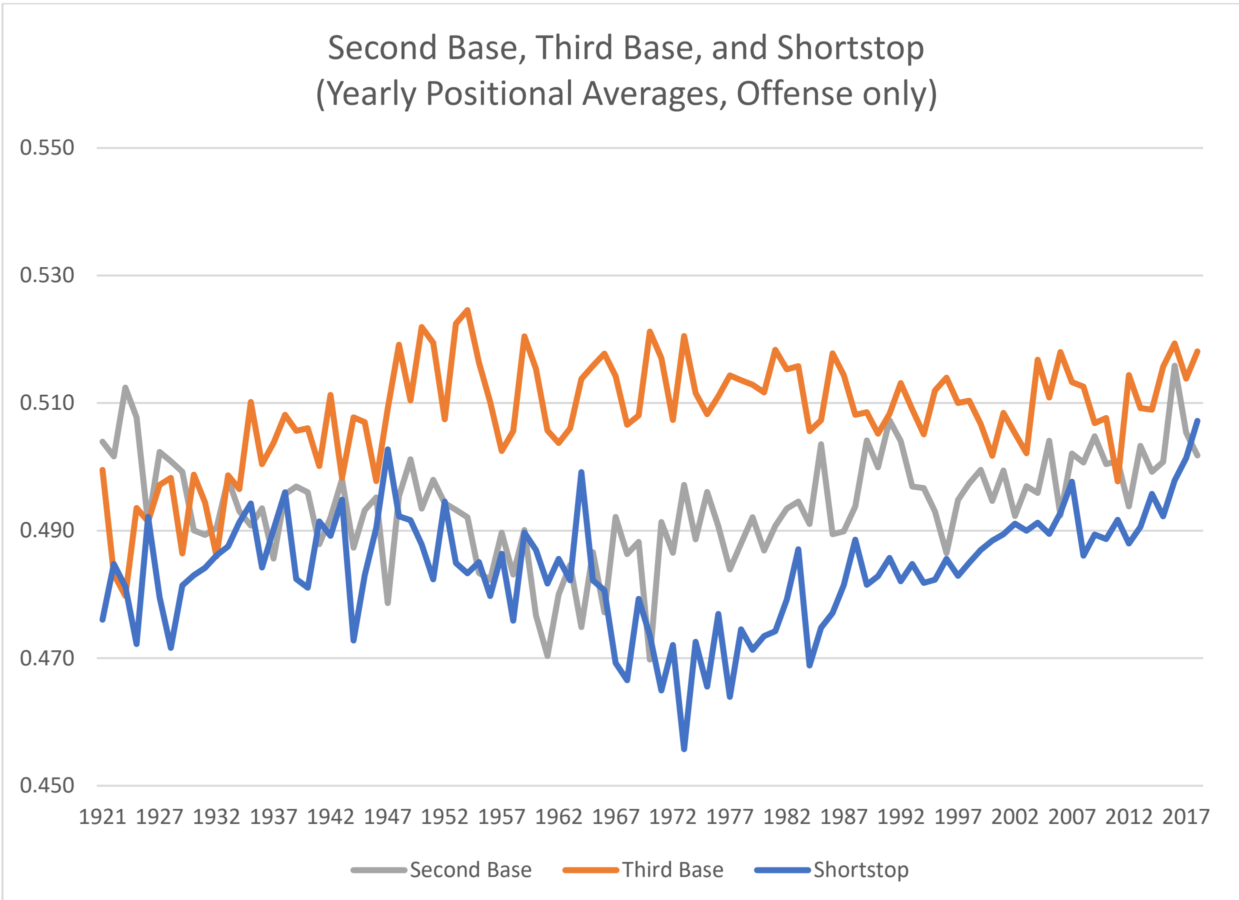


The numbers here are interesting. The results were fairly stable through 1965 (0.482) or 1966 (0.481) before dropping sharply – 0.469 in 1967 – and remaining low through 1984 (0.469). Results then bounced back up to the long-run average from 1985 through 1997 (as well as 1983) before trending upward since then. This trend has been especially sharp since 2015 (0.492). Since then, the positional average for shortstop has been 0.498, 0.501, and 0.507, respectively.

The results here do not really surprise me. I think it is fairly well known that the shortstop position was historically weak offensively through the 1970s, for example. And the more recent trend coincides with the emergence of several higher-offense shortstops, starting with the 1980s trio of Robin Yount, Cal Ripken, and Alan Trammell, continuing in the mid-to-late 1990s with Alex Rodriguez, Nomar Garciaparra, and Derek Jeter, and continuing through this century with Miguel Tejada, Jimmy Rollins, Troy Tulowitzki, Francisco Lindor, Carlos Correa, and others.

One hypothesis I have heard is that the increasing use of Astroturf in the late 1960s and through the 1970s increased the defensive requirements for shortstops – ground balls reach the infielders quicker on turf than on grass. As in other cases, I think it is more reasonable to say that the expectations for what constitutes an average shortstop have changed over time as opposed to asserting that modern shortstops are mostly above average while virtually all of the shortstops in the 1970s were below average.

- Infielders



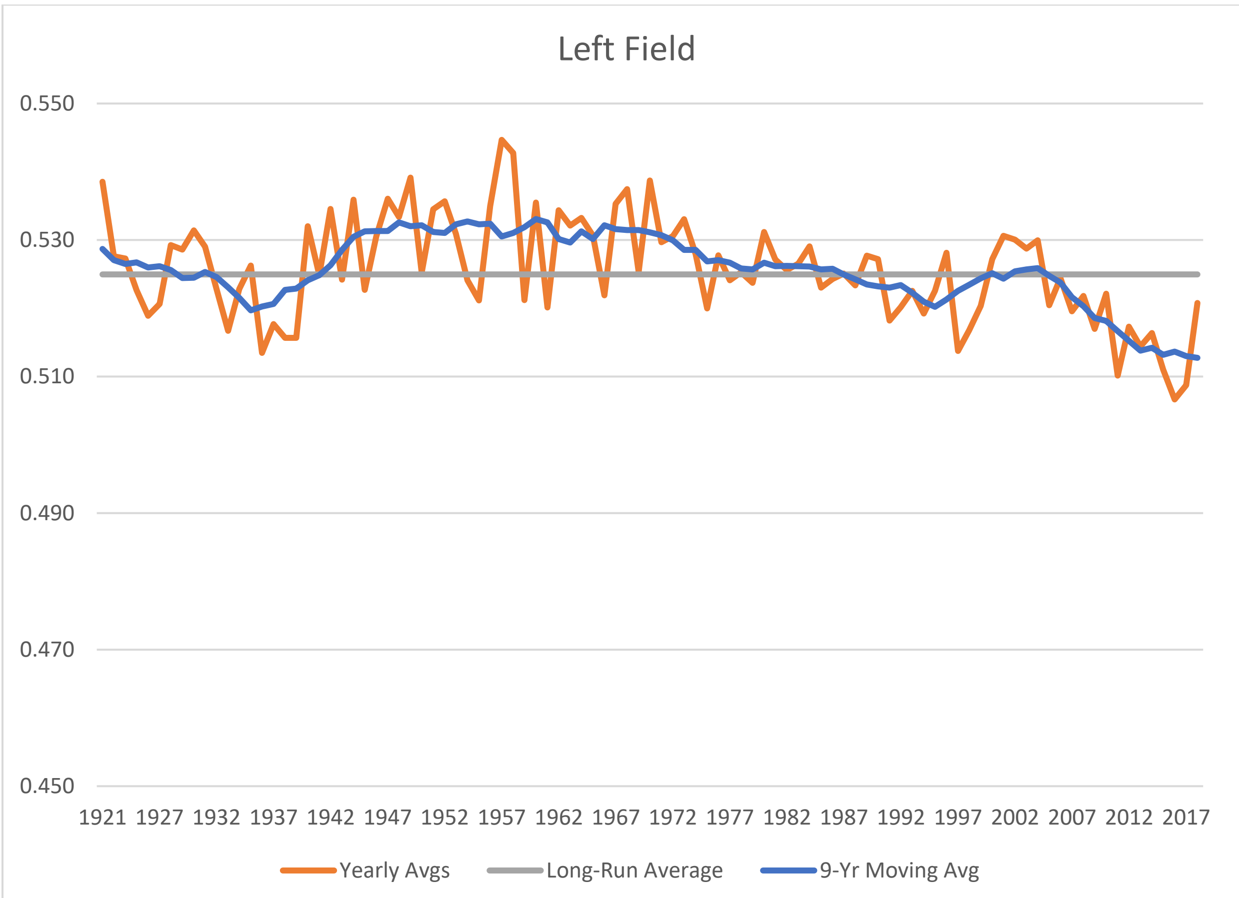
The above graph, then, combines the results from the previous three graphs and looks at how the positional averages at second base, third base, and shortstop have shifted over time relative to one another. The gray line is second base, the orange line is third base, and the blue line is shortstop.

Second base and third base traded positions in the 1920s. Since then, third base has been consistently higher than second base and shortstop almost without exception. Second base and shortstop were very similar in positional average from about 1934 through 1966. Over those 33 seasons, the positional average for second base was 0.489 vs. 0.487 for shortstop. The positional average for shortstop dropped off at this time and had stayed below second base every season since then until 2018 when shortstops amassed a positional average of 0.507 vs. 0.502 for second basemen (note that both numbers come in just below 0.500 when the DH-adjustment is factored in).

Overall, the differences between the three infield positions have narrowed in recent years. Across the entire 98-year period, the positional averages are 0.509 for third base, 0.494 for second base, and 0.484 for shortstop. The separation between the three was sharpest from, say, 1967 to 1986. Over these twenty years, the positional averages were 0.513, 0.490, and 0.472. Since 2011, these averages are 0.512, 0.503, and 0.496.

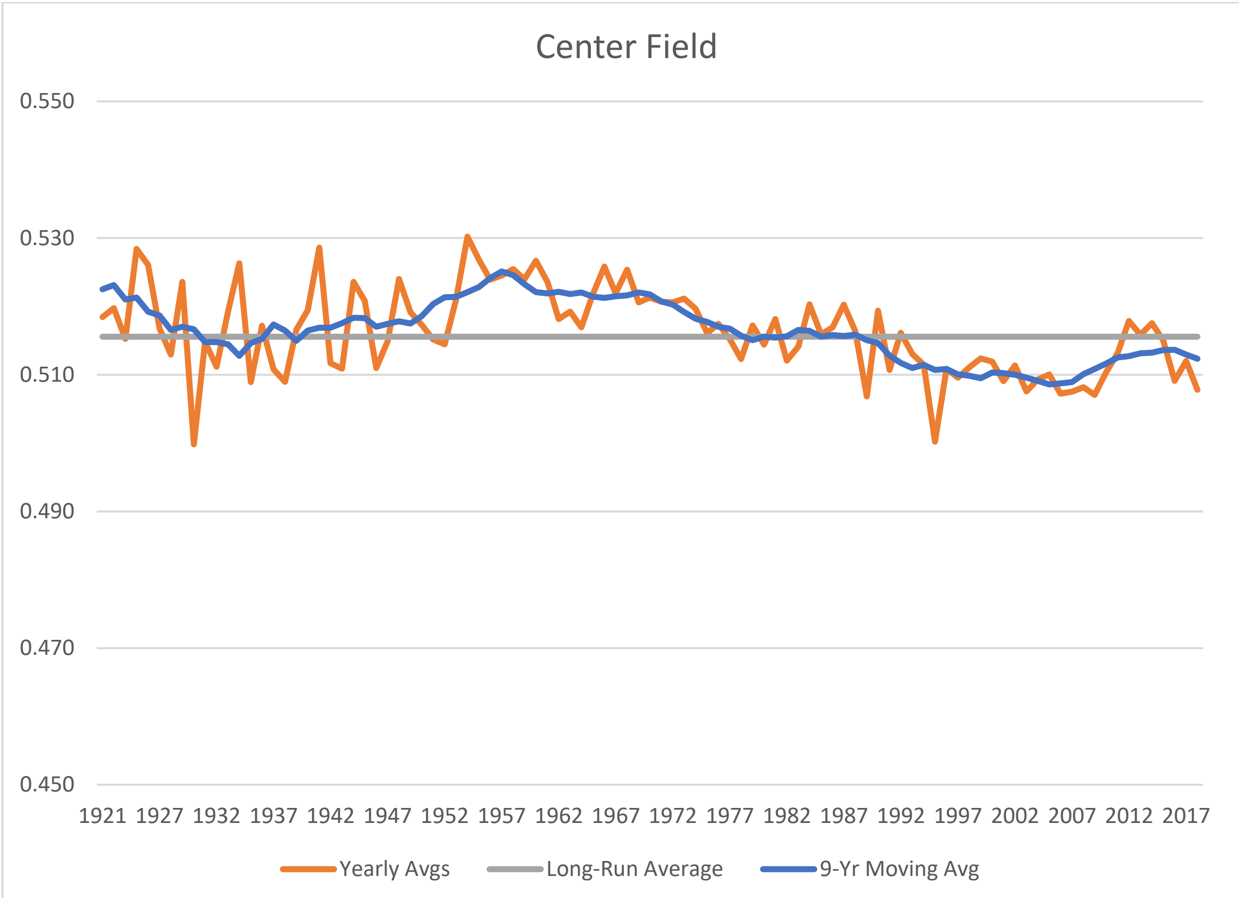
One possible explanation for this could be that infield positions have gotten more fungible over time, perhaps due to shifting which has made the definitions of the specific positions more amorphous. It will be interesting to watch these trends unfold over the next few years. If infield positions become more similar, it may become more difficult to compare modern infielders for whom expectations may be similar across all three non-first base infield positions to historical infielders for whom the defensive spectrum was more clearly defined.

- Left Field



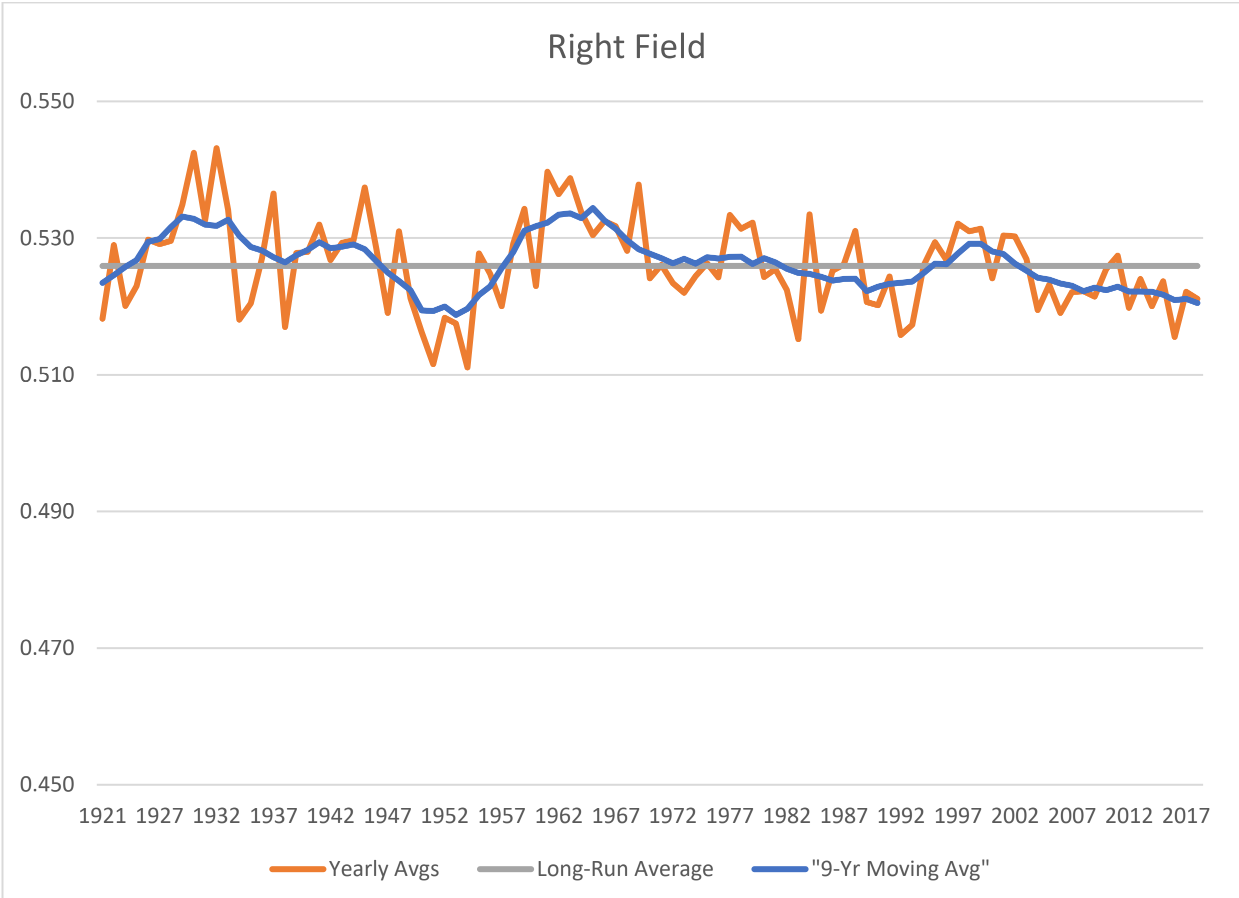
Left field was an especially strong position from about 1940 through 1973 with an average value of 0.531. Seasonal positional averages have been below the long-run average (0.526) every year since 2005 (the season after Barry Bonds won the last of his seven MVPs, all of which he won as a left fielder). The positional average for left field from 2005 – 2018 has averaged 0.516.

- Center Field



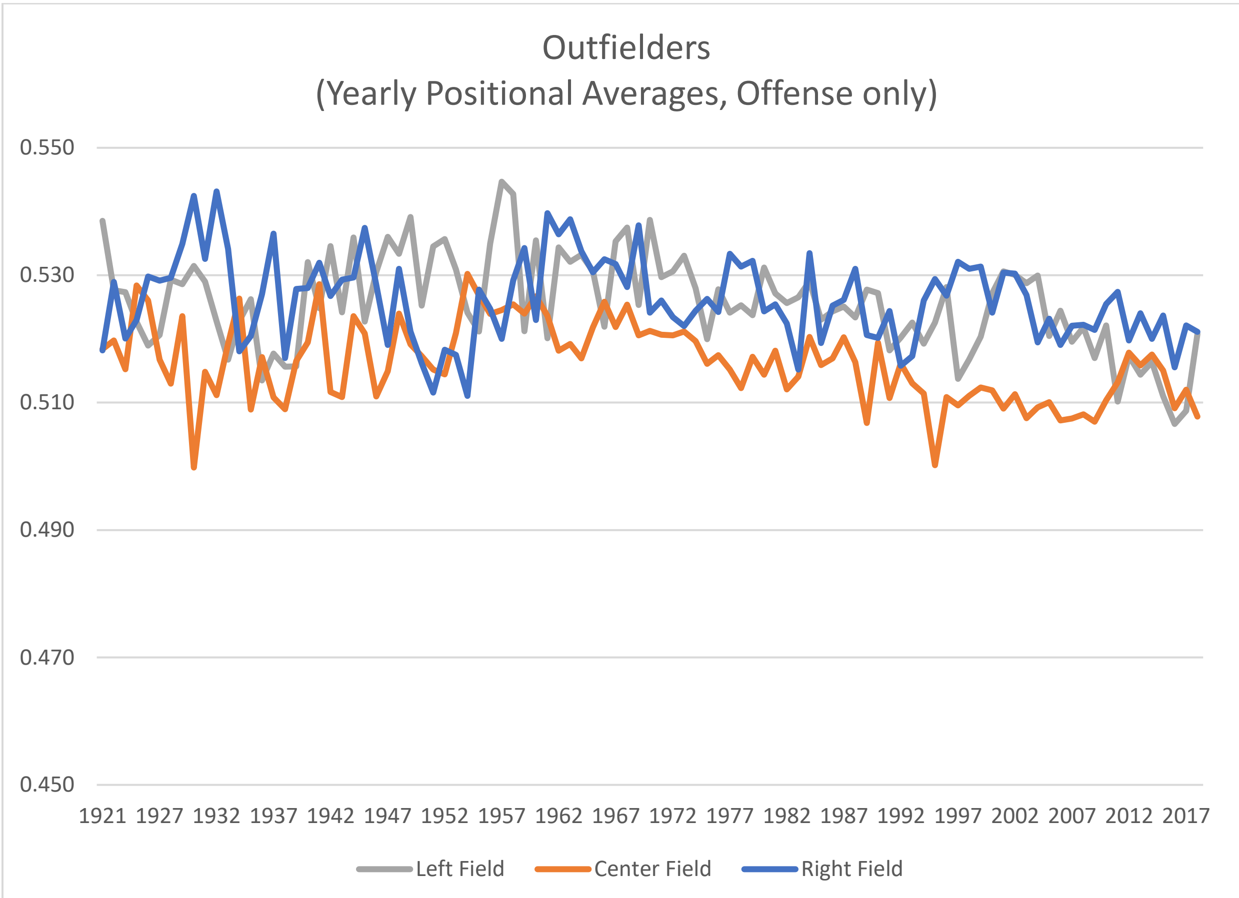
Positional averages for center field peaked in the 1950s – Willie Mays, Mickey Mantle, Duke Snider, Larry Doby, et al. Like left field, recent positional averages are below the long-run level. The positional average over the full 98 years is 0.516. This has fallen to 0.511 since 1994.

- Right Field



Eyeballing the graph, right field seems perhaps a bit more stable than left field. Like both left field and center field, recent positional averages have been below the long-run average. In the case of right field, the relevant time period seems to be 2004 – 2018 with a positional average of 0.522 (versus a 98-year average of 0.526).

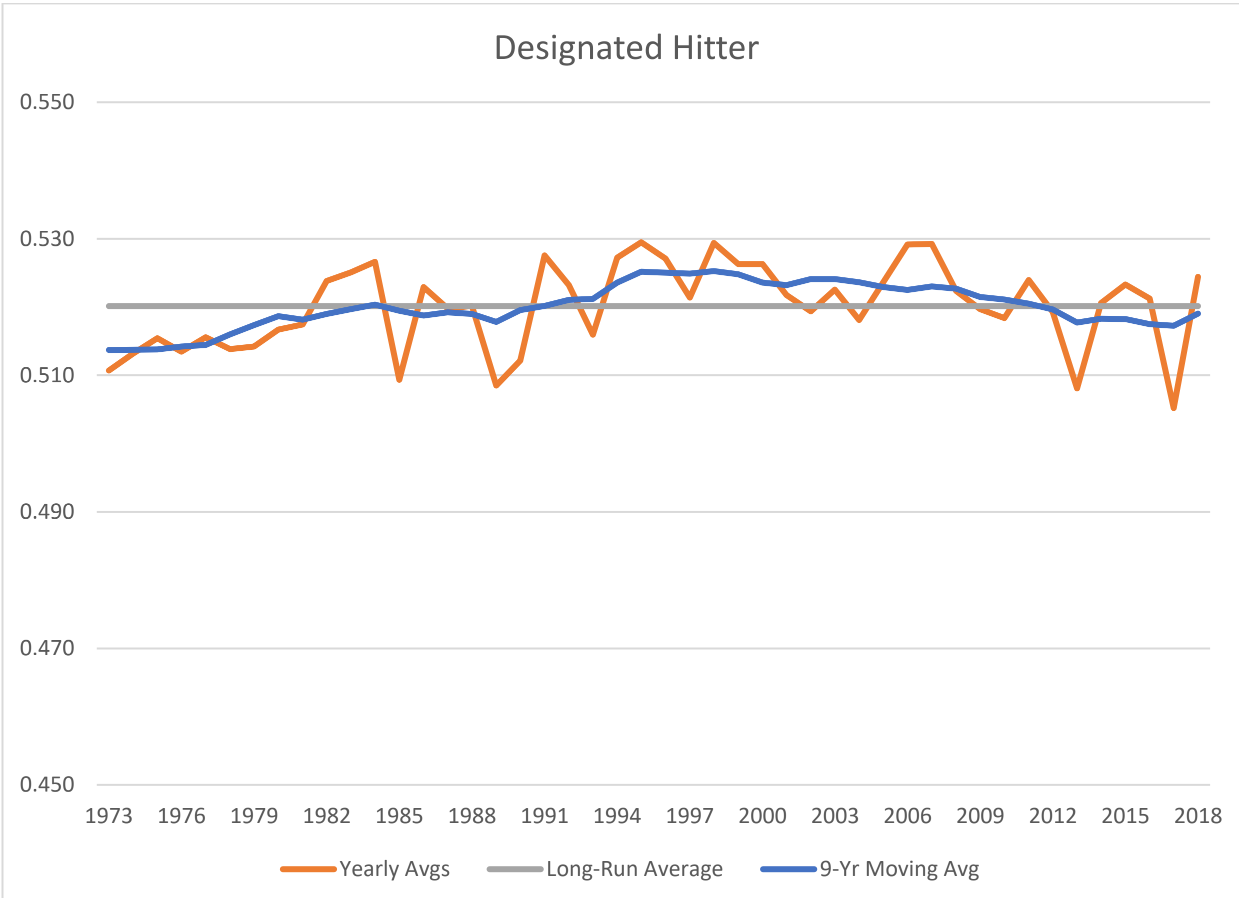
- Outfielders



The above graph, then, combines the results from the previous three graphs and looks at how the positional averages at left, center, and right field have shifted over time relative to one another. The gray line is left field, the orange line is center field, and the blue line is right field.

Center field has tended to have a lower positional average than the corners although this has not always been true. Left field and right field, however, have tended to track one another. The 98-year average for right field is 0.526; for left field, 0.525.

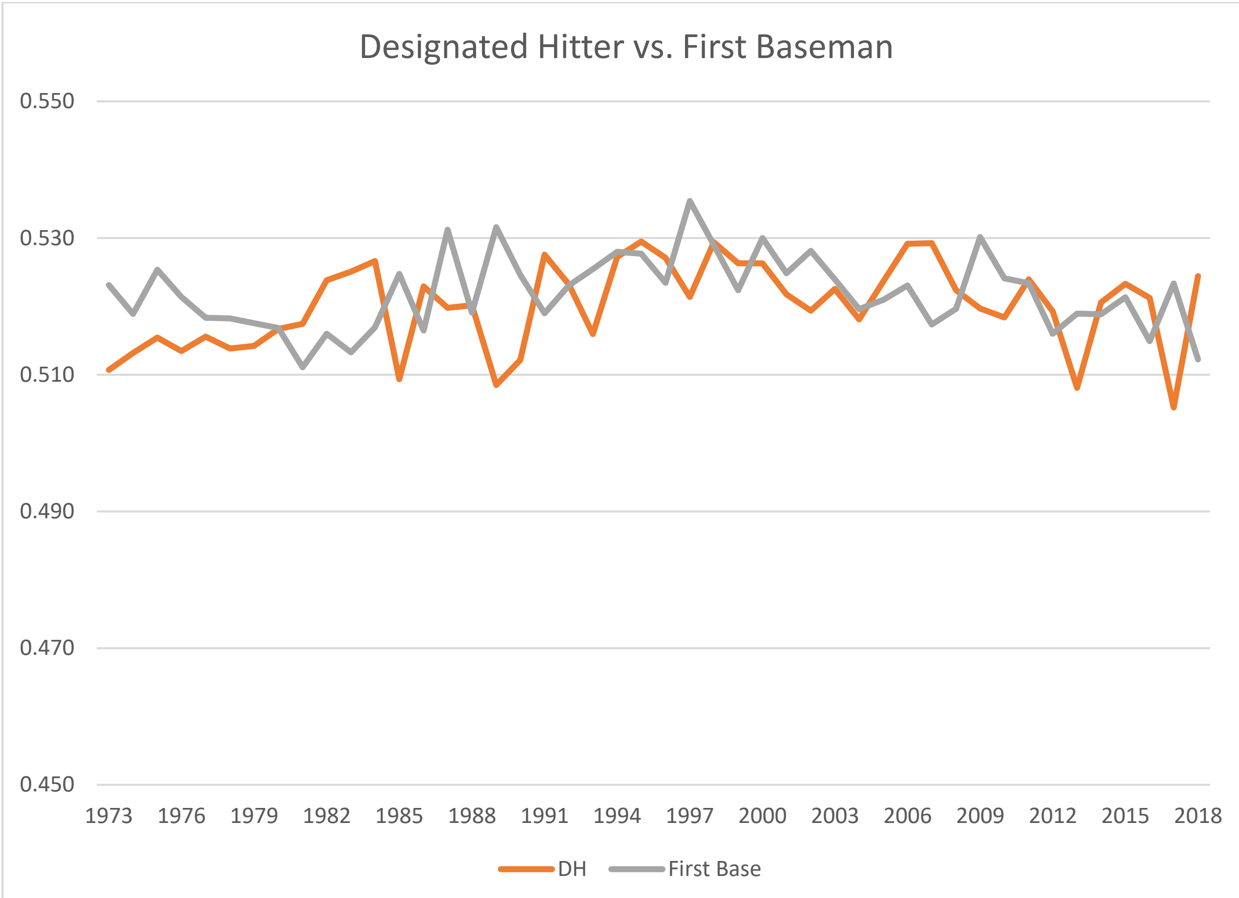
- Designated Hitter



The numbers here only go back to 1973, because the DH rule only goes back to 1973. And, obviously, the positional averages here apply to DH-leagues only.

There is a clear and rather strikingly smooth positive trend in the positional average for DH from 1973 into the early 1980s. The positional average for DH was 0.510 in 1973 and increased in 9 of the next 11 seasons, peaking at 0.526 in 1984. Since that time, the positional average for DH has held relatively stable. The long-run average here is 0.520. The average since 1982 is 0.521.

- DH vs. 1B

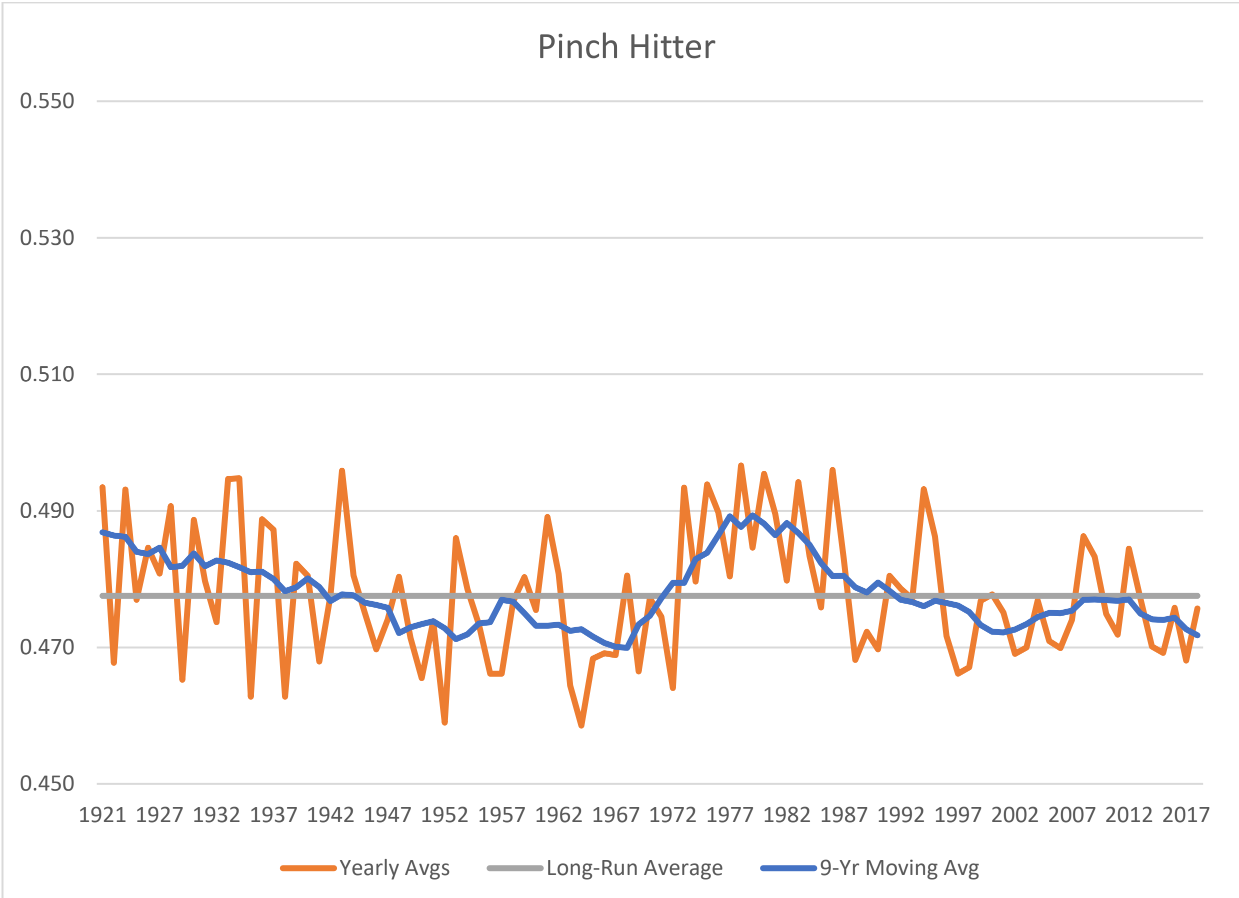


The above graph compares positional averages for designated hitters and first basemen. The first base numbers here include the DH-adjustment and so are about one percent (0.010) lower than the first base numbers shown earlier.

The ramp-up of DH averages that I noted earlier can be seen here and coincides with a mirror-image ramp-down of first base averages over the same time period.

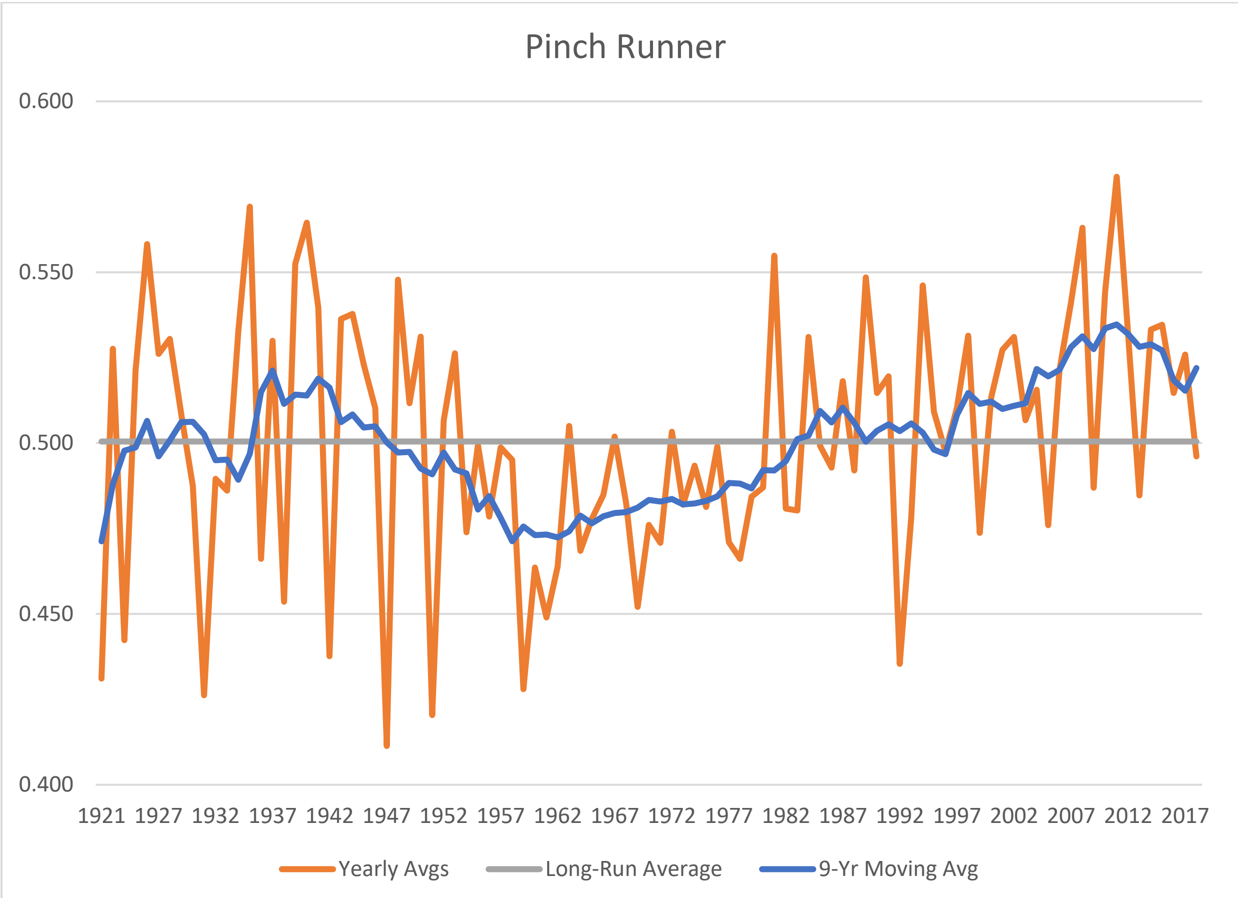
By 1980, the two numbers are almost literally identical (0.5167 for first base, 0.5162 for DH) and remain essentially equal thereafter. From 1980 through 2018, the positional average for DH is 0.521 and for first basemen in DH-leagues, the positional average is 0.522.

- Pinch Hitter



The results here are erratic. The blue line helps to show more subtle patterns that are largely overwhelmed by the year-to-year volatility of the orange line. Looking at the blue lines, there appears to have been a negative trend in pinch hitter average from 1921 (0.493) through 1972 (0.464) and again from 1978 (0.497) through the present (0.476 in 2018). In between these two negative trends, the positional average for pinch hitters jumped from 0.464 in 1972 to 0.493 in 1973. Not coincidentally, 1973 was the year in which the DH was introduced in the American League. In a DH league, the standard for pinch hitting is higher than in a non-DH league. In the latter, most pinch hitters bat for the pitcher and (not to minimize the difficulty of hitting major-league pitching) it doesn't require all that much for a major-league hitter to hit better than a pitcher. But in non-DH leagues, the only pinch hitters will be batting for other position players which may require a better set of hitters to make the pinch hitting worthwhile.

- Pinch Runner



I mentioned earlier that the range on the y-axis for all the positional average graphs from catcher through pinch hitter are the same, 0.450 to 0.550. The range here is twice as large, 0.400 to 0.600. The reason for this is, I think, obvious if you look at the graph. There were several years where the positional average for pinch runners was below 0.450. And there were several years where the positional average for pinch runners was above 0.550. Hence, it should go without saying that the results for pinch runners are more volatile than for any other position. In fact, they are vastly more volatile than for any other position.

One way to measure the volatility of a series is to calculate the standard deviation of the series. The largest standard deviation for any of the eight non-pitcher fielding positions is shortstop at 0.0091. As a rough explanation, this means that approximately two-thirds of shortstop positional averages are within 0.009 of the long-run average, or within a range of 0.475 to 0.493 (69 of 98 positional averages for shortstop fall within this range). The standard deviation for pinch hitters is somewhat greater than for shortstops, 0.0098. The standard deviation for pinch runners is 0.0359 – four times greater than for shortstops!

It is hard to see within the volatility but there appears to be a positive trend in pinch-runner average since perhaps as far back as the 1940s (the minimum value here, 0.411, was in 1947; the peak, 0.578, was in 2011). That said, the long-run average here is 0.500 which is not terribly different from the 2018 average of 0.496.

- Pitchers

So, having worked through the offensive positions, that leaves us with pitching. As I said earlier in this essay, I believe that there are three possible ways to calculate positional averages for pitchers.

(1) Pitching is pitching: the positional average is 0.500 for all pitchers for all seasons by construction.

(2) Starting pitchers and relief pitchers should have different positional averages, which should be calculated empirically each season. That is, in 2018, the overall winning percentage for starting pitchers (excluding their offense) was 0.497; the overall winning percentage for relief pitchers (again, only on defense) was 0.504. Those are your positional averages for 2018.

(3) Starting pitchers and relief pitchers should, indeed, have different positional averages, but option (2) assumes that the pool of starting pitchers and the pool of relief pitchers are equal. What we should do, instead, is focus on pitchers who pitched as both starters and relief. Doing this produces positional averages for 2018 of 0.487 for starting pitchers and 0.518 for relief pitchers.

In my first book, *Player Won-Lost Records in Baseball: Measuring Performance in Context* (McFarland, 2017), I chose option (3). Last summer, I had an e-mail conversation with a SABR member named Bob Sawyer who objected to my choice here. He raised some good points, which prompted me to re-think this choice. Initially, this re-thinking led me to choose option (1) instead. But now, in keeping with my general philosophy of allowing as much user freedom in how people can use my data as possible, I have decided to give people the option of any of these three choices or a weighted average of them.

While allowing people as much freedom to use my Player won-lost records to evaluate data as one sees fit, let me share some thoughts on pitching positional averages.

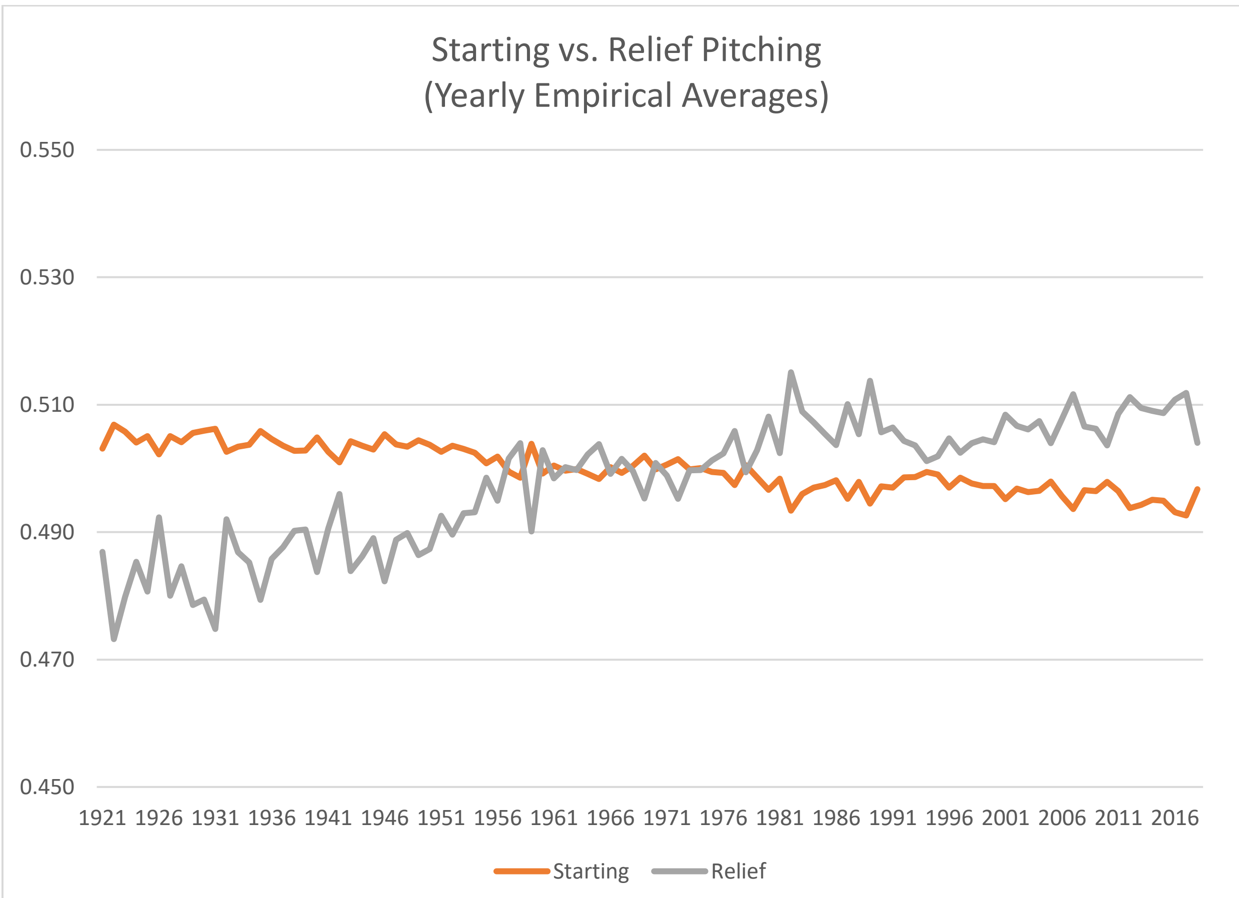
First, let's think about pWins. The underlying point of Player won-lost records is to explain team wins at the game level. From the team's perspective, what is the difference between starting pitchers and relief pitchers? Ultimately, nothing. For my 50th birthday (June 6, 2018), I attended a Cubs game against the Philadelphia Phillies at Wrigley Field. In that game, the Cubs led 3 – 0 with two outs and runners on first and second base in the top of the sixth inning. The Cubs replaced starting pitcher Jose Quintana with relief pitcher Steve Cishek. The next batter, Aaron Altherr, hit a three-run home run tying the score at 3 – 3. Would anything have been any different if Quintana had stayed in the game and given up the home run to Altherr? No. But if the positional average for starting pitchers differs from the positional average for relief pitchers, doesn't that imply that the expectations are different for a starting pitcher than for a relief pitcher? Specifically, if the positional average for starting pitchers is lower than for relief pitchers (as is the case with either option (2) or (3) in 2018), that implies, at some level, that the home run would have been less costly if Quintana had given it up instead of Cishek. That doesn't make sense.

The 2018 season has also introduced a new wrinkle here: the opener. Why is the positional average lower for starting pitchers? In the case of option (3), it is because pitchers perform better when they pitch in relief than when they pitch as starters. Why? Because relief pitchers don't have to pace themselves; they don't face the same batters more than once in a game; they likely have the platoon advantage more often. But now, with the "opener", the starting pitcher – i.e., the first pitcher of the game – gets these advantages and it's the second pitcher of the game who is then expected to work through the lineup multiple times and pitch multiple innings.

Shouldn't openers be treated like "relief" pitchers? But how do you distinguish between a starting pitcher who pitched one inning because that was the plan versus a starting pitcher who pitched one inning because he was pulled from the game early because he was ineffective or injured? I don't know that you can.

There is also, I think, another problem. As the use of relief pitchers has changed so, too, has the use of starting pitchers. Teams are making more and more effort to avoid having starting pitchers face batters a third time in the same game. This is likely making the job of a starting pitcher easier over time.

The next graph shows empirical averages for starting pitchers and relief pitchers – i.e., option (2) above.

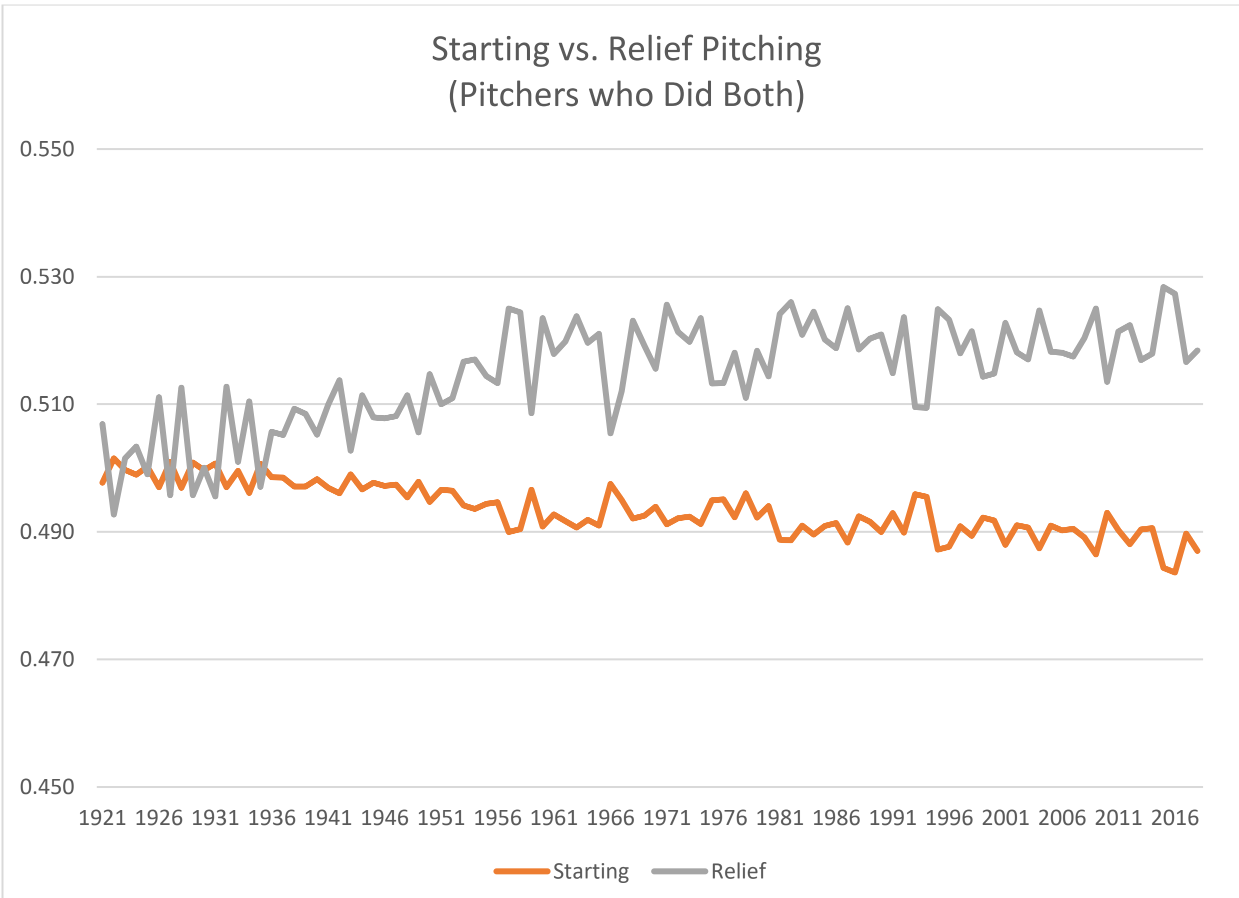


The average winning percentage of starting pitchers has generally trended uniformly downward over the past 98 years. The average winning percentage of relief pitchers trended upward – somewhat less uniformly – from, say, 1931 (0.475) through, say, 1977 (0.506). The average winning percentage for relief pitchers has bounced around a fair bit since 1977 but has not exhibited a long run trend. The average winning percentage for relief pitching from 1977 through 2018 has been 0.5065. For 2018, the figure was 0.504.

Actually, that's a bit of an oversimplification. The relief pitcher average from 2011 through 2017 was 0.510 and for starting pitchers was 0.494. In 2018, both of these averages shifted noticeably toward 0.500 – specifically, to 0.504 and 0.497, respectively.

One possible explanation for this one-year spike could be the effect of "openers" by the Tampa Rays. Openers, of course, are starting pitchers but have a role that more closely resembles that of a relief pitcher. Off-season chatter suggests that the use of the opener is likely to increase in 2019. This is a development worth following.

The next graph shows positional averages calculated by looking at pitchers who both started and relieved in the same season.



The numbers are interesting here. Two things, in particular, are striking about the first 15 years or so here. First, the relief pitcher numbers are ***extremely*** unstable, likely because relief pitching was much less common. Second, looking beneath the instability, there was essentially no difference in pitchers' performance starting versus relieving. From 1921 – 1935, the average starting pitcher average here was 0.499; for relief pitchers, 0.502. In the earliest days of relief pitching, the role of relief pitcher was very different than it is today. Relief pitchers were brought in when the starting pitcher failed (or was pinch-hit for) and the expectation was that the relief pitcher would likely finish the game. Sometimes, this translated into only pitching an inning. But sometimes, if the starting pitcher was knocked out in the first or second inning, this meant a relief pitcher entering a game with an expectation of pitching 6, 7, or 8 innings.

Since the mid-1930s, however, the spread between the starting and relief averages for pitchers who did both has consistently increased. In the 1940s, the gap was 0.011 (0.508 vs. 0.497). By the 2010s, the gap was up to 0.031 (0.520 vs. 0.489). There really isn't any evidence of the "opener" here. From 2010 – 2017, the average for starting pitching was 0.489. In 2018, it was 0.487. For relief pitchers, the numbers were 0.520 from 2010 – 2017, 0.518 in 2018.

- Impact of Different Positional Averages

The most extreme difference between positional averages for pitchers would be between option (3), which was the default positional average at my website for most of its history, and option (1), which was the default positional average at my website for the past month or so. Comparing these two options, shifting from option (3) to option (1) lowers the positional average for relief pitchers – thereby increasing relief pitchers' wins over positional average (WOPA) and replacement level (WORL) – and increases the positional average for starting pitchers – thereby decreasing starting pitchers' WOPA and WORL. As the last graph showed, option (3) positional averages have been trending away from 0.500 over time for both starting and relief pitchers, so the gap between options (1) and (3) is greatest for more recent pitchers.

The next table shows the top 10 pitchers with the most extreme differences in career pWOPA between options (1) and (3), using single-season positional averages. As you would expect, the top 10 pitchers in increased pWOPA are long-career relief pitchers. The top 10 pitchers in decreased pWOPA are long-career starting pitchers.

Top Increases in Career pWOPA

				pWins over Positional Average		
	Player	pWins	pLosses	Option 3	Option 1	Difference
1	Lee Smith	111.3	78.2	13.3	16.9	3.56
2	Goose Gossage	131.2	101.8	11.7	15.2	3.48
3	Mariano Rivera	126.6	60.8	29.6	32.9	3.36
4	John Franco	103.7	74.3	11.6	14.9	3.35
5	Rollie Fingers	121.0	97.2	9.8	13.1	3.25
6	Jeff Reardon	91.3	71.3	7.1	10.3	3.19
7	Hoyt Wilhelm	138.6	116.0	10.4	13.5	3.10
8	Frankie Rodriguez	90.0	56.9	13.6	16.6	2.94
9	Trevor Hoffman	100.7	62.3	16.5	19.5	2.93
10	Gene Garber	88.1	78.4	2.7	5.6	2.89

Top Decreases in Career pWOPA

				pWins over Positional Average		
	Player	pWins	pLosses	Option 3	Option 1	Difference
1	Nolan Ryan	356.8	328.0	24.6	19.7	-4.88
2	Greg Maddux	328.5	271.3	44.8	39.9	-4.84
3	Roger Clemens	318.1	228.1	51.1	46.3	-4.84
4	Tom Glavine	279.4	249.5	29.0	24.8	-4.26
5	Randy Johnson	281.2	220.9	38.5	34.2	-4.26
6	Bartolo Colon	212.8	205.4	9.9	5.6	-4.22
7	Steve Carlton	337.9	303.9	31.7	27.6	-4.16
8	Bert Blyleven	291.2	260.0	22.2	18.0	-4.16
9	C.C. Sabathia	213.6	179.2	22.2	18.1	-4.13
10	Don Sutton	320.8	295.3	24.6	20.5	-4.08

- Should Positional Averages Be Based on Players' Offensive or Defensive Performances?

The positional averages presented and discussed so far here are based on players' offensive performances. This may seem wrong to some people. Specifically, many people will object that players do not bat as shortstops or left fielders, they bat as batters, i.e., players' offensive performance is not necessarily tied to a defensive position.

In fact, however, having decided that the way to compare players at different positions is by measuring players against different positional averages, I think that this is the correct way to calculate positional averages. This is because what we are trying to control for are the offensive expectations of players at different positions. And while it's true, in one sense, that a player does not bat as a "shortstop" or a "left fielder"; he bats as a batter, there is another sense in which this is not true. From the perspective of the team, the lineup must be filled with players who will each field one of the 8 or 9 defensive positions (with the exception of the DH in the American League). If, then, say, the Texas Rangers can find a player who can play shortstop adequately while hitting better than the players the other teams are playing at shortstop (e.g., Toby Harrah in the first five years of the Texas Rangers' existence, 1972 – 1976), this is a real advantage for the Rangers.

Having said that, there is a way to calculate positional adjustments based on fielding, and, in fact, I believe this is the basis for the positional adjustments used by Baseball-Reference in their calculation of WAR. This can be done by looking at the fielding performances of players who play multiple positions within the same season. For example, across all seasons for which I have estimated Player won-lost records, players who played both left field and center field within the same season had an average winning percentage of 0.488 in center field and 0.510 in left field. From this, one could reasonably conclude that center field is a more difficult position to play and one could also use this difference as a basis for adjusting these winning percentages to reflect a common base.

Comparisons of this type were done for all of the infield and outfield positions. Pitchers and catchers are not considered here. In the case of pitchers, this is because pitchers virtually never play a different position. This is also true, although to a lesser extent, of catchers. More problematic, however, in the case of catchers, is the fact that the skill set needed to be a good major-league catcher is not really the same skill set needed to be a good fielder at any other position (the same is true to a lesser extent, of course, when comparing infielders to outfielders, and, really, is true to at least some extent in every case here).

Average Winning Percentage at Position X

	1B	2B	3B	SS	LF	CF	RF
1B		0.527	0.522	0.541	0.505	0.504	0.504
2B	0.491		0.492	0.497	0.486	0.487	0.485
3B	0.481	0.496		0.500	0.480	0.477	0.477
SS	0.482	0.489	0.487		0.484	0.482	0.484
LF	0.487	0.503	0.496	0.509		0.510	0.501
CF	0.484	0.492	0.490	0.498	0.488		0.490
RF	0.480	0.490	0.489	0.494	0.493	0.506	

This table is read as follows. For a player who played both first base and second base, the average winning percentage at first base is shown in the top row, 0.527 - this is the average winning percentage of second basemen when they are playing first base. The average winning percentage of first basemen when they are playing second base is shown in the first column, 0.491. In all cases here, average winning percentages are calculated as weighted averages where the weights used are the harmonic mean between the player decisions at the two fielding positions being compared.

The average “normalized” winning percentage for a player at position Y when playing other positions can then be calculated as the weighted average of the numbers down the relevant column. The weights used to calculate these averages were the number of games upon which the comparison was based, which, as noted above, was the harmonic mean of the number of Player decisions accumulated at the two positions being compared.

After a few steps to re-center the winning percentages to average to 0.500 overall, produces the following average winning percentages by fielding position:

1B	0.472
2B	0.501
3B	0.504
SS	0.512
LF	0.488
CF	0.517
RF	0.499

Several aspects of these results are noteworthy. First, the range of winning percentages is narrow, outside of first basemen; much narrower than the range in offensive performance across positions. Another interesting comparison, I think, is that third base appears to be a tougher position to play than second base. Bill James discusses this in his *Win Shares* book, where he discusses the historical shift of the defensive spectrum with second base becoming more important than third base over time. As Bill James puts it:

"Third basemen need quicker reactions, since they are nearer the batter, and they need a stronger arm, since they are further from first base. Without the double play, third base is obviously the more demanding position." (*Win Shares*, p. 183)

The results here confirm this. Second base is, in one sense, the more valuable position, with approximately 19 percent more player decisions accumulated at second base than at third base, a difference which comes entirely from what I call component 7 (double plays). Yet, comparing how well fielders do when they play both second base and third base in the same season, third base is the more difficult position.

The next table, then, compares these results with relative fielding winning percentages implied by average offensive performances by position.

	Implicit Fielding Win Percentage	
	Relative Fielding	Offense
1B	0.472	0.394
2B	0.501	0.534
3B	0.504	0.501
SS	0.512	0.547
LF	0.488	0.470
CF	0.517	0.488
RF	0.499	0.467

Combining the results in the first column of that table with the average offensive winning percentages by position yields the following average winning percentages by defensive position:

1B	0.519
2B	0.495
3B	0.506
SS	0.492
LF	0.511
CF	0.514
RF	0.515

Looking at total player game points by position for non-catchers reveals an interesting contrast. First basemen and outfielders turn out to be better than 0.500 on average (0.515 to be exact) while second basemen, third basemen, and shortstops turn out to be slightly below average as a group (0.497). So, is this a mistake? Not necessarily. One possible explanation for this that I have seen suggested is the fact that, in major league baseball, left-handed throwers are only employed as pitchers, first basemen, and outfielders. Because other defensive positions are limited to those who throw right handed, however, these positions draw from a more limited pool of candidates. I can see some merit to this, but am also skeptical. In terms of "value" as opposed to talent, the fact remains that a team has to play a 2B, a 3B, and a SS, and (right or wrong), these players have to throw right-handed.

So, which methodology produces better results?

For my work, I have chosen to calculate my positional averages based on relative offensive performances by position. I do this for several reasons which, I believe, make this a better choice for my purposes.

First, the mathematics here, attempting to normalize winning percentages across fielding positions, is fairly murky. In contrast, simply setting the positional average equal to the average winning percentage compiled at that position seems to me to be much cleaner and more elegant mathematically.

Second, I believe that limiting the analysis only to players who have played more than one position in the same season, as is done here, may lead to issues of selection bias. That is, we are not looking at the full population of all major-league players here, since most major-league players never played a game at shortstop, for example, or a random sample of major-league players. Instead, we are looking at a selected sample of major-league players, who were selected, in part, on the basis of exactly what we are attempting to study: i.e., the players considered here are self-selected for their ability to play multiple positions similarly well. Truly bad players at offense-first positions - think Frank Thomas at first base, Manny Ramirez in left field - are so bad that nobody would ever consider trying to play Frank Thomas at third base or Manny Ramirez in center field. But, at the other end, great defensive players at defense-first positions are so great defensively that, for example, Ozzie Smith never played a single inning of major-league baseball at any defensive position besides shortstop; Willie Mays never played a corner outfield position until he was 34 years old. If a player is significantly better at one position than another, it almost certainly makes more sense for a team to simply play that player at his better position – or, if he is blocked at his best position, to trade him for a player who better fits the team's needs. Because of this, I think that looking only at the defensive performance of players who played multiple positions understates the true spread in positional averages – and, I suspect, dramatically so.

Finally, I believe that setting positional averages based on actual empirical winning percentages is more consistent with what I am attempting to measure with my Player won-lost records. Player won-lost records are a measure of player value. At the bottom-line theoretical level, every team must field a player at all nine positions. If one team has a second baseman that is one win above average and another team has a left fielder who is one win above average, then these two teams will win the same number of games (all other things being equal). Hence, in some sense, not only is it a reasonable assumption to view an average second baseman as equal in value to an average left fielder, it is, in fact, a necessary assumption.

All of that said, I understand that this is a subjective decision and I would hate for somebody to be hesitant in accepting my Player won-lost records because of a subjective decision that, at the end of the day, does not actually affect the key output of my work, Player wins and Player losses.

This segues nicely, then, into the next section of this essay, where I look at the options which I allow for calculating positional averages.

- Positional Average Options: Cases For and Against and Potential Uses

I allow for four possible choices of positional averages, or a weighted average of any (or all) of these choices.

  - 0.500

I allow for the possibility of simply measuring all players against a standard of 0.500. This is the one of the four options that I allow which I would argue is not correct and this is the one of the four options which I do not include in the positional averages which display by default at my website. Let me make my case against 0.500 by example.

Mark Grace had a fine major-league career. Over the course of his 16-year career, Grace played 2,245 games with 9,290 career plate appearances and 18,587 defensive innings (18,586 at first base, 1 as a pitcher). Grace won four Gold Gloves and had a career batting line of .303/.383/.442 with 173 career home runs. He was a very good major-league player.

Alan Trammell had a 20-year major-league career. He played 2,293 games with 9,376 plate appearances and 18,731 defensive innings – all very similar to Mark Grace. Trammell also matched Grace in career Gold Gloves, with four. Alan Trammell was also similar, but probably somewhat inferior to Mark Grace as a batter with a career batting line of .285/.352/.415 with 185 career home runs.

As measured by Player won-lost records, Alan Trammell had a career batting record of 174.1 – 164.7, +9.4 net wins. Mark Grace had a record of 187.1 – 162.6, +24.5 net wins. Trammell made up some of that gap in baserunning: Trammell was 18.9 – 16.9 (+2.0 net wins); Grace was 15.1 – 14.3 (+0.8 net wins). Both players were excellent fielders: Trammell was 84.1 – 77.8 (0.519 winning percentage, +6.3 net wins); Grace was 40.2 – 36.3 (0.525, +3.9 net wins). They were very similar across the board. Add it all up, and Alan Trammell's career eWins and eLosses were 280.9 – 259.1; Grace had an eWin – eLoss record of 241.6 – 212.4.

Using a positional average of 0.500, Mark Grace beats Alan Trammell in career eWins over "positional" average (eWOPA), 14.6 – 10.9.

Does that make sense? All of the analysis thus far has ignored one thing. Mark Grace was a first baseman. He was usually one of the best hitters on his team, but most first basemen are among the best hitters on their team. And many – maybe not most, but many – first basemen regularly out-hit Mark Grace. Alan Trammell, on the other hand, was a shortstop. And except for a few well-known exceptions (Cal Ripken, Robin Yount), Alan Trammell was one of the best hitting shortstops in baseball.

Mark Grace was named to three All-Star teams in his career. He appeared on one Hall-of-Fame ballot, receiving 22 votes (4.1%) – which, I'll be honest, is 20 more votes than I would have guessed. Alan Trammell was named to six All-Star teams in his career. In his first year on the Hall-of-Fame ballot, he received 74 votes – which is perhaps 300 fewer than he should have received. It took far too long, but Alan Trammell is a Hall-of-Famer; Mark Grace is not.

And the entire difference between their careers was essentially the positional average against which their careers are measured. Using one-year positional averages, Trammell beats Grace in eWOPA 21.5 – 1.7; using long-run positional averages, Trammell wins 19.8 – 3.4. Mark Grace was a pretty good major-league baseball player, above average at his best. Alan Trammell was a Hall-of-Famer. You need positional averages to fully appreciate the statistical difference between them.

Now, having said that, one could perhaps use 0.500 as a way to regress the positional averages toward 0.500. Personally, I do not think that it would be appropriate to regress the positional averages toward 0.500, if for no other reason than it would adversely affect comparisons between players in DH and non-DH leagues (this is part of what makes Grace look better than Trammell when measured against 0.500). Nevertheless, as I discussed above, if one were inclined to calculate positional averages by looking at how players' fielding varies across positions, the result would be a narrower band of positional averages than my calculations based on offensive performance. One could, for example, use a weighted average of 0.500 and one (or more) of the other positional average options as a way to regress the positional averages toward 0.500 as a way to mirror fielding-based positional averages.

- One-Year Positional Averages

The positional averages used in my first two books and, until very recently, on my website were one-year positional averages. I have implicitly made my argument for one-year positional averages throughout this essay. Let me try to summarize it in one paragraph here.

For pWins, where we are trying to explain team wins at the game (and season) level, I believe it is completely appropriate to use single-season positional averages and let the results fall where they may. If everybody has a first baseman who can hit the crap out of the ball, you'd better find a first baseman who can hit if you want to compete. But, on the other hand, if everybody else has a shortstop who can't hit his weight, then finding a shortstop who can hit is going to be a big help in beating other teams. And, of course, at the team level, you can trade these things off: a team that is one win below average at first base but two wins above average at shortstop (net: +1 win) should do better than a team that is average at both positions.

Do I believe that this is a strong and valid argument? I do. Is it an objectively unassailable argument? It is not. Hence, my decision to let people choose their own positional averages (from among a finite set of choices).

- Nine-Year Positional Averages

The obvious advantage of nine-year positional averages over one-year positional averages is that they are smoother with far fewer and less dramatic year-to-year spikes. Nine-year positional averages will generally maintain smoother trends, peaks, and valleys within the data, however.

For example, positional averages for second base have been above their long-term average (0.494) in recent years. Starting in 2010, the one-year positional average for second base by season was 0.500, 0.501, 0.494, 0.503, 0.499, 0.501, 0.516, 0.505, and 0.502, respectively. The 0.516, which was in 2016, sticks out like a sore thumb. Averaging the 2016 value with the four years on either side (well, actually just the two years since – the nine-year positional averages for 2016 will change with the release of 2019 and 2020 data) reduces the positional average that year to 0.503. The nine-year positional averages for the years surrounding 2016 go up somewhat as they incorporate the unusual 2016 value. The nine-year positional averages for second base from 2010 – 2018 are 0.500, 0.501, 0.502, 0.503, 0.502, 0.503, 0.503, 0.504, and 0.505. The fact that the one-year positional averages in 2017 and 2018 remained high, relative to historical norms shows up in the nine-year positional averages for not only these seasons (0.504, 0.505) but also for the seasons just prior to 2016. But the nine-year positional average for second base never changes year-to-year by more than 0.001 during this time period, as opposed to the +0.015 and -0.011 changes in the one-year averages in 2016 and 2017.

Why 9 Years?

When I first started working through this exercise of allowing people to choose their own positional averages, I planned to use five-year positional averages as my middle option. Having worked up some five-year positional averages, though, I felt like they didn’t really smooth the numbers out as much as I had hoped they would. Basically, if you choose too short a time period, the results are not different enough from simple one-year positional averages; if you choose too long a time period, the results are not different enough from long-run positional averages. I feel like nine years hits a nice sweet spot between those two.

- Long-Run Positional Averages

At the opposite extreme of one-year positional averages is long-run positional averages. No fluke seasons, no occasional flips of the defensive spectrum. But also, no allowance for changes in relative positional averages.

One-year positional averages posit that any changes in relative performance across positions is “real” and should be accounted for in assessing player value.

Nine-year positional averages posit that any changes in relative performance across positions that persist across several seasons are “real” and should be accounted for in assessing player value.

Long-run positional averages posit that any changes in relative performance across positions between seasons or even between long groups of seasons are probably just noise and should be treated as such. Personally, I would be hesitant to rely exclusively on long-run positional averages. But I do think that they can be useful as part of a set of weighted-average positional averages as a way of hedging against what may be anomalous periods of several seasons.

For example, consider the shortstop position. The long-run positional average for shortstop is 0.484. But the one-year positional average for shortstop was below 0.484 every season from 1965 through 1982. In my online discussions of positional averages, I think this time period has probably generated the most controversy. Why did the shortstops of the late 1960s and 1970s hit so poorly? I have heard, in effect, two hypotheses, which are not necessarily mutually exclusive, but which clearly suggest different positional averages.

One hypothesis is that the introduction and rapid expansion of Astroturf in major-league infields dramatically increased the fielding requirements for shortstops. The ball reaches infielders much quicker on Astroturf with the result that speed became much more important. The increased need for speed on defense then led to an increase in the speed of players on offense, which further increased the difficulty of playing shortstop faced with a group of batter who were better able to beat out infield hits.

The second hypothesis essentially says that managers over-reacted to the perceived need for better fielders at shortstop and sacrificed too much offense at the position. Over-simplifying somewhat, the hypothesis is that major-league managers, as a group, were kind of stupid in the 1960s and 1970s.

Now, measured in terms of wins (as Player won-lost records are, by definition), major-league baseball is a zero-sum game. If everybody else is playing a lousy hitter at shortstop, it doesn't really hurt me if I play an equally lousy hitter at shortstop. Certainly, I'm missing out on an opportunity to exploit everyone else's mistake. And in real life, it's probably not entirely coincidental that the two teams who combined to win five straight World Series from 1972 through 1976 had two of the only three shortstops in the 1970s who were worth a damn offensively (exaggerating somewhat – but probably not too much – for effect). But if we want to compare Bert Campaneris and Dave Concepcion (and Toby Harrah, who was a pretty good hitter but never played for a playoff team) to shortstops from other eras, should the 1970s guys essentially get bonus points for playing in an era when major-league managers didn't know what they were doing?

You can choose the weight you want to apply to long-run positional averages that gets you to your answer for that question.

- A Caveat

I want to point out one caveat here that may be obvious. The weights chosen for positional averages will be applied equally to all positions. It is not possible, for example, to set positional average for third base constant since 1947, use a straight nine-year positional average for shortstops, and a straight single-year positional average for pitcher offense.

The primary reason for this is simply that the math is too complicated. I will say, though, that I also worry that using different standards for different positions may create distortions such that the average player within some season(s) may be above or below average, not merely at one or more positions (which will be true using anything but straight one-year positional averages) but across all players, which seems non-sensical to me: it is plausible that a league's third basemen were above-average in aggregate in some season; saying that the major-league players were above-average in aggregate in a season seems to misunderstand what the word "average" means.

I do, however, allow one to make position-specific adjustments in calculating all-encompassing "uber-statistics". This is discussed in the last section of this essay.

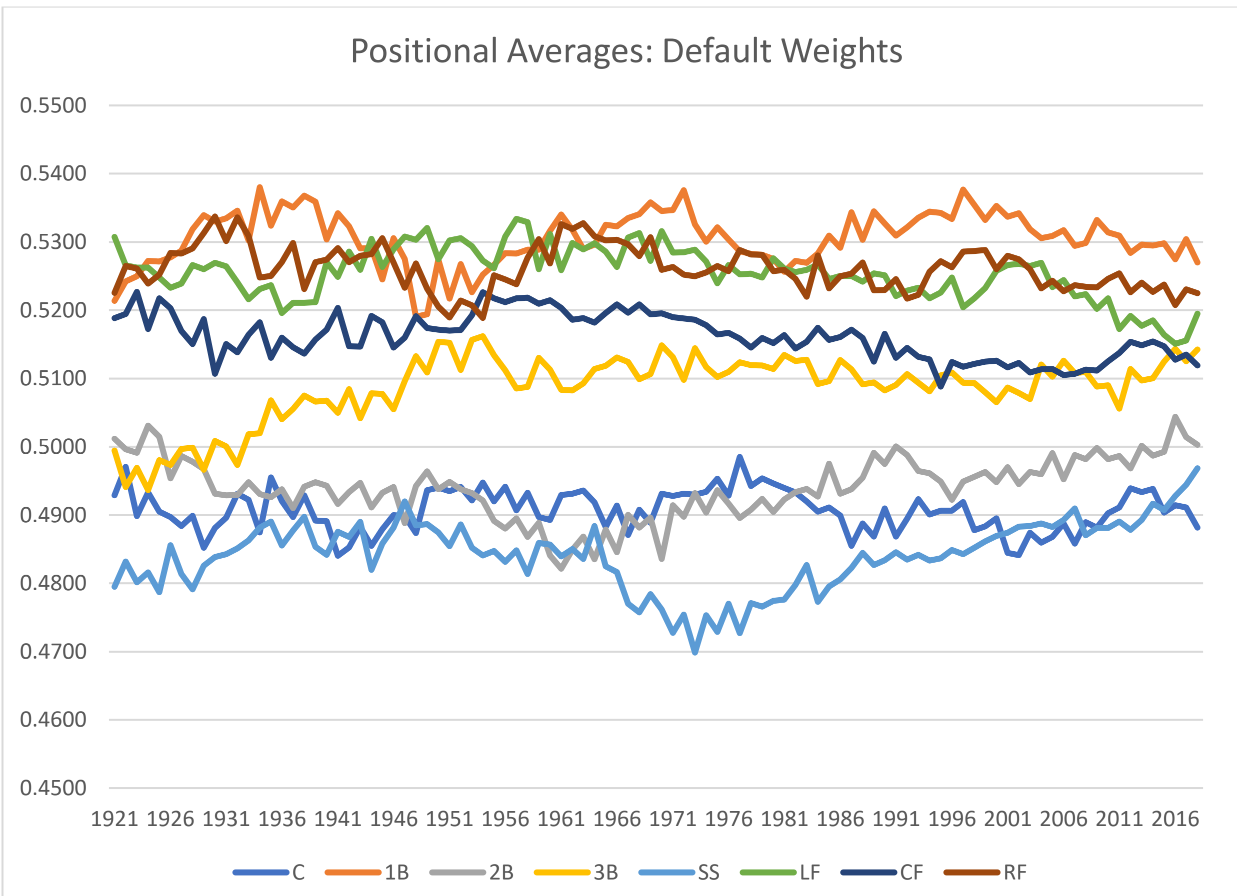
Finally, here are some examples of the potential impact of one's positional average choices.

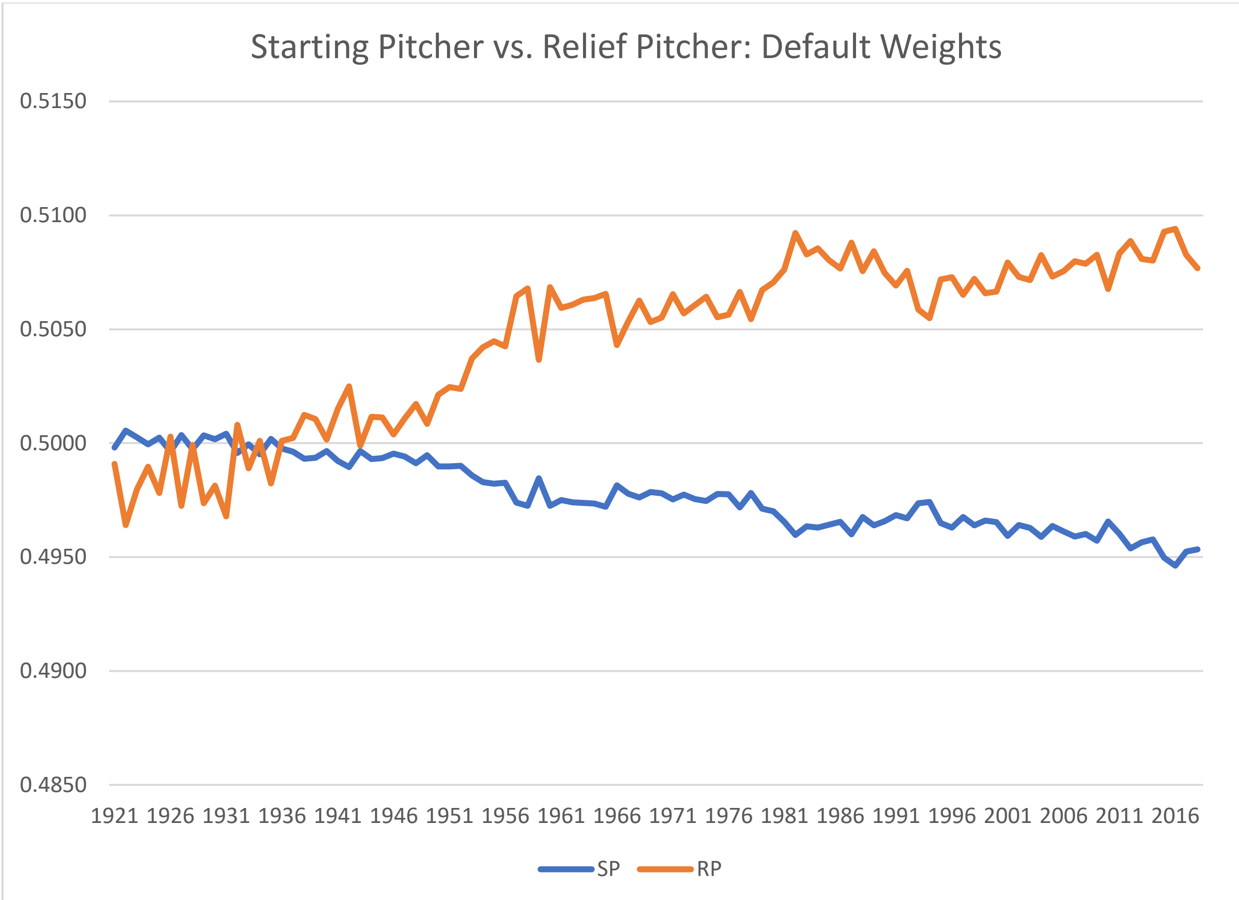
- Selecting Positional Averages: Some Examples and the Potential Impact

The default weights for positional averages on the website, if one does not choose one's own, are to give equal weight (one-third) to each of the one-year, long-run, and nine-year positional average. Working through some math, that works out to giving about 37.4% weight to the current year (approximately 3/8), 4.0% weight to the four years on either side of the current year, and 0.34% weight (one-third of one percent) to all other seasons for which I have calculated Player won-lost records.

For pitchers, the default gives equal weight to the three options for treating starting and relief pitchers: 0.500, empirical, and differences based on pitchers who do both. The one-year, nine-year, and long-run versions of these three calculations are weighted the same as for non-pitchers.

The resulting positional averages are shown in the next two graphs. For the fielding positions, the numbers shown here are offense only in non-DH leagues.





As you can see, there is still considerable variation, both in terms of long-run peaks and troughs as well as in terms of year-to-year variation. To reduce the latter of these, one could reduce the weight on one-year positional averages; to reduce the former, one could reduce the weight on nine-year positional averages. It's completely up to you.

I wanted to end this section, then, with a few player comparisons where one's choice of positional averages makes a difference. The comparisons made here are based on eWins, which are context neutral, so as to avoid potential issues with different team contexts in which these players played – i.e., I want to make sure that the issue being highlighted here is positional average and only positional average.

- Oakland A's Shortstops: Bert Campaneris vs. Miguel Tejada

Who was the greatest shortstop in the history of the Oakland A's? It is possible that I am forgetting somebody, but I think there are only two possible answers to this question (I think this is true whether one limits oneself to the Oakland A's or includes the Philadelphia and Kansas City A's in one's answer).

Dagoberto Campaneris debuted in 1964 with the Kansas City A's. He played the first 13 seasons of his 19-year career with the A's, moving from Kansas City to Oakland with the team in 1968, and was the starting shortstop for the A's five consecutive AL West division champions and three consecutive World Series champions from 1971 through 1975. Campaneris was named to six All-Star teams, starting three of them. Campaneris never won a Gold Glove, but was, I believe, well regarded defensively and scores as above-average defensively over his career by most, if not all, defensive metrics. Campaneris had a career batting line of .259/.311/.342 with 79 career home runs (two of which he hit in his major-league debut), which looks unimpressive, although he played most of his career in a pitchers' park during some of the lowest run-scoring environments in major-league history. Still, even controlling for that, Campaneris was a below-average hitter over the course of his career. He was, however, an excellent baserunner, leading the AL in stolen bases six times with 649 career stolen bases (14th all-time). Adding his baserunning to his batting, Campaneris was about average offensively for his career, a bit above average through his prime.

Put all of that together and Bert Campaneris's career eWin – eLoss record was 288.7 – 282.7, a winning percentage of about 0.505.

Miguel Tejada debuted for the A's 33 years after Campaneris did, in 1997. He was then the A's starting shortstop for six seasons before leaving for the Baltimore Orioles in free agency. Tejada never won (or even played in) a World Series, but the A's made the playoffs in each of Tejada's last four seasons with the team. Miguel Tejada was probably not as good a fielder as Campaneris and he was definitely not as good a baserunner (he had 85 career stolen bases). But Tejada was a much better hitter. His career batting line was .285/.336/.456 with 307 career home runs. Certainly, Tejada played in much higher run-scoring environments than Campaneris, but even controlling for that, Tejada was clearly the better hitter of the two. Tejada matched Campaneris in All-Star selections with six, although five of Tejada's six All-Star selections came after he left Oakland.

For his career, I have Miguel Tejada's career eWin – eLoss record at 278.5 – 273.1, a winning percentage of about 0.505.

Remarkably close. And you can see how one's choice between these two players is likely to come down to one's choice of positional average.

Before going any further, I should point out that Tejada spent a bit more time at positions other than shortstop. For his career, Tejada played 163 games at third base and 27 games as a DH. Overall, Tejada played shortstop in 90.9% of his career games played and about 91.5% of his career defensive innings. In contrast, Campaneris played shortstop in 91.9% of his career games and 93.9% of his career defensive innings. Not a huge difference, but an "advantage" to Campaneris.

Over the 98 seasons for which I have calculated Player won-lost records, the long-run positional average for shortstop (offense-only, non-DH leagues) has been 0.484. Using long-run positional averages, Tejada beats Campaneris in career eWins over positional average (eWOPA), **10.4 – 10.0**. Very close – as you'd expect given their essentially identical career eWin percentages – but **advantage, Tejada**.

Single-season positional averages for shortstop have ranged from a low of 0.456 to a high of 0.507. The former of these occurred in 1973, when Bert Campaneris played 151 games for the World Champion A's. Campaneris played 100 or more games at shortstop for 13 consecutive seasons, from 1965 through

1977 (he was a Texas Ranger in the last of these seasons). The highest one-year positional average in any of these 13 seasons was 0.482 (in 1965). In other words, the single-season positional average for shortstop was below the long-run average every season of Bert Campaneris's career as a regular shortstop. The average one-year positional average for shortstop over these 13 seasons was 0.471.

Miguel Tejada played 100 or more games at shortstop for 12 consecutive seasons, from 1998 through 2009. The lowest one-year positional average for shortstop for any of these 12 seasons was 0.485 (in 1998). The single-season positional average for shortstop was above the long-run average in every season in which Miguel Tejada was a regular shortstop. The average one-year positional average for shortstop over the 12 seasons in which Tejada was a regular shortstop was 0.490.

Using one-year positional averages, the positional average for shortstops during Miguel Tejada's career was nearly two percentage points higher than during Bert Campaneris's career. Which leads to Bert Campaneris beating Miguel Tejada in career eWOPA, using one-year positional averages, **14.6 – 8.6. Clear advantage, Campaneris**.

So, who was better, Campaneris or Tejada? You decide.

- Gil Hodges vs. Rafael Palmeiro

Gil Hodges is something of the patron saint of one-year positional averages. I already discussed him earlier comparing him to Todd Helton. But he makes such a good example that I decided to do a second comparison, this time with Rafael Palmeiro (who I also mentioned briefly earlier in this essay).

Gil Hodges came closer to making the Hall of Fame than any other non-Hall-of-Famer in major-league history. Hodges' Hall-of-Fame case essentially rests on two pillars: he was the manager of the 1969 Miracle Mets and he was the best first baseman of the 1950s. In many ways, this is similar to the Hall-of-Fame case of Jack Morris whose case rested on Game 7 of the 1991 World Series (and pitching for two other World Series winners) and being the best pitcher of the 1980s.

The big difference between the cases of Hodges and Morris, in my opinion, is that Hodges really has an argument for being the best first baseman of the 1950s. The problem, though, is that the second-best first baseman of the 1950s was either Stan Musial – who was clearly better than Gil Hodges (and almost everybody else in the major leagues) but only played 100 games at first base in 4 of the 10 seasons of the 1950s (and only played 103 and 110 in two of those seasons) – or Ted Kluszewski – who was a very good hitter but who received only 3.1% of the vote in his first year on a Hall-of-Fame ballot, which (a) would have gotten him kicked off the ballot under modern rules (for some reason, Kluszewski stayed on the ballot for 15 seasons, peaking at 14.4% of the vote), and (b) seems about right for his career: he was really good for four years, pretty good for a few more, but that was about it.

Anyway, Gil Hodges played first base at a time when there were not a lot of other great (or, arguably, even good) first basemen. The question is, how much credit should he get for that?

I calculate Gil Hodges to have had a career eWin – eLoss record of 219.1 – 179.1, a winning percentage of about 0.550 which is quite good, although it's worth remembering that he played his entire career in non-DH leagues (which boosts the winning percentage of non-pitcher hitters).

Setting aside the steroid issue, Rafael Palmeiro has much better career statistics than Gil Hodges. Palmeiro is one of four players with at least 3,000 career hits and 500 home runs (3,020 and 569, respectively). Hodges had fewer than 2,000 hits and fewer than 400 home runs (1,921 and 370). And simply looking at eWins and eLosses, the difference is clear. Palmeiro had a career eWin – eLoss record of 310.7 – 263.9, a winning percentage of about 0.541. This is a bit lower than Hodges' primarily because Palmeiro played most of his career in a DH league. As I discussed somewhat earlier, offensive winning percentages for non-pitchers tend to be around 0.010 lower in DH leagues than in non-DH leagues.

But Rafael Palmeiro played in an era where it seemed like every team had a first baseman who could either hit 0.300 (e.g., John Olerud, Mark Grace, Sean Casey), hit 30 home runs (e.g., Mark McGwire, Fred McGriff, Carlos Delgado), or do both (e.g., Jeff Bagwell, Frank Thomas, Mo Vaughn).

Gil Hodges was named to the All-Star team eight times in his career and won three Gold Gloves. Rafael Palmeiro was named to the All-Star team only four times in his career as he was competing against a much deeper pool of good-hitting first basemen. Palmeiro also matched Hodges with three Gold Gloves, one of which he rather infamously won in a season in which he only played 28 games at first base (1999). Palmeiro almost certainly didn't deserve that Gold Glove but in a way, it highlights the issue here: in the late 1990s, everybody was focused on finding big hitters to play first base, leaving less room for some of the weaker hitting glove-first guys who played during Hodges' career (e.g., Vic Power, who I talked about earlier).

From 1948 – 1961, the seasons in which Gil Hodges was a major-league regular, the average single-season positional average (offense only, non-DH league) for first basemen was 0.524. From 1988 – 2005, 0.535. The long-run positional average for first basemen across all seasons for which I have calculated Player won-lost records is 0.531.

Using long-run positional averages, Rafael Palmeiro beats Gil Hodges in career eWOPA, **13.5 – 10.4**. Using one-year positional averages, Hodges beats Palmeiro, **12.9 – 11.7**.

- Richie Ashburn

My last player example is not a comparison but a look at a single player, Richie Ashburn. Richie Ashburn was the center fielder for the Philadelphia Phillies primarily through the 1950s. This means that Ashburn played center field during an era in which there were only 16 major-league teams, three of whose center fielders were Willie Mays, Mickey Mantle, and Duke Snider.

Needless to say, having 20% of the other center fielders in the league be Hall-of-Famers (actually, slightly more than one-quarter, as Larry Doby was also a center fielder during this decade) – two of them arguably among the top 10 players in major-league history – will affect the positional average for center field. The long-run positional average for center field (offense only, non-DH league) is 0.516. From 1948 – 1962, Richie Ashburn's career, the average single-season positional average for center field was 0.522. From 1954 – 1960, the average single-season positional average was 0.526.

I have calculated Richie Ashburn's career eWin – eLoss record at 283.5 – 272.8, a winning percentage of about 0.510. Using one-year positional averages throughout his career, this translates into −3.1 career eWOPA for Ashburn. Using long-run positional averages, Ashburn's record improves to −0.6 eWOPA.

Why have I chosen this example? For this reason. Yes, Richie Ashburn looks a fair bit worse when measured against single-season positional averages than he would if measured against long-run positional averages. But that is not the primary reason why Richie Ashburn fares poorly in Player won-lost records. The fact is, Player won-lost records are somewhat unimpressed by Richie Ashburn. The Philadelphia Phillies of the 1950s allowed a very high number of doubles and triples and, as the team's center fielder, Ashburn bears some of the blame for that. From 1954 – 1959, the Phillies led the National League in doubles allowed four times, in triples allowed four times, and in combined doubles and triples allowed four times. In 1960, Ashburn moved to Chicago and the Cubs allowed 67 triples to lead the National League, 27 more than the Cubs had allowed in 1959. Player won-lost records are also somewhat unimpressed by Ashburn's offensive profile – specifically, his lack of power (29 career home runs, career slugging percentage of 0.382).

If Player won-lost records are missing something about Ashburn's career, it's hard to see that it translated into real wins. Ashburn has 4.4 fewer pWins than eWins in his career.

My point with this example is not to denigrate the career of Richie Ashburn – he was an above-average major-league center fielder in his 8-year prime (regardless of the choice of positional average) – but to emphasize that you can't use positional averages to get wherever you might want to go. There is an underlying objective reality embedded in the eWins and the pWins (and the eLosses and the pLosses) and not ***every*** case of a player who looks unexpectedly good or bad is the result of subjective analytical choices. There are, I believe, real truths to be revealed by Player won-lost records.

By letting people choose their own positional averages (within limits) as well as some other choices which will be discussed next, I hope to ensure that people will be more accepting of these "real truths" when they arise.

- **Player Comparisons: Beyond Positional Averages**

Many people (including me) like to distill player values down to a single number, so that they can create lists and rankings of players. The ability to express player values in a single number is a frequent feature of Hall-of-Fame debates, MVP discussions, and trade evaluations. It forms the core of putting together alternate Halls-of-Fame.

Despite my affinity for this type of list-making and ranking, I think the real value of my player won-lost records is the fact that they do not simply present a single number and leave it at that. Looking at all of the underlying numbers - wins and losses, contextual and context-neutral, comparing to positional averages and replacement levels, broken down by component - helps one to better put a player's value within the context in which it was accumulated. I also think that a weakness of some of the most prominent “single-number” statistics out there – WAR, Win Shares – is precisely that they distill everything into a single number, which forces one to accept the subjective adjustments that go into creating that number – some of which may not be immediately obvious or well explained, with the result, in the case of WAR, that you can go to (at least) three different websites and find three different statistics called “WAR” which give three (sometimes very) different answers.

Having said that, I think there can be a place for trying to condense everything down to one single number. And when condensing everything down to one number, I think there's a lot to be said for allowing as much flexibility as possible, letting people construct their one number however they want to. I have facilitated this in three locations on my website.

To calculate an “uber-statistic” for a single player, click on the “Value Decomposition” link on the player’s basic player page. This page allows one to enter a set of “uber weights” which are combined into a table of player-specific uber-statistics by season in the final table of this page.

To compare two players in a user-specified “uber-statistic”, one can click on the “Uber-Statistic” option of the Player Comps page (which can be accessed either through the home page of my website or via any player page). This brings up a special version of the Player Comps page which allows one to specify a set of uber-weights. The two players chosen are then compared by “uber-statistic” by season.

Finally, I have a page, accessible through the “Historical Leaders and Trailers” link on my home page, which creates a ranking of the top N players (where N is chosen by the user) over a user-specified time period based on a user-specified uber-stat.

There is an article on my website at https://baseball.tomthress.com/Articles/UberStats.php which looks at the factors which the user can choose to help construct this uber-statistic. The remainder of this essay looks at several of the primary sets of factors which go into the construction of these uber-statistics.

- pWins vs. eWins

I calculate two measures of Player won-lost records: pWins & pLosses - which are tied to team wins - and eWins & eLosses - which control for the quality of a player's teammates and the context in which he performed.

The choice between pWins and eWins will likely depend on one's purposes in putting together a list. One could think of pWins as measuring what actually happened, while eWins perhaps measure what should have happened. Personally, I think both of these measures provide us with useful and interesting information. Because of this, my default weights are equal to 0.5 each for both pWins and eWins.

- Wins vs. WOPA vs. WORL vs. WO*
    - Wins

Player wins are calculated such that the players on a team earn 3 Player decisions per game. I calculate two sets of Player wins: pWins are tied to team wins - the players on a winning team earn 2 pWins and 1 pLoss, the players on a losing team earn 1 pWin and 2 pLosses - while eWins attempt to control for the context in which they were earned, as well as controlling for the abilities of a player's teammates.

Player wins end up being on a similar scale to traditional pitcher wins: 20 wins is a good season total, 300 wins is an excellent career total.

- Wins over Positional Average (WOPA)

Positional average was the subject of the first section of this essay.

- Standard Deviations

After calculating positional average, I calculate one additional number for players, a standard deviation. Standard deviation is a function of average, so standard deviation will vary based on one's choice of positional average. Specifically, standard deviations are calculated over the same time period as positional averages, so, for example, if one uses nine-year positional averages, the standard deviations will also be calculated over a moving nine-year average. Final standard deviations are then equal to the weighted average of standard deviations, using the same weights as are used in calculating positional averages.

I calculate separate standard deviations for pitchers and non-pitchers. For pitchers, if separate positional averages are calculated for starting pitchers and relief pitchers, I also calculate separate standard deviations for starters and relievers. If a single positional average of 0.500 is used for starting and relief pitchers, a single standard deviation would be used for both starting pitchers and relief pitchers.

For non-pitchers, I distinguish between fielding positions and non-fielding positions. The latter have a higher standard deviation because there is no opportunity for such players to make up for poor hitting with strong fielding. In the process of introducing the changes discussed in this essay, I experimented with different standard deviations for catchers versus non-catchers. This tended to produce higher standard deviations for catchers than for the other fielding positions, which translated into somewhat higher wins over replacement level (WORL). I abandoned that and reverted to a single standard deviation for non-pitcher fielding positions, however, as I felt like the higher standard deviation for catchers was not really accomplishing what I had hoped it would (which was to get catchers' numbers closer to non-catchers in aggregate). Hence, for non-pitchers, standard deviations are calculated for two groups: fielding positions and non-fielding positions. Differences between fielding positions (e.g., catchers versus non-catchers) can be incorporated via position weights which are discussed later in this essay.

Standard deviations are then used to create the final two measures of player value which are presented on player pages and which may be weighted in constructing uber-statistics.

- Wins over Replacement Level (WORL)

Replacement Level is the level of performance which a team should be able to get from a player who they can find easily on short notice - such as a minor-league call-up or a veteran waiver-wire pickup. The theory is that major league baseball players only have value to a team above and beyond what the team could get from basically pulling players off the street. That is, there's no real marginal value to having a third baseman make routine plays that anybody who's capable of playing third base at the high school or college level could make, since if a major-league team were to lose its starting third baseman, they would fill the position with somebody and that somebody would, in fact, make at least those routine plays at third base. This is similar to the economic concept of Opportunity Cost.

For my work, I define Replacement Level as equal to a winning percentage one standard deviation below Positional Average. This works out to an average Replacement Level of about 0.444. A team of 0.444 players would have an expected winning percentage of 0.331 (54 - 108 over a 162-game season).

- Wins over Star (WO*)

In my second book *Baseball Player Won-Lost Records: 150 Players, 50 Years* (Amazon, 2018) For my Uber-Stat, I introduced a fourth potential measure of player value, WO*, or Wins over Star Level. "Star Level" is set equal to one standard deviation above positional average: the mirror image of replacement level. That is, WORL will give value to players who are below-average but not necessarily terrible, whereas WO* will only give value to players who are not merely above average but are "stars". The purpose of introducing WO* is to provide a means of giving players credit for their peak performance above and beyond the extent to which a strong peak translates into strong career numbers.

I also give the option of zeroing out negative seasonal values of WOPA, WORL and/or WO*. In my default calculations, I zero out negative values for all three of these.

I would be extremely cautious about not zeroing out negative values for WO* as the vast majority of player-season WO* values are negative and such values are likely to overwhelm a player’s positive values in other measures. Also, be aware that multiplying negative numbers by numbers greater than one will result in values that are more negative. So, for example, if one were to set one or more positional weights greater than one – say, catchers, who have shorter careers than players at other positions – allowing negative WO* values (or WOPA or WORL values, although these are, of course, rarer) would actually penalize catchers. I have thought about simply not allowing one to consider negative WO* values (and, for example, I do not show them on the player pages on the website) but decided to leave it as a user choice. But it is one that I would strongly argue against using.

- Distribution of Wins vs. WOPA vs. WORL vs. WO*

The table below looks at how the scales of these four statistics compare over the past fifty years (1969 – 2018, with negative single-season values zeroed out).

Distribution of Wins vs. WORL vs. WOPA vs. WO*
(1969 - 2018)

Total for Player #

	1	100	200	300	500	1,000	Total
Wins	462.3	254.8	216.0	189.0	153.3	104.3	329,897.8
WORL	98.7	37.2	29.4	24.7	19.3	12.2	39,631.7
WOPA	61.4	15.5	11.2	8.8	6.3	3.4	12,036.6
WO*	27.1	2.9	1.9	1.2	0.7	0.3	1,571.0

To explain, briefly, the numbers in the first column are the values of the four statistics which were earned by the player with the highest total of each of these four statistics over this time period. The second column is the values of the four statistics which ranked 100th over this time period, etc. Note that the player who ranked 100th in Player wins (the numbers here are calculated based on an average of pWins and eWins) may not have been the same player who ranked 100th in WO* - although the first column is entirely Barry Bonds, who ranks first in all four measures over this fifty-year period. The final column shows the totals across all player seasons over these fifty seasons (again, with any negative WORL, WOPA, and WO* seasons zeroed out).

The next table, then, normalizes the numbers in the previous table by dividing the numbers in each column by the value of WORL in the relevant column.

Distribution of Wins vs. WORL vs. WOPA vs. WO*
(Normalized Relative to WORL)

Total for Player #

	1	100	200	300	500	1,000	Total
Wins	0.2135	0.1461	0.1363	0.1306	0.1259	0.1174	0.1201
WORL	1.0	1.0	1.0	1.0	1.0	1.0	1.0
WOPA	1.61	2.39	2.62	2.80	3.08	3.63	3.29
WO*	3.6	12.8	15.8	20.2	27.7	44.5	25.2

In my second book, *Baseball Player Won-Lost Records: 150 Players, 50 Years* (Amazon, 2018), I chose a set of weights based on (a slightly different version of) the preceding table that attempted to balance the value of all four values. Specifically, I chose weights such that the 300th-best career total in the four statistics (giving equal weight to pWins and eWins and zeroing out negative seasonal numbers for WORL, WOPA, and WO*) over the 50-year period I looked at in the book (1961 – 2010) would end up having the same value.

In looking at this table, however, one thing that is striking to me is that the drop-off from 1st to 100th to 500th, etc., gets sharper the higher the baseline against which the statistic is measured. So, for example, the 1,000th-most wins are 22.5% as large as the most wins, but the #1,000 figure for WORL is only 12.4% as large as the #1 figure. For WOPA, the #1,000 figure is only 5.5% as large as the #1 figure – and this is even with zeroing out negative WOPA values. For WO*, again, even zeroing out negative values, the #1,000 figure, 0.3, is only equal to 1.1% of the top figure (27.7).

I interpret this to mean that one should be cautious about how one weights these values, especially WO*. If one thinks of the numbers in the above table as candidate weights, the range of possible weights for WO* is huge and one's choice from between weights of, say, 3.6 and 44.5 could have a huge impact on where some players may end up ranking.

I would also caution that, while WOPA and WORL are, in some sense, "real" measures of player value – this is how much more valuable a player was than an average player or a "replacement-level" player (granting that the latter is a theoretical concept) – WO* really is not. It is not clear to me that there is actual marginal value in having a player who is a "star" versus a player who is not a star. To be perfectly honest, I created WO* when I was putting together my second book as a way to boost the ranking of Sandy Koufax. Without question, at his best, Sandy Koufax was among the most valuable players in the history of major-league baseball. But I am not entirely sure that WO* is the best way to measure this.

I would also caution against weighting raw wins too heavily. While it is certainly true that wins (specifically pWins) are the object of the game, I think that it is important to put wins into some kind of context that also takes player losses (and player position) into account.

- Combining Wins, WOPA, WORL, and WO*

The choice between wins, WOPA, WORL, and WO* will likely depend on exactly what one is looking for. And there's no reason to limit oneself to just one of these three.

As I just said, I have some reservations about using raw wins and WO* as measures of player value. For this reason, the default weights on the website are zero for wins and WO* with weights of 1.0 for replacement level and 3.0 for wins over positional average. But, of course, the point of the exercise is that you can choose whatever weights you would like.

Let me look briefly at a couple of other weighting schemes that may have some appeal.

- Changing Replacement Level

It may be the case that somebody dislikes my choice for replacement level. One could approximate an alternate replacement level by weighting Wins, WOPA, and/or WORL.

My replacement level works out to around 0.444. Wins over positional average (WOPA) works out to 0.500 on average. Knowing this, we can make two general statements:

(1)	WORL - WOPA	=	.056*(Player Decisions)
(2)	Wins - WORL	=	.444*(Player Decisions)

Suppose you wanted to set replacement level at 0.472. You want to add 0.028*(Player Decisions) to WOPA, or, from (1) above: 0.5*(WORL - WOPA), i.e.,

Wins over 0.472 = WOPA + 0.5*(WORL - WOPA) = 0.5*WOPA + 0.5*WORL

Suppose you wanted to set replacement level at 0.344. You want to add 0.100*(Player Decisions) to WORL, or, from (2) above: (.1/.444)*(Wins - WORL), i.e.,

Wins over 0.344 = WORL + (.225)*(Wins - WORL) = (.775)*WORL + (.225)*Wins

I should point out one thing here, which will become more relevant below. In my work, player-level replacement level is not the same as team-level replacement level. This is because I award three player decisions for each team game – two pWins and one pLoss for a team win, one pWin and two pLosses for a team loss. To calculate a team's record, then, from the sum of its player's pWins, one has to subtract out what I call the "background" wins – the one pWin that the players on a team earn just for showing up and playing the game.

Doing a little algebra, one can convert from a player-level replacement level to a team-level replacement level by subtracting one-third (0.333) from the player replacement level and multiplying the result by three. And, conversely, one could convert from team-level replacement level to player replacement level by dividing the team replacement level by three and adding one-third (0.333).

- Replicating WAR

Probably the most popular "uber-stat" for measuring baseball players' value is Wins above Replacement (WAR). Measures of WAR are presented on player pages at Baseball-Reference.com (which I call bWAR here), Fangraphs.com (which I call fWAR here), and TheBaseballGauge.com (which I don't discuss any further here, but which could be called gWAR). I have compared my Player won-lost records to WAR in two articles on my website, the first of which was published as an essay in the Fall 2016 issue of SABR's *Baseball Research Journal* (Volume 45, number 2) and as a chapter in my first book, *Player Won-Lost Records: Measuring Performance in Context* (McFarland, 2017).

As I discussed at that time, eWins over positional average (eWOPA) does not relate to team wins over .500 on a one-to-one basis (as bWAA and fWAA do), but instead on a two-to-one basis, i.e.,

bWAA ~ 2*eWOPA

But the difference between WAA (Wins above Average) and WAR (Wins above Replacement) and the difference between WOPA (Wins over Positional Average) and WORL (Wins over Replacement Level) are on the same scale, i.e.,

(bWAR - bWAA) ~ (eWORL - eWOPA)

Hence, one can calculate the pWin or eWin-based equivalent of WAR by adding WOPA plus WORL.

Except for one detail: bWAR and fWAR are both calculated setting team replacement level at .294 (approximately 48-114 over a 162-game season). In contrast, my player-level replacement level is approximately 0.444, which translates into a team-level replacement level of 0.331. Converting eWORL from a team-level replacement level of 0.331 to .294 (to match bWAR and fWAR) can be done following the algebra laid out in the previous section. Skipping the gory details, a team-level replacement level of 0.331 translates into a player-level replacement level of 0.431, and converting from a player replacement level of 0.444 to 0.431 is done via the following formula:

$$\text{Wins over } 0.431 = 0.0276*\text{Wins} + 0.9724*\text{WORL}$$

Combining the earlier result, then, that WAR ~ WOPA + WORL, we can get the Player won-lost version of WAR - eWAR - by fitting the following formula:

$$\text{"eWAR"} = 0.0276*\text{eWins} + \text{eWOPA} + 0.9724*\text{eWORL}$$

One could do this with either eWins or pWins (or a weighted average of the two) and, of course, the final value of eWAR (or pWAR) will depend on the positional averages that one chooses.

The next table shows some “WAR” values calculated using eWins and either one-year, nine-year, or long-run positional averages as well as values using pWins and long-run positional averages and compares these to the WAR values shown by Baseball-Reference and Fangraphs. The players shown here were semi-randomly chosen.

	bWAR	fWAR	$eWAR_1$	$eWAR_9$	$eWAR_{l-r}$	$pWAR_{l-r}$
Barry Bonds	162.8	164.4	170.5	170.9	170.6	169.3
Greg Maddux	106.7	116.7	129.6	129.5	123.1	121.4
Joe Morgan	100.6	98.8	123.8	124.2	119.4	128.5
Rickey Henderson	111.2	106.3	107.7	108.4	109.3	105.3
Eddie Murray	68.7	72.0	61.2	61.3	62.5	76.8
Tommy John	62.0	79.4	67.0	67.0	68.0	81.8
Andruw Jones	62.8	66.9	60.9	61.2	58.1	65.0
Jim Rice	47.7	50.8	47.1	46.9	48.2	54.9
Jack Morris	43.9	55.8	41.4	41.5	41.6	51.8

I don’t really have any points to make with any of these players, so I’ll just let the numbers speak for themselves.

- Differences by Position

The final set of weights which I allow in creating an uber-statistic are perhaps the most interesting. They also tie the most closely to the extensive discussion of positional averages which made up the bulk of this essay.

Specifically, I allow one to specify position-specific weights for the fourteen positions for which I create positional averages: the eight fielding positions, DH, pinch hitters and runners, pitcher offense, starting pitching, and relief pitching. Unlike positional averages, which treat offense and fielding differently, the positional weights chosen here will be applied to all of a player’s performance in proportion to the amount of time he played the particular position.

One's purpose in putting together a list may prompt one to weight some positions more highly than others, or even to exclude some positions altogether. For example, to produce a list of only position players, one could set the weights for starting and relief pitchers equal to zero. If somebody hates the designated hitter rule strongly enough, he or she can assign a weight of zero to the DH position.

For the most part, differences in value between positions should be accounted for by positional averages. There are a few caveats associated with that, however. For example, my positional averages are based on differences in offensive performance by players across different positions. Alternately, one could use differences in defensive performance by players who play multiple positions. As I discussed earlier in the essay, this produces a narrower range of positional averages – which can be mirrored by giving some weight to positional averages of 0.500 for all positions. It also, however, produces a result that suggests the aggregate winning percentage of first basemen and outfielders may be greater than 0.500 while the aggregate winning percentage of catchers and non-first-base infielders may be less than 0.500. I am skeptical of this approach for reasons which I already discussed. But if one wanted to incorporate this information in their analysis, they could add positive positional adjustments for first basemen and outfielders and/or negative adjustments for other fielding positions.

Even with positional adjustments, some people may want to make adjustments for some positions based on differences in career values across positions. The next table breaks down player decisions – based on pWins – by position in several ways. The numbers shown here span the most recent 50 years, 1969 – 2018.

Distribution of Player Decisions Across Positions

			Uber-Statistic	
Position	**pW + pL**	**pWORL**	**All Players**	**Top 1,000**
Catcher	5.9%	4.7%	5.2%	4.2%
First Base	6.9%	6.0%	6.3%	7.5%
Second Base	8.0%	7.1%	6.6%	6.6%
Third Base	7.6%	6.8%	6.8%	7.8%
Shortstop	7.9%	7.1%	6.5%	6.7%
Left Field	8.6%	6.5%	6.7%	7.0%
Center Field	8.3%	6.5%	6.4%	6.5%
Right Field	8.5%	6.8%	6.8%	7.8%
Fielding Positions (non-Pitcher)	**61.6%**	**51.6%**	**51.2%**	**54.1%**
Designated Hitter	2.6%	2.6%	2.7%	3.5%
Pinch Hitter	1.8%	0.9%	1.2%	0.8%
Pinch Runner	0.11%	0.04%	0.07%	0.02%
Offense-Only Positions	**4.5%**	**3.5%**	**4.0%**	**4.2%**
Pitcher Offense	1.3%	1.6%	1.6%	1.6%
Starting Pitcher	23.8%	27.8%	27.8%	31.0%
Relief Pitcher	8.8%	15.5%	15.5%	9.1%
Pitchers (incl. Offense)	**33.9%**	**44.9%**	**44.9%**	**41.7%**
Pitchers (Defense Only)	**32.6%**	**43.3%**	**43.3%**	**40.0%**

The first column shows the simple sum of pWins and pLosses. Pitchers earn about one-third of player decisions – which translates to about one decision per game, which puts pWins on the same scale as traditional pitcher wins. Differences across fielding positions are for two possible reasons: some fielding positions have more fielding decisions than others and some fielding positions are more likely to bat in lineup positions which generate more decisions. Outfielders benefit from both of these factors while catchers are hurt by both.

The second column shows pWins over replacement level. For this table, the positional averages used were my default weights (equal weights for one-year, nine-year, and long-run positional averages) although the results here are not overly sensitive to the choice of positional averages. Shifting to pWORL yields two worthwhile results. First, differences across fielding positions narrow – with the exception of catchers. Second, pitchers' share of pWORL is about 10 percentage points higher than their share of raw player decisions.

This latter result is due to two reasons. First, there is something of a multiplicative impact of player value on team victories: being a little bit above average helps a lot in producing team victories. This effect is more pronounced for pitchers than for position players, because pitchers concentrate their performance into fewer games. Second, pitchers are somewhat harder to replace than position players, which is reflected in a somewhat lower replacement level for pitchers (approximately 0.425) than for non-pitchers (approximately 0.453).

Excluding pitcher hitting, the percentage of wins over replacement level (pWORL) earned by pitchers, 43.3%, is virtually identical to pitchers' share of major-league payroll (e.g., 43.4% in 2014), which is encouraging to me as an economist, suggesting that my numbers here are generally reasonable.

The last two columns of the table calculate results for a default version of my uber-weights. These weight pWins and eWins equally, include postseason decisions, zero out negative single-season values, and use the default weights for wins, WORL, WOPA, and WO* that I specified earlier (0, 1, 3, and 0, respectively). These results will, of course, be dependent on the specific weights used although I think the results here are potentially insightful regardless of what weights one ends up using.

The first of these two columns (the third column overall) sums the uber-stats across all players. One could, perhaps, use these numbers to create some position weights. For example, catchers remain lower, in aggregate (5.2%), than other positions (the other seven fielding positions average 6.6%). A position weight for catchers of 1.25 (+25%) may be desirable, then, to raise catchers' representation to the unweighted representation among the other seven (non-pitcher) fielding positions.

The final column, then, looks at only the top 1,000 players. In the past, I have based my default position weights on a comparison of these last two columns. So, for example, catchers earn 5.2% of total "decisions" but account for only 4.2% of the top 1,000 players, suggesting, perhaps, a boost of perhaps 20-25% (i.e., a position weight of 1.20 - 1.25) to boost their representation among the top 1,000 players to mirror their unweighted representation among all players.

I believe that such adjustments could make sense, depending on one's purpose in constructing one's uber-statistic. I do, however, think that I had a tendency in the past to perhaps over-adjust across all positions. Hence, I have scaled back my default position weights, adjusting only for catchers and being somewhat conservative in my adjustments here with a default weight for catchers of 1.15.

Of course, the whole point of the uber-statistic pages on my website is that you can – and should – choose your own weights.

I hope you enjoy the new options. Thank you for reading and enjoy the website!