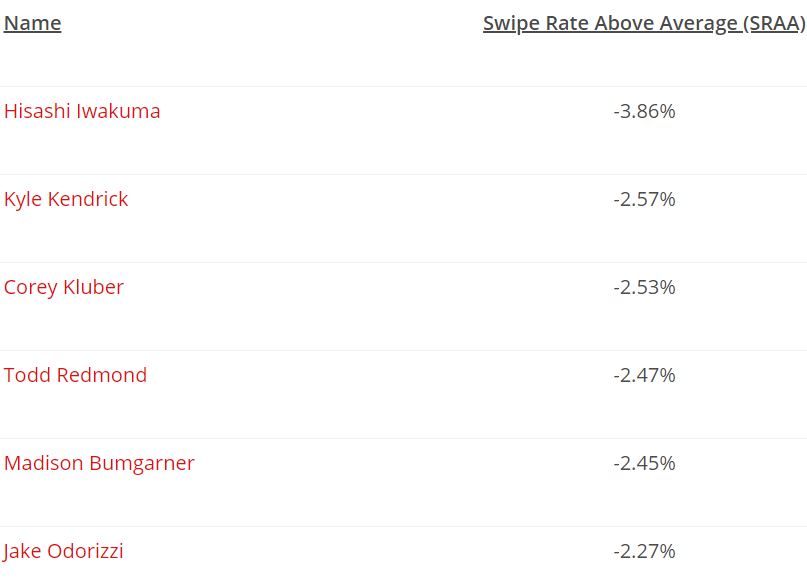
Deserved Run Average

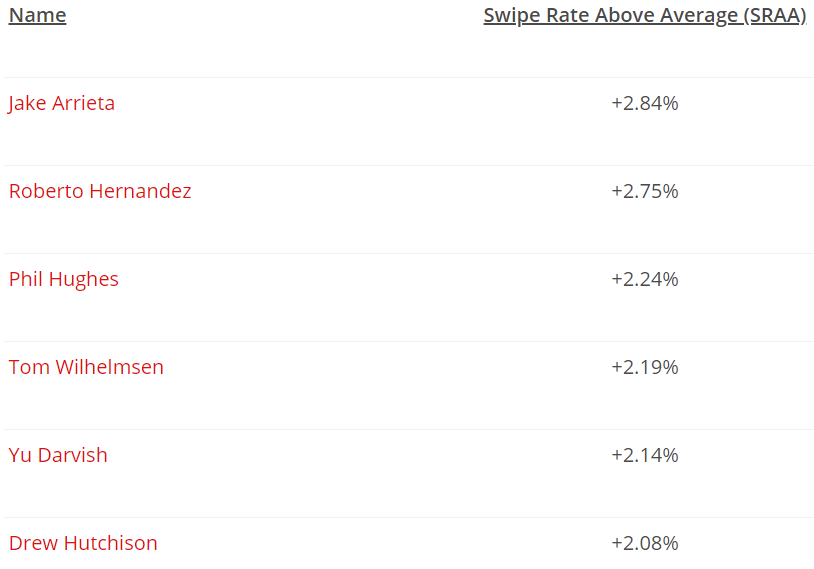
* *Prospectus Feature: Introducing Deserved Run Average (DRA) And All Its Friends*

<https://www.baseballprospectus.com/news/article/26195/prospectus-feature-introducing-deserved-run-average-draand-all-its-friends/>

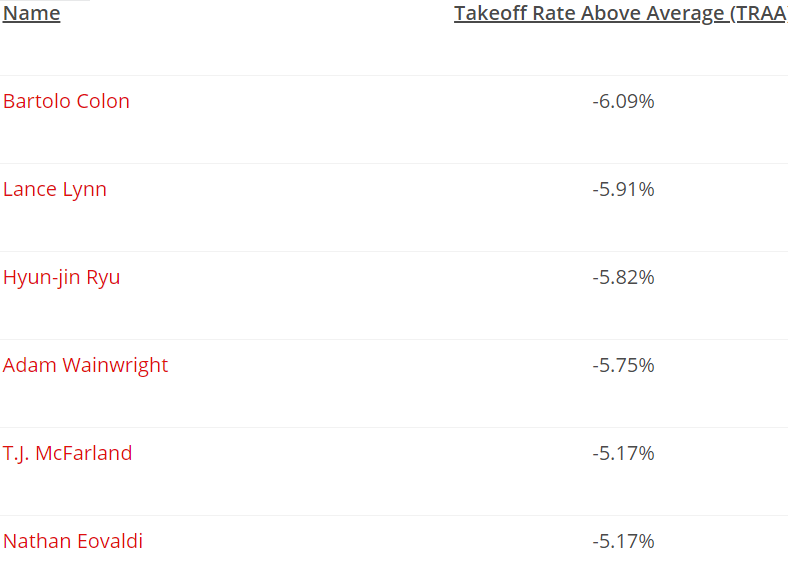
* **ERA** = benchmark by which pitchers have been judged for a century == runs pitcher gives up, on average, every 9 innings pitched
* If gave up a bunch of runs = probably terrible; if gave up very few = assume he’s pretty good.
* ERA problem = blames/credits pitcher for everything, simply b/c he threw the pitch that started the play
* Sometimes fair 🡺wild pitch, right down the middle (Not too many catchers request those)
* However, more often things happen = NOT pitcher’s fault at all = strikes umpire incorrectly calls balls, inducing grounders infielders aren’t adept enough to grab, routine fly ball leaves park on a hot night at a batter-friendly stadium
* *ERA doesn’t account for any of that* = just tells us, in summary fashion, how many runs were “charged” to the pitcher “of record.”
* Starting pitcher who departs w/ runner on 1st = charged w/ that run even if reliever walks next 3 batters + same starter gets charged if reliever makes a good pitch, but shortstop can’t turn a double play
* None of these runs count at all if “unearned”— an exclusion by which home team’s scorer decides whether a fielder demonstrated “ordinary effort.”
* *The list of problems goes on* 🡺 Pitchers who load bases but escape = treated same as pitchers who strike out the side, pitchers w/ great catchers get borderline calls, guys who can’t catch a break for months show immense “improvement”, average guys 1 year wash out the next.
* **ERA, in short, can be a bit of a mess, particularly when we have only a few months of data to consider.**
* The problem is this 🡺 We know which runs came across the plate, but **can’t tell, just from ERA, which runs were actually the pitcher’s fault.**
* Need a *reliable* way to determine which runs pitcher deserves to be charged w/ = **Deserved Run Average (DRA)**
* Baseball researchers have spent past few decades trying to figure out a better way to measure pitcher quality.
* Voros McCracken = popularly credited for discovering pitchers have varying (+ often little) control over results of balls **put in play.**
* Tom Tango proposed metric of **Fielding Independent Pitching (FIP) =** looks only at pitcher’s HRs, SOs, hit batsmen, + walks.
* From these 4 stats alone, **FIP** can account for ~50% of variance in RA by pitchers each year.
* *At same time, most plays in baseball do NOT involve a SO or HR, so many researchers have tried to improve on FIP’s formula.*
* Colin Wyers = **Labeled Fair Run Average** (“**FAIR RA**” or “**FRA**”) = tried to adjust for, among other things, what was considered to be a “fair” # of actual innings pitched + assigned a “fair” # of runs allowed for each pitcher as a result.
* **FRA** = not succeeded b/c while some of its assigned values make sense, others do not.
* Many researchers noted what appears to be bias in **FRA** against pitchers who generate a lot of groundballs (generally thought to be desirable)
* **FRA** = not caught on, +, more importantly, understanding of tools for measuring performance have advanced since the time **FRA** was conceived.
* New metric for evaluating pitcher’s responsibility for runs = **Deserved Run Average**, or **DRA**.
* Leveraging recent applications of “**mixed models**” to baseball stats, **DRA controls for context in which each event of a game occurred, thereby allowing more accurate prediction of pitcher responsibility, particularly in smaller samples.**
* DRA goes well beyond SO, walks, hit batsman, + HR, + considers all available batting events.
* DRA does NOT explain everything by any means, but its estimates appear to be more accurate + reliable than alternatives.
* As such, DRA allows us to declare how many runs a pitcher *truly* deserved to give up, + to say so W/ more confidence than ever before.
* DRA: Step by step:
* Step 1: Compile individual value of all baseball batting events in a season.
* Batter in box = # of different events can ultimately occur, from a SO to a single to a double play to a HR.
* Over course of a season, those events each, as a category, tend to result in some average # of additional (or fewer) runs.
* Ex: HR, on average, results in ~1.4 runs, b/c of runners on base or not
* By the same token, a DP tends to cost a team ~3/4 of a run (can sometimes allow a run to score = runner on 3rd w/ no outs, but far more often ends innings/empties bases)
* In the world of baseballs stats, **average seasonal value** of these events = a “**linear weight**”
* To understand ultimate effect of batting events, 1st must assign **typical value** of those events
* So, DRA begins by collecting *every single baseball batting event in a given season* + assigning the **average linear weight** for the outcome of that play.
* Step 2: Adjust each batting event for its context.
* Once we have the **average value of each play** in a season 🡺 start making adjustments.
* HRs depend, among other things, on stadium, temp, + quality of opposing batter, ball + strike calls tend to favor home team, **likelihood** of a hit depends on quality of opposing defense, pitcher’s success depends on how far he is ahead in count, + both a catcher’s framing ability + size of umpire’s strike zone help get him there.
* So, DRA begins by adjusting for the **average effect** of *these factors beyond the pitcher’s control* in each plate appearance, using a **linear mixed model.**
* These environmental factors include:
* Overall friendliness of stadium to scoring, accounting for handedness of batter (using park factors from Baseball Prospectus);
* Identity of opposing batter + of catcher and umpire;
* Effect of catcher, umpire, + batter on likelihood of called strike (e.g., framing / umpire strike zone, from 1988 onward);
* Handedness of batter;
* # of runners on base + # of outs @ time of the event;
* Run differential between teams @ time of the event;
* Inning + also half of inning during which event is occurring;
* Quality of defense on field for each individual play (assessed through BP’s **FRAA** [1])
* Whether defense is home or away + Whether pitcher is home or away;
* Whether pitcher started game or is a reliever;
* Temp of the game at opening pitch (from 1998 onward).
* 2 other aspects that affect how DRA scores pitchers.
* 1st, rather than grade pitchers purely on # of outs (like ERA does), **DRA grades them on basis of each plate appearance.**
* Thus, pitchers who escape a bases-loaded jam = no longer treated same as pitchers who retire all 3 batters faced just b/c they both got 3 outs.
* 2nd; DRA judges pitchers on **run expectancy** of each play, rather than runs that happen to cross the plate.
* Ex: Starter put a man on 1st + was replaced, he would NOT be penalized *entire* run if reliever subsequently allowed that score.
* Rather, would be penalized **only by likelihood said player would’ve scored from 1st, on average, w/ reliever getting charged difference between average likelihood + full value of scores.**
* Likewise, when starter loads bases but reliever gets team out of it, *reliever doesn’t simply get credit for an out or 2 but gets a bonus for all runs that were expected to score from a bases-loaded situation in an average situation, but didn’t.*
* Here, true “stopper” relievers get more fairly recognized for accomplishments + we more accurately forecast their “deserved” runs allowed.
* DRA component that emerges from all these adjustments = **value/PA** = **average value of each plate appearance which pitcher *completed* during season.**
* Step 3: Account for base-stealing activity.
* Understanding avg. weight of a batting event = essential, but run-scoring also depends on who happens to be on base @ the time
* Some more likely to score when on base than others, all other things being equal, + certain pitchers hold runners better than others.
* Afraid of being picked off = fewer steal attempts, runners who stay closer to base should have harder time scoring, runners thrown out trying to steal = erased from basepath
* To account for these situations + provide some insight into effect of baserunning on each event, 2 additional stats created, looking @ base-stealing success + frequency w/ which baserunners attempt to steal = both (potentially) part of DRA, but also useful themselves
* B/c looking @ how pitchers compare to other pitchers in controlling baserunners, these stats are **Swipe Rate Above Average (SRAA)** + **Takeoff Rate Above Average (TRAA)**
* **SRAA** judges each participant in a base-stealing attempt for his likely effect upon its success
* Using a **generalized linear mixed model 🡺** simultaneously weigh all participants involved in attempted steals against each other + then determine **likelihood** of base being stolen, as compared to involvement of a **league-average** pitcher, catcher, or lead runner, respectively
* **SRAA** allows us to evaluate how good catcher’s arm is while controlling for inherent ability of pitcher to hold runners + quality of runners on base.
* Likewise, evaluate ability of individual pitchers to hold runners while controlling for possibility they may be throwing to a catcher w/ a subpar arm.
* For baserunners in particular, now have something much more accurate to evaluate base-stealing ability than base-stealing %.
* Base-stealing %, by itself = not very useful 🡺 using straight %’s, an elite base-stealer who swipes 90% of attempts + tries to steal 40 times a year ranks lower than a catcher who had 1 lucky steal all year (+ therefore has a 100% base-stealing %)
* In same way **Controlled Strikes Above Average (CSAA)** controls for effect of other factors on catcher framing, SRAA **regresses** baserunners’ steal-success rates against both themselves + others to provide a more accurate assessment of each participant’s effect on likelihood of a stolen base.
* Factors considered by **Swipe Rate** are:
* inning in which runner was on base; stadium for game; underlying quality of pitcher, as measured by Jonathan Judge’s **cFIP** stat; pitcher + catcher involved; lead runner involved.
* B/c SRAA rates pitchers above/below average in preventing SBs, average = 0, + pitchers generate either positive (bad) or negative (good) #’s
* 2014 pitchers who were hardest to steal a base on:



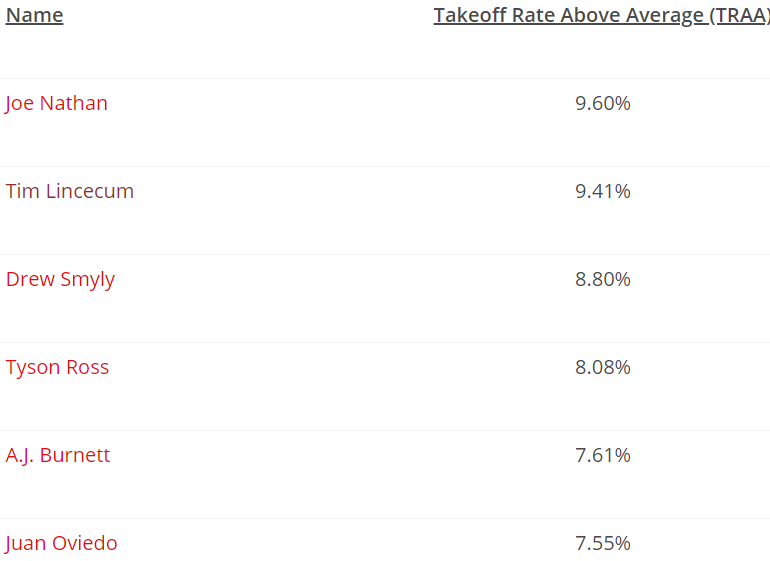
* 2014 pitchers that baserunners exploited the most:



* **TRAA (Takeoff Rate Above Average)** = similar to SRAA, but more complicated.
* W/ TRAA, don’t care whether baserunner actually *succeeds* in stealing, but care about *that he made an attempt*.
* **Hypothesis = stealing attempts = connected w/ pitcher’s ability to hold runners.**
* When baserunners are not afraid of a pitcher, they will take more steps off the bag. Baserunners who are further off the bag are more likely to beat a force out, more likely to break up a double play if they can’t beat a force out, and more likely to take the extra base if the batter gets a hit.
* Takeoff Rate stats consider the following factors:
* The inning in which the base-stealing attempt was made;
* The run difference between the two teams at the time;
* The stadium where the game takes place;
* The underlying quality of the pitcher, as measured by Jonathan Judge’s cFIP statistic;
* The SRAA of the lead runner;
* The number of runners on base;
* The number of outs in the inning;
* The pitcher involved;
* The batter involved;
* The catcher involved;
* The identity of the hitter on deck;
* Whether the pitcher started the game or is a reliever.
* Takeoff Rate Above Average is also scaled to zero, and negative numbers are once again better for the pitcher than positive numbers. By TRAA, here were the pitchers who worried baserunners the most in 2014.
* Pitchers who emboldened baserunners in 2014:
* Current 2015 ratings for Takeoff Rate Above Average are on our leaderboards. We don’t’ have enough data yet to release Swipe Rate Above Average, but we expect it will have enough to work with in another month or so.
* Step 4: Account for Passed Balls/Wild Pitches.
* Baseball’s scoring rules: **wild pitch** = pitch deemed too difficult for catcher to control w/ **ordinary effort**, allowing a baserunner (including a batter, on a 3rd strike) to advance a base.
* **Passed ball** = pitch a catcher ought to have controlled w/ ordinary effort, but nonetheless gets away, allowing a baserunner to move up a base
* Difference between wild pitch + a passed ball = @ discretion of official scorer = can be inconsistency in applying these categories, so *we prefer to consider them together.*
* Dan Brooks + Harry Pavlidis introduced a **regressed probabilistic model** that combined Harry’s pitch classifications from PitchInfo with a **With or Without You (WOWY)** approach.
* **RPM-WOWY** measured pitchers + catchers on # + quality of passed balls or wild pitches (**PBWP**) experienced while involved in the game.
* We’ve updated this approach to a **mixed model** as well w/ a new term to see if it is easier to communicate these concepts = **Errant Pitches Above Average (EPAA) = compares pitchers and catchers in these events**
* Unfortunately, this mixed model only works from 2008 forward, when **PITCHf/x** data became available.
* Before that time, will rely solely on WOWY to measure PBWP, which is when pitch counts were 1st tracked officially.
* For time being, won’t calculate EPAA before 1988 at all, + it will not play a role in calculating pitcher DRA for those seasons
* But, from 2008-2014, + going forward, factors that EPAA considers:
* identity of pitcher; identity of catcher; likelihood of pitch being an Errant Pitch, based on location + type of pitch, courtesy of PitchInfo classifications.
* Errant Pitches = much smaller list of relevant factors than other statistics.
* 2014, pitchers w/ best (most negative) EPAA scores:



* 2014 Pitchers model said were most likely to generate a troublesome pitch were:



* Step 5: Calculate DRA (Deserved Run Average).
* Now got components, so it’s time to calculate each pitcher’s DRA.
* Steps:
* 1) Put all identified components (value/PA, SRAA, TRAA, EPAA) together into a new **regression**, looking for their **combined effect** on **run expectancy**. [2]
* Added 2 more variables though to be relevant: % of each pitcher’s plate appearances that came as a starter vs. as a reliever (**Starter Pitcher Percentage**, or **SPP**) + **total # of batters faced.**
* 6 potential predictors for each pitcher to come up w/ DRA for a season.
* Regress these using **MARS method** (check In-Depth article)
* 2) To smooth out season-to-season variation + to tease out most accurate connection between these variables + RA, train model on the previous 3 seasons.
* From this we derive the most accurate connection between potential predictors + actual RA by pitchers in the current run environment.
* 3) Take connections determined by model + use them to calculate each pitcher’s DRA for current season = **Deserved Runs Average per 9 innings.**
* DRA does NOT distinguish between earned + unearned runs, b/c that distinction can be arbitrary + over the course of a season, tends to obscure rather than reveal differences between pitchers.
* Therefore adjust DRA so it’s on the scale of **RA per 9 innings** (**RA/9**) rather than ERA
* ERA is what many are used to, but once you get over that, you’ll be much happier.
* We do ensure that, in converting runs per plate appearance to runs per 9 innings, we use each pitcher’s individual ratio of batters-faced to innings pitched, rather than just a league average.
* This allows us to credit most efficient pitchers + avoid over-crediting pitchers getting lucky w/ baserunners.
* Latter pitchers category don’t “deserve” the lower RA #’s they might (temporarily) be putting up
* What It Means
* So there you have it: DRA, explained. Most of you really don’t care how we got there; you just care that DRA will be easy to look up and be a good evaluator of pitcher performance. In both respects, you are in luck.
* As for the first issue, past DRA is available on our leaderboards right now. In-season DRA during 2015 will be calculated each night after the previous day’s games have concluded. You will be able to use DRA not only to put past pitching performances in context[3] but also to monitor the value of pitchers as we progress through the 2015 season, and beyond. As with our other statistics, DRA will be available for you to download and use for your own comparisons and work.
* As for the second issue, rest assured that your time spent reading this article was not in vain. DRA does a very good job of measuring a pitcher’s actual responsibility for the runs that scored while he was on the mound—certainly better than any metric we are aware of in the public domain. And only DRA gives you the assurance that a pitcher’s performance is actually being considered in the context of the batter, catcher, runners on base, as well as the stadium and stadium environment in which the baseball game occurred.
* The detailed explanation of DRA’s effectiveness is saved for the accompanying In Depth article. But since you’ve made it this far, we’ll give you the Reader’s Digest version. There are two measures of accuracy that we pay particular attention to in evaluating the accuracy of a new metric.
* First, we look at how close, mathematically, the metric’s prediction is to the actual number of runs allowed with the pitcher on the mound. If the pitcher actually allowed four runs per nine innings, we test our alternative metric by how close it comes to that same number. The most commonly used calculation that does this is called the Root Mean Square Error or RMSE.
* The second test looks at how accurately the metric ranks the various pitchers relative to each other. Why do we care about rank? Because we know that all pitcher run estimates are a bit off from their actual runs allowed, and more so early in the season. So as a check, we test whether it is at least ranking the pitchers correctly relative to each other. In other words, if the metric can’t estimate runs allowed down to the exact decimal point, the least it can do is tell the difference between Max Scherzer and Ricky Nolasco. This second approach is called the Spearman Correlation.
* To judge DRA’s accuracy, we’ll compare it to the leading brand: FIP. We know FIP does a reasonable job of predicting a pitcher’s actual runs allowed in a season. Does DRA do a better job than FIP? It does.
* We compared how well FIP and DRA predicted each pitcher’s RA/9 in each of the past four major-league seasons. We looked at their performance with all pitchers, and then two subsets: pitchers who faced at least 170 batters (about the workload of an established major-league reliever, or 40 IP), and pitchers who faced at least 660 batters (about 162 innings, which is a qualified major-league starter).
* We then averaged the results over four seasons (2011­–2014) to get a consistent (and recent) picture of each metric’s performance. Here is how it ended up:
* Metric
* Minimum BF
* RMSE
* (lower is better)
* Spearman Correlation
* (higher is better)
* FIP
* 0
* 3.59
* 0.64
* 170
* 1.14
* 0.70
* 660
* 0.67
* 0.72
* DRA
* 0
* 2.65
* 0.76
* 170
* 0.96
* 0.76
* 660
* 0.54
* 0.78
* DRA is consistently superior to FIP at all sample sizes. By accounting for the context in which the pitcher is throwing, DRA allows us to determine which runs are most fairly blamed on the pitcher. DRA is particularly effective with smaller samples. Even for pitchers with only a few batters faced, DRA is already separating the good pitchers from the bad with superior accuracy.
* In the end, of course, we are not satisfied simply to have brought you DRA. In addition to being useful in and of itself, DRA has become the new foundation of Pitcher Wins Above Replacement Player (PWARP) here at Baseball Prospectus. By integrating DRA into WARP, we can do a better job than ever of evaluating how much value individual pitchers delivered to their teams, both during the current season and as compared to past pitchers in other seasons and eras. The new PWARP figures featuring DRA are also available on the leaderboards, under the column “DRA\_PWARP.”
* Just for fun, here are the 25 best qualified starters by DRA over the past 25 years. You’ll note that in some cases their DRA basically matches their RA/9; in others, it does not. Our position, of course, would be that when DRA and RA/9 disagree, you should go with DRA, as it tells you how well the pitcher really pitched. Without further ado:
* Rank
* Season
* Name
* DRA
* RA/9
* 1
* 2000
* Pedro Martinez
* 1.03
* 1.87
* 2
* 2004
* Jason Schmidt
* 1.23
* 3.30
* 3
* 1997
* Pedro Martinez
* 1.49
* 2.46
* 4
* 1995
* Greg Maddux
* 1.55
* 1.62
* 5
* 2004
* Randy Johnson
* 1.64
* 3.29
* 6
* 2009
* Zack Greinke
* 1.80
* 2.54
* 7
* 2009
* Tim Lincecum
* 1.87
* 2.79
* 8
* 2013
* Jose Fernandez
* 1.89
* 2.48
* 9
* 2013
* Max Scherzer
* 1.90
* 3.09
* 10
* 2013
* Matt Harvey
* 1.93
* 2.33
* 11
* 2013
* Clayton Kershaw
* 1.93
* 2.14
* 12
* 2007
* Erik Bedard
* 1.98
* 3.29
* 13
* 2011
* Justin Verlander
* 2.01
* 2.66
* 14
* 1997
* Roger Clemens
* 2.04
* 2.26
* 15
* 2004
* Johan Santana
* 2.05
* 2.79
* 16
* 1992
* Curt Schilling
* 2.11
* 2.63
* 17
* 1995
* Randy Johnson
* 2.14
* 2.79
* 18
* 2011
* Josh Beckett
* 2.15
* 3.07
* 19
* 2009
* Chris Carpenter
* 2.17
* 2.32
* 20
* 2003
* Pedro Martinez
* 2.17
* 2.52
* 21
* 2014
* Clayton Kershaw
* 2.18
* 1.94
* 22
* 2009
* Josh Johnson
* 2.20
* 3.38
* 23
* 1997
* Greg Maddux
* 2.23
* 2.29
* 24
* 1992
* Juan Guzman
* 2.27
* 2.84
* 25
* 2002
* Curt Schilling
* 2.27
* 3.24
* One caution: DRA is not (presently) adjusted for run-scoring across different eras. Rather, it is adjusted to the average runs-allowed by the league for that season. So, please don’t directly compare Pedro’s DRA of 1.03 in 2000 to somebody else’s DRA in 1985 or some other season.[4] A DRA metric that compares players across eras will be coming soon.
* A second caution: DRA corrects for what is known as survival bias: the tendency of better pitchers to pitch more innings in a season. Applying the full DRA model early on can result in some extreme values. To avoid that, we will keep the model simple at first during the season, and model only value/pa to RE24. As we get further along, we’ll allow the full model to operate and achieve the best explanation of each pitcher’s performance.
* Conclusion
* We are excited about DRA, as well the other statistics we have introduced: Swipe Rate Above Average (to measure base-stealing success), Takeoff Rate Above Average (to measure base-stealing attempts), and Errant Pitches Above Average (to measure passed balls and wild pitches).
* Three final things to remember.
* First, while DRA accounts for a great many things, DRA doesn’t need to be complicated for fans. DRA is on our leaderboards. Just look up the pitcher(s) that interest you, and you’ll have the best estimate of how good they’ve been in a particular season. If you want to leave the details to us, feel free.
* Second, remember that DRA was created to evaluate past performance. If you want to project future performance of a pitcher, use PECOTA. And if you want to evaluate how talented the pitcher is regardless of his performance to date, use cFIP, which is also on our leaderboards. In fact, cFIP is in the same table to DRA so you can compare recent results with the likelihood of future improvement (or decline).
* Finally, DRA is now the foundation for Pitcher Wins Above Replacement (PWARP) here at Baseball Prospectus. For the time being, if you want to see how many wins a pitcher has been worth in a particular season, check — you got it — our leaderboards and the column DRA\_PWARP. (WARPs that appear on pitchers' player pages remain, for now, the old, FRA-based WARPs.) We’ll change the description it to plain old WARP once people have gotten used to the new idea.
* We welcome your comments, and hope you find DRA as useful as we do.
* Special thanks to Rob McQuown for research assistance; to Rob Arthur, Rob McQuown, and Greg Matthews for their collaboration; to Stephen Milborrow for modeling advice; and to Tom Tango and Brian Mills for their review and insights.
* [1] Please note that FRAA (Fielding Runs Above Average) is different than FRA (Fair Run Average), which we are replacing for everyday purposes with DRA.
* [2] We use RE24/PA: the average effect of the pitcher on run expectancy per batter faced over the course of a season.
* [3] We have populated DRA back to 1953.
* [4] In fact, don’t try to suggest Pedro’s 2000 season is comparable to what anyone else has ever done at any time. It’s probably very unfair to the other player.

[Why FIP?](http://www.fangraphs.com/library/pitching/fip/)

* <https://www.fangraphs.com/library/pitching/fip/>
* Fielding Independent Pitching (FIP) measures what a player’s ERA would look like over a given period of time if the pitcher were to have experienced league average results on balls in play and league average timing. Back in the early 2000s, research by Voros McCracken revealed that the amount of balls that fall in for hits against pitchers do not correlate well across seasons. In other words, pitchers have little control over balls in play and assuming short-term fluctuations in BABIP are attributable to the pitcher is likely incorrect. McCracken outlined a better way to assess a pitcher’s talent level by looking at results a pitcher can control directly: strikeouts, walks, hit by pitches, and home runs.
* FIP is a measurement of a pitcher’s performance that strips out the role of defense, luck, and sequencing, making it a more stable indicator of how a pitcher actually performed over a given period of time than a runs allowed based statistic that would be highly dependent on the quality of defense played behind him, for example. Certain pitchers have shown an ability to consistently post lower ERAs than their FIP suggests, but overall FIP captures most pitchers’ true performance quite well. For this reason, FanGraphs’ version of Wins Above Replacement (WAR) for pitchers is based on FIP rather than on ERA and even analysts who prefer a different method of determining WAR find FIP to be extremely useful and informative.
* Calculation:
* Here is the formula for FIP:
* FIP = ((13\*HR)+(3\*(BB+HBP))-(2\*K))/IP + constant
* The constant is solely to bring FIP onto an ERA scale and is generally around 3.10. You can find historical FIP constant values here, or you can derive the constant yourself. Because FIP is designed so that league average ERA and league average FIP are the same, to find the constant for any year, all you need to do is the following:
* FIP Constant = lgERA – (((13\*lgHR)+(3\*(lgBB+lgHBP))-(2\*lgK))/lgIP)
* Knowing how to calculate the constant can be especially useful if you’re interested in doing some of your own calculations for data spanning multiple seasons. The individual weights for home runs, walks/HBP, and strikeouts are based on the relative values of those actions with respect to run prevention.
* Why FIP:
* Ultimately, we want to use statistics that allow us to isolate the performance of the player we are attempting to analyze. ERA or RA9 do a terrific job telling us how many runs were scored while the pitched was on the mound, but they do not necessarily tell us how well the pitcher performed because the number of runs a pitcher allowed is also dependent on their defense, luck, and the order in which events happened (often called sequencing).
* FIP is an attempt to isolate the performance of the pitcher by using only those outcomes we know do not involve luck on balls in play or defense; strikeouts, walks, hit batters, and home runs allowed. Research has shown that pitchers have very little control on the outcome of balls in play, so while we care about how often a pitcher allows a ball to be put into play, whether a ground ball goes for a hit or is turned into an out is almost entirely out of their control.
* As a result, a statistic that estimates their ERA based on their strikeouts, walks, hit batters, and home runs while assuming average luck on balls in play, defense, and sequencing is a better reflection of that pitcher’s performance over a given period of time. This is highly related to the reasons why we care so much about Batting Average on Balls in Play (BABIP), specifically the fact that pitchers have very little control over their BABIP allowed.
* Imagine two pitchers who always throw the same quality pitches to identical hitters, but one pitcher throws in front of a vastly superior defense. The pitcher with the better defense will allow fewer hits, and therefore fewer runs, but the two pitchers performed identically.
* Additionally, the order of events (sequencing) can have a big impact on runs allowed, even though there is no evidence that pitchers are capable of influencing their sequencing. If you have two outs and allow a single, single, home run, and out, you have just allowed three runs. If you have two outs and allow a home run, single, single, and then an out, you have just allowed one run, even though the four events were identical.
* Essentially, FIP is an attempt to measure how well a pitcher actually performed independent of factors outside of his control that contribute to runs allowed based statistics. FIP is not perfect and there are certain pitchers who have the skills to allow fewer runs than their FIP suggests, but they are reasonably rare and FIP remains highly accurate and extremely simple at the same time.
* How To Use FIP:
* In one sense, using FIP is extremely easy because it’s designed to look exactly like ERA. This means that you can read and use FIP exactly like you would typically use ERA. If a pitcher has a 3.15 FIP, that’s just like saying they have a 3.15 ERA as far as making comparisons among players is concerned. You don’t have to learn a new scale to interpret a player’s FIP.
* On the other hand, using FIP requires a bit of caution and it is best to think of it as a starting place for the analysis of pitcher performance, especially if you are interesting in determining how a pitcher is likely to perform in the future. In the long run, the majority of pitchers will have ERAs and FIPs that are very close together, but over the course of a season they could vary a great deal. Typically, people attribute the difference between the two to luck on balls in play, but there are other factors that can lead to a difference.
* For example, pitchers with the ability to limit the running game or generate fly balls at the expense of line drives or ground balls are more likely to beat their FIP than the average player. This doesn’t mean that every lefty fly ball pitcher will do so, but simply that holding runners and generating a type of batted ball that falls for hits less frequently are legitimate skills that might allow you to limit your runs allowed.
* If you have to bet on a pitcher’s ERA or their FIP, FIP is the better bet, but FIP tells you about a subset of a pitcher’s results which means that it is possible that it is missing something important about that pitcher’s profile that allows them to run consistently high or low BABIPs.
* FIP is a terrific statistic, but as always, looking into the components of the statistic and other measures of pitcher performance will help you understand how a pitcher is truly performing.
* Context:
* Please note that the following chart is meant as an estimate, and that league-average FIP varies on a year-by-year basis so that it is always the same as league-average ERA. To see the league-average FIP for every year from 1901 to the present, check the FanGraphs leaderboards.
* Fielding Independent Pitching (FIP)
* Rating FIP
* Excellent 3.20
* Great 3.50
* Above Average 3.80
* Average 4.20
* Below Average 4.40
* Poor 4.70
* Awful 5.00
* Based on 2016 Run Environment
* ● Voros McCracken’s research was called Defense Independent Pitching Theory (DIPS Theory). It’s the building block of much of today’s pitching analysis. It can be a tricky concept to understand and counter-intuitive for most baseball fans. Refer to our sections on DIPs, BABIP, and Luck for more information.
* ● FIP does a better job of predicting the future than measuring the present, as there can be a lot of fluctuation in small samples. It is less effective in describing a pitcher’s single game performance and is more appropriate in a season’s worth of innings. That doesn’t mean it isn’t a retrospective statistic, simply that it requires more than a handful of innings to be a reliable indicator of performance, just like any statistic.
* ● FIP is not league or park adjusted meaning that pitchers in good pitcher’s parks will have consistently lower FIPs and pitchers who pitch during eras of lower run scoring will have consistently lower FIPs. To control for both of those factors, FanGraphs offers FIP-, which is a park and league adjusted version of the statistic.
* ● FanGraphs’ WAR for pitchers is based on FIP, but we also offer a statistic called RA9-WAR if you are interested in using a different input.

[Everything FIP](http://sports.yahoo.com/mlb/blog/big_league_stew/post/Everything-you-always-wanted-to-know-about-FIP?urn=mlb,206286)

* **What it stands for**: **F**ielding **I**ndependent **P**itching, devised by Tom Tango, based on research by Voros McCracken.
* **How they calculate FIP**: Tango's simplest version is expressed by the following equation.
* https://web.archive.org/web/20100106005516im_/http:/upload.wikimedia.org/math/6/4/c/64cce9b9812b7fbc1dabe63152818531.png
* Usually a constant is added to scale the FIP to the league average, so the formula winds up looking like:
* (13 \* HR + 3 \* BB - 2 \* K)/IP + 3.20
* (Not to confuse everyone right off the bat, but BB can also be represented as BB + HBP - IBB, which forgives a pitcher for walks his manager ordered but penalizing him for beanball.)
* There is also a stat called xFIP (aka expected FIP) that utilizes many more pitching components in an attempt to normalize for the luck in home runs allowed, but it's outside the scope of this article.
* **What FIP is good for**: FIP is a stat that's supposed to look like ERA, but it sure isn't ERA. The formula is shocking in its simplicity, a result of McCracken's revolutionary finding that pitchers only have control over home runs, walks, and strikeouts and that nearly every other outcome — from hits to errors to runs that score on anything other than a homer — is essentially a product of random chance.
* [Here's the sentence that shook sabermetrics](https://web.archive.org/web/20100106005516/http:/www.baseballprospectus.com/article.php?articleid=878), all the way back in 2001:
* There is little if any difference among major-league pitchers in their ability to prevent hits on balls hit in the field of play.
* (There are other things that are a pitcher's fault — hit batters, balks, wild pitches, etc. — and it's possible to create more nuanced formulas that account for exactly how much of each possible outcome is the pitcher's fault; this is the simplest form.) As the Hardball Times says:
* "FIP helps you understand how well a pitcher pitched, regardless of how well his fielders fielded."
* Put another way, it tells you how well he should have done, not how well he actually did.
* **How FIP works**: You're probably wondering where the coefficients came from — 13 \* HR, 3 \* BB, and 2 \* K. The simple answer is that [Tom Tango](https://web.archive.org/web/20100106005516/http:/www.tangotiger.net/bsrexpl.html) created a matrix with run values for each play outcome. The coefficients attempt to adjust for how much each home run and walk contribute to the other team's runs scored and how much each strikeout contributes to preventing the other team's runs scored.
* Basically, the fundamental difference between old-school stats and new-school stats is that old-school stats measure what happened at the surface level — batting average, earned run average, wins. New-school stats try to measure each player's contribution to those surface stats, while filtering out the contributions of their teammates and the random fluctuations of chance. Because of this, newer stats tend to do a much better job of predicting the future. FIP is a much better predictor of future performance than ERA, wOBA is a much better predictor of future performance than batting average, and so forth.
* However, like the Oracle at Delphi, they do a much better job of predicting the future than measuring the present. On a year-to-year basis, due to luck or defense, a player may vastly overperform or underperform his FIP.
* [Javier Vazquez](https://web.archive.org/web/20100106005516/http:/sports.yahoo.com/mlb/players/5947/)[(notes)](https://web.archive.org/web/20100106005516/http:/sports.yahoo.com/mlb/players/5947/news) is particularly notorious for this. He's had an ERA higher than his FIP for four of the last five years and by a fairly substantial margin. From 2004 to 2008 he had a 4.50 ERA, but only a 4.08 FIP. In 2009, he had the best season of his career and had his first top-four Cy Young finish, surprising many people who just looked at his ERA and won-loss record. But from his FIP, it was clear that he was a better pitcher than many people thought.
* **When FIP doesn't work**: The one thing FIP can't tell you is how many runs the other team scored. So it is a nice tool for player evaluation, but the one thing it doesn't measure is the one thing the team cares about most — whether runs were actually prevented. Fairly or unfairly, Javier Vazquez disappointed a lot of fans in New York and the Bronx. A lot of runs scored on his watch, whether they were his own fault or the fault of something like [Derek Jeter's](https://web.archive.org/web/20100106005516/http:/sports.yahoo.com/mlb/players/5406/)[(notes)](https://web.archive.org/web/20100106005516/http:/sports.yahoo.com/mlb/players/5406/news) poor range at shortstop.
* **Why we care about FIP**: It's one of [Cy Young winner Zack Greinke](https://web.archive.org/web/20100106005516/http:/mlb.mlb.com/news/article.jsp?ymd=20091118&content_id=7682348&vkey=news_mlb&fext=.jsp&c_id=mlb)'s favorite stats! Look, man cannot live by FIP alone. Ultimately, baseball is about runs, runs scored and runs prevented. FIP tells you something about how much a pitcher contributed to his team's cause, but you obviously still need to look at your favorite old reliables: ERA, Runs Scored/Runs Allowed, and so on.
* Still, FIP is a great tool to have. It's a version of ERA that tells you how well a pitcher actually pitched, rather than how well his team did. There are always more complicated measures of a pitcher's exact contributions on the mound, but certainly none as simple.

[Everything WAR](http://sports.yahoo.com/mlb/blog/big_league_stew/post/Everything-you-always-wanted-to-know-about-WAR?urn=mlb,211211)

The world of advanced baseball statistics can be an intimidating place for those of us who slept our way through advanced algebra or haven't been a follower of the Bill James revolution from the beginning.

Still, that doesn't mean that we should feel left out when it comes to another way of understanding and appreciating the game we all love. With that in mind, BLS stat doctor ***Alex Remington*** will explore a new advanced statistic each week during the offseason, providing a simple primer for the uninitiated.

**Today's statistic:** WAR

**Good god, what does it stand for?**: Wins Above Replacement. [Say it again!](https://web.archive.org/web/20100106115433/http:/www.youtube.com/watch?v=01-2pNCZiNk) Simply put, Wins Above Replacement means: how many wins did that player contribute to his team's win total above and beyond what they would have gotten from a "replacement value" player, someone they could have picked up off the scrap heap for next to nothing?

**How they calculate WAR**: WAR is probably the most popular total value stat out there today, a single number that attempts to quantify a player's worth by looking at his offense, defense (or pitching), defensive position, and the context of the year and league. (Bill James' Win Shares is  another total value stat, but it has been eclipsed in popularity by WAR.)

It's also one of the hardest to calculate, so hang with me here.

As Jabberwocky explains [here](https://web.archive.org/web/20100106115433/http:/www.beyondtheboxscore.com/2009/6/12/906943/war-lords-of-the-diamond-position) on Beyond the Boxscore, WAR basically aggregates a lot of methods and insights from other advanced stats, like [wOBA](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/blog/big_league_stew/post/Everything-you-always-wanted-to-know-about-wOBA?urn=mlb,208135), [FIP](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/blog/big_league_stew/post/Everything-you-always-wanted-to-know-about-FIP?urn=mlb,206286) and UZR, all of which basically express a player's contributions to the team in terms of runs added or runs prevented. These run measures are then adjusted for that player's value "Above Replacement." A replacement player is defined as someone who is below average and should be easily obtainable, the sort of fringy cup-of-coffee guy you can find in AAA, on the waiver wire, or acquire for a PTBNL, a warm body who hurts the team the more he plays. With WAR anyone on the 25-man roster should have an implicit value just in keeping those fringy AAAA guys down on the farm and off the team.

After calculating the run measures, a positional adjustment is added to account for the relative importance of different positions and a league adjustment to account for the relative strength of the American League — which has a higher run environment thanks to the DH, higher salaries thanks to the presence of the Red Sox and Yankees and currently a greater concentration of talent — compared to the National League.

For a position player, you then add their wRAA, UZR, positional adjustment, and replacement adjustment to get their Runs Above Replacement. Then you scale all those contributions to be expressed in terms of total team wins. The usual scale is that 10 runs is equal to one team win. So WAR is equal to RAR divided by 10.

Believe it or not, it's much more complicated for a pitcher. In order to derive an equivalent of RAR from FIP, FIP must be rescaled to Runs Allowed rather than ERA (because unearned runs count for the purposes of RAR), then adjusted to the pitcher's innings pitched, and then the result must be adjusted for replacement level, for the run environment that the pitcher is creating, and for park effect. Dave Cameron wrote a seven-part series explaining the methodology and showed the formula in action [here](https://web.archive.org/web/20100106115433/http:/www.fangraphs.com/blogs/index.php/pitcher-win-values-explained-part-seven), but even he doesn't fully explain all the constants in his final formula. There's another way to derive it from Pitcher Winning Percentage, which is a variant on Bill James's Pythagorean Win Expectation formula. Devil\_Fingers [explains it here](https://web.archive.org/web/20100106115433/http:/www.drivelinemechanics.com/2009/3/10/787510/how-good-is-oakland-s-2009).

**What WAR is good for**: Absolutely nothing! Oh, excuse me. I've just been informed by the Committee on Obvious Humor not to make any more references to that song.

WAR is basically one attempt at a Grand Unified Stat, a single number that usefully expresses a player's worth and can be compared to all other players. Dave Cameron [just wrote a fascinating piece](https://web.archive.org/web/20100106115433/http:/insider.espn.go.com/mlb/insider/news/story?id=4742365) for ESPN which used WAR analysis to argue that [Chase Utley](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7072/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7072/news) is actually a more valuable player than [Albert Pujols](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/6619/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/6619/news), because of the comparative value of their defensive positions and team contracts. Because WAR takes all of this into account, it's a great starting point for argument.

Also, as the Utley article illustrates, WAR is also an interesting way to look at player salaries. A full discussion on properly valuing player salaries is outside the scope of this article (and Colin Wyers did a great job [here](https://web.archive.org/web/20100106115433/http:/www.hardballtimes.com/main/article/the-root-part-1/), [here](https://web.archive.org/web/20100106115433/http:/www.hardballtimes.com/main/article/the-root-part-2/) and [here](https://web.archive.org/web/20100106115433/http:/www.hardballtimes.com/main/article/the-root-part-3/)) but the basic assumption is this: there is a [linear relationship](https://web.archive.org/web/20100106115433/http:/baseballanalysts.com/archives/2009/11/war_salary_and.php) between WAR and salary, and by multiplying a player's WAR by the average value of a win that year, you can come up with an approximate figure of how much money his performance that year was truly worth.

**How WAR works**: Conceptually, it's simple: WAR is a sum of the win value of a player's offense, defense, pitching, adjusted for that player's defensive position, playing time (thus keeping the replacement level players off the field) and year, park, and league context. The heavy lifting occurs in the individual calculations of the values and constants — as usual, I really just recommend you just use FanGraphs.

The fuzziest of all of these is the concept of the "replacement player." Tom Tango [defines it](https://web.archive.org/web/20100106115433/http:/www.insidethebook.com/ee/index.php/site/article/how_to_calculate_war/) as "the talent level for which you would pay the minimum salary on the open market, or for which you can obtain at minimal cost in a trade." On the other hand, we've all seen our teams struggle with players below replacement level, like [Emilio Bonifacio](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/8112/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/8112/news) and [Yuniesky Betancourt](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7511/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7511/news) (who had WAR of -0.4 and -2.1 last year, literally below replacement level). In the majors last year, there were eight players with a negative WAR. So "replacement level" is more of a theoretical conception rather than a concrete reality.

The positional adjustments are fixed: catchers, center fielders and shortstops are significantly more valuable than designated hitters and first basemen, so if you look at two players with similar offensive and defensive production at their positions, a catcher with those skills is worth about two more wins than a first baseman.

Generally speaking, the very biggest stars in the game, both pitchers and hitters, post a WAR of 7 or better. [Ben Zobrist](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7829/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7829/news) led all of baseball this year with a WAR of 8.6, which goes to show just how valuable his incredible versatility in the field was. Anything above 9 is extremely rare. (Since 2001, the first year that we have all of the data available for WAR calculations, only two players have ever posted a WAR above 10: [Adrian Beltre](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/6039/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/6039/news), the year he exploded to hit .334 and 48 homers with world-class defense at third, and [Barry Bonds](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/3918/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/3918/news), who did it every year from 2002-2004. Very good players are between 4 and 6 WAR; players between 1 and 3 are useful, and those with a WAR less than 1 are, by definition, easily replaceable.

**When WAR doesn't work**: An awful lot of fixed constants are used in the calculation of this stat, both in the individual run wOBA and FIP calculations and then in the positional, replacement, and league adjustments. Every model relies on assumptions, and WAR is no exception. But that simplicity comes at the price of the ability for the model to predict all the variance we see. Fortunately, a number of sabermetricians on the web have continued to go back and rejigger the constants every year, as more games are played and more data is added, so that the assumptions don't become outdated. Still, the caveat remains the same: there are a lot of different ways of calculating a player's worth, and a lot of different ways of choosing which on-field events to include in the model (intentional walks? double plays?) and which to ignore (temperature? wind?).

In addition, UZR is a useful defensive stat, but it's far from perfect; it frequently contradicts the findings of Plus/Minus, from John Dewan's Fielding Bible, which results from a video analysis of every play made by every defensive player, and there's no easy way to reconcile the contradiction. UZR can show some major fluctuations year to year — [Nate McLouth](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7513/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7513/news) had a UZR of -13.8 in 2008, then +3.6 in 2009, for example — which resulted in a nearly two-win swing in his defensive WAR. Ultimately, WAR is a terrific shorthand for a player's worth, but it's by no means the final word on a player.

**Why we care about WAR**: Two reasons. First, whatever its flaws, it's a stat that lets you express just about everything you can say about a player. (Hey, Yankee fans: [Derek Jeter](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/5406/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/5406/news) had a higher WAR than [Hanley Ramirez](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7488/)[(notes)](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/7488/news)last year, 7.6 to 7.4. It's true!) Second, because it lends itself to contract discussions, it's a great way for fans to visualize roster construction in terms of what each player contributes to the team. It's not the perfect stat — no single measure ever will be — but it's impressive in its scope, and awfully fun to use. Good God!