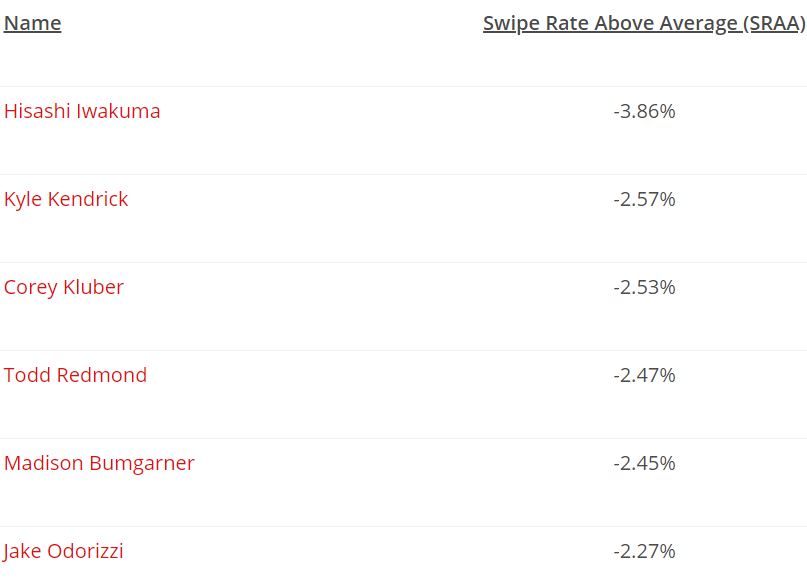
Deserved Run Average

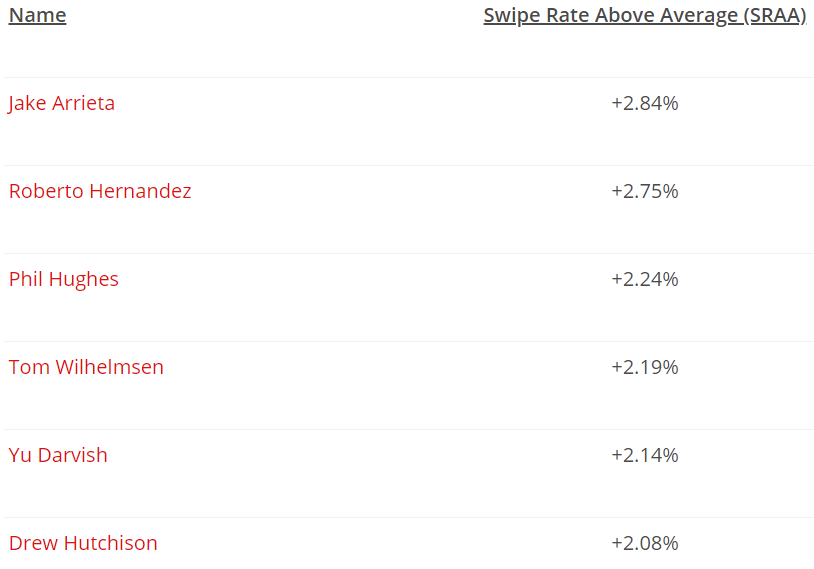
* *Prospectus Feature: Introducing Deserved Run Average (DRA) And All Its Friends*

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

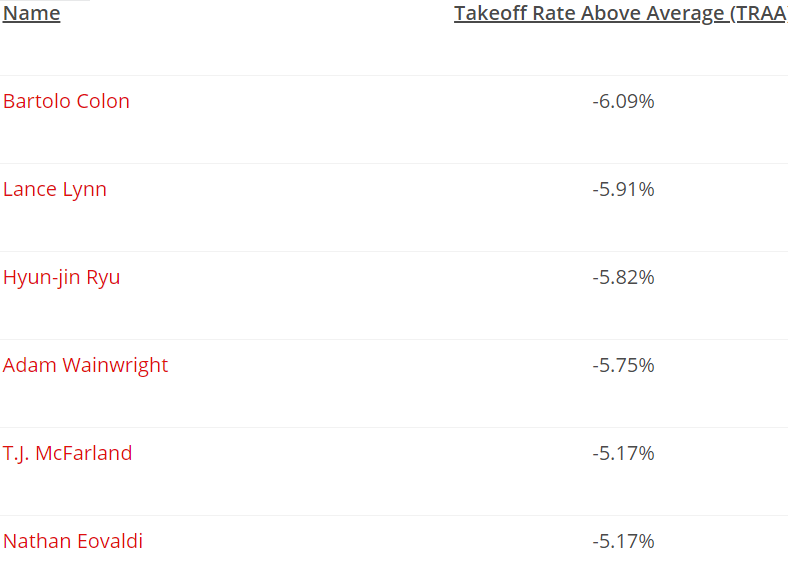
* **ERA** = benchmark by which pitchers have been judged for a century == runs pitcher gives up, on average, every 9 innings pitched
* If gave up a bunch of runs = probably terrible; if gave up very few = assume he’s pretty good.
* ERA problem = blames/credits pitcher for everything, simply b/c he threw 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:



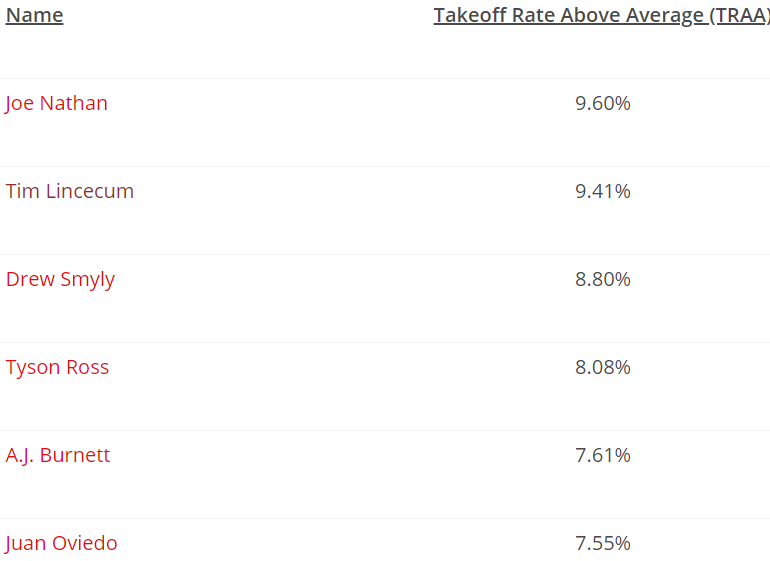
* 2014 pitchers that baserunners exploited the most:



* **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:



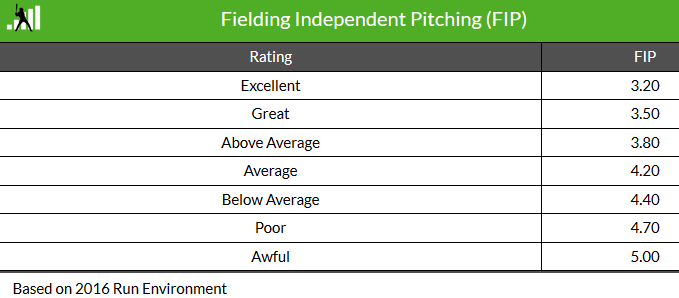
* Step 5: Calculate DRA (Deserved Run Average).
* Now got components, so it’s time to calculate each pitcher’s DRA.
* Steps:
* 1) Put all identified components (value/PA, SRAA, TRAA, EPAA) together into a new **regression**, looking for their **combined effect** on **run expectancy**. [2]
* Added 2 more 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
* In-season DRA = calculated each night after previous day’s games have concluded.
* Will be able to use DRA not only to put past pitching performances in context [3] but also to monitor value of pitchers through season + beyond.
* DRA does very good job of measuring a pitcher’s actual responsibility for runs scored while on the mound, certainly better than any metric in the public domain + only DRA gives assurance a pitcher’s performance is actually being considered in context of batter, catcher, runners on base, as well as stadium + stadium environment for the game
* Detailed explanation of DRA’s effectiveness = In Depth article
* *Reader’s Digest version* 🡺 **2 measures of accuracy to pay particular attention to in evaluating accuracy of a new metric.**
* 1) How close, mathematically, metric’s prediction is to actual # of RA w/ pitcher on the mound.
* If pitcher actually allowed 4 R per 9 innings, test our alternative metric by how close it comes to that same #
* Most commonly used calculation that does = **Root Mean Square Error (RMSE)**
* 2) Test to look @ how accurately metric ranks various pitchers relative to each other (**the Spearman Correlation)**
* Why do we care about rank? 🡺 b/c all pitcher run estimates are a bit off from actual RA, + more so early in season, so as a check, test whether it is at least ranking pitchers correctly relative to each other.
* In other words, if metric can’t estimate RA down to the exact decimal, least it can do is tell the difference between 2 pitchers=
* To judge DRA’s accuracy, can compare it to leading brand: **FIP** = does reasonable job of predicting a pitcher’s actual RA in a season.
* *Does DRA do a better job than FIP? 🡺* ***It does***
* Compared how well FIP + DRA predicted each pitcher’s RA/9 in each of the past 4 MLB seasons + looked @ performance w/ all pitchers + then 2 subsets: pitchers who faced >= 170 batters (about workload of an established MLB reliever, or ~40 IP) + pitchers who faced >= 660 batters (~162 IP, which is a qualified MLB starter).
* Then averaged results over 4 seasons (2011­–2014) to get a consistent (+ recent) pic of each metric’s performance.:

` 

* **DRA = consistently superior to FIP @ all sample sizes**.
* By accounting for *context* in which pitcher is throwing, **DRA allows us to determine which runs are most fairly blamed on the pitcher** + is **particularly effective w/ smaller samples**.
* Even for pitchers w/ only a few batters faced, DRA = already separating good from bad w/ superior accuracy
* In addition to being useful in + of itself, DRA = become new foundation of **Pitcher Wins Above Replacement Player (PWARP)** @ Baseball Prospectus.
* By integrating **DRA** into **WARP =** ***DRA\_PWARP*** =can do a better job than ever of evaluating how much value individual pitchers delivered to teams, both during current season + as compared to past pitchers in other seasons + eras.
* 5 best qualified starters by DRA over past 25 years (in some cases DRA basically matches RA/9; in others, doesn’t, so when DRA + RA/9 disagree, go w/ DRA = tells you how well pitcher really pitched)
* 
* 1 caution: DRA = *NOT* (presently) adjusted for run-scoring *across different eras*.
* Rather, it is **adjusted** to the average RA by the league for that season 🡺 don’t directly compare Pedro’s DRA of 1.03 in 2000 to somebody else’s DRA in 1985 or some other season
* DRA metric that compares players across eras will be coming soon.
* 2nd caution: DRA corrects for what is known as **survival bias** = tendency of better pitchers to pitch more innings in a season.
* Applying full DRA model early on can result in some extreme values + to avoid that, model was kept simple at first during the season, + model only **value/pa to RE24.**
* Further along, allow full model to operate + achieve best explanation of each pitcher’s performance.
* Conclusion
* **DRA** + the other statistics introduced: **Swipe Rate Above Average** (to measure base-stealing success), **Takeoff Rate Above Average** (to measure base-stealing attempts), + **Errant Pitches Above Average** (to measure passed balls and wild pitches).
* 3 final things to remember.
* 1) While DRA accounts for a great many things, DRA DOESN’T need to be complicated for fans.
* DRA = best estimate of how good pitcher has been in a particular season + you can leave details to Baseball Prospectus
* 2) DRA was created to evaluate PAST
* To project *future* performance of a pitcher, use **PECOTA** + to evaluate how talented pitcher is regardless of performance to date, use **cFIP**
* cFIP can be used w/ DRA to compare recent results w/ likelihood of future improvement/decline
* 3) DRA is now the foundation for Pitcher Wins Above Replacement (PWARP) @ Baseball Prospectus.
* For the time being, to see how many W a pitcher has been worth in a particular season, look @ **DRA\_PWARP**.
* [1] Please note **FRAA (Fielding Runs Above Average)** = different than **FRA (Fair Run Average),** which we are replacing for everyday purposes with **DRA**.
* [2] We use **RE24/PA: average effect of pitcher on run expectancy per batter faced over course of season**
* [3] We have populated DRA back to 1953.

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

* <https://www.fangraphs.com/library/pitching/fip/>
* **Fielding Independent Pitching (FIP) = what a player’s ERA would look like over given period of time if pitcher were to have experienced league avg. results on balls in play + league avg. timing**
* Back in early 2000s, research by Voros McCracken revealed **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** + assuming short-term fluctuations in **BABIP** are attributable to pitcher is likely INCORRECT.
* McCracken outlined a better way to assess a pitcher’s talent level = looking @ results a pitcher CAN control directly: SO, walks, HBP, + HR.
* **FIP** = measurement of pitcher’s performance that *strips out role of defense, luck, + sequencing*, making it a *more stable indicator of how a pitcher actually performed over a given period of time than a RA-based statistic that would be highly dependent on other factors*
* Certain pitchers have shown an ability to consistently post lower ERAs than FIP suggests, but overall, FIP captures most pitchers’ true performance quite well.
* FanGraphs’ version of **Wins Above Replacement (WAR)** for pitchers is based on **FIP** rather than on ERA + even analysts who prefer a different method of determining WAR find FIP to be extremely useful + informative.
* **FIP = ((13\*HR)+(3\*(BB+HBP))-(2\*K))/IP + constant**
* constant = solely for bringing FIP onto an ERA-scale + is generally ~3.10
* B/c FIP = designed so that league avg. ERA + league avg. FIP = 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 interested in doing your own calculations for data spanning multiple seasons.
* Individual weights for HRs, walks/HBP, + SOs are based on the RELATIVE values of THOSE actions w/ respect to run prevention.
* Why FIP:
* Ultimately, **want statistics that isolates performance of a player we’re attempting to analyze**
* **ERA** or **RA9** = terrific job telling us how many runs were scored while pitcher was on the mound, but do not necessarily tell us *how well* pitcher performed, RS = also dependent on defense, luck, + the order in which events happened (**sequencing**).
* **FIP** = an attempt to isolate performance of pitcher by using only outcomes known to not involve luck on BIP or defense 🡺 SOs, walks, HBP, + HRs.
* Research has shown **pitchers have very little control on the outcome of BIP**, 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 stat that **estimates ERA based on SOs, walks, HBPs, + HRs while *assuming average luck on BIP, defense, + sequencing*** = better reflection of performance over a given period of time.
* This is highly related to reasons why we care so much about BA on Balls in Play (**BABIP**), specifically b/c **pitchers have very little control over their BABIP allowed.**
* Imagine 2 pitchers who always throw same quality pitches to identical hitters, but one throws in front of a vastly superior defense 🡺 pitcher w/ better defense = allows fewer hits, + therefore fewer runs, but 2 pitchers performed identically.
* Additionally, order of events/**sequencing** can have a big impact on RA, *but there’s no evidence pitchers = capable of influencing their sequencing.*
* W/ 2 outs 🡪 allow a single, single, HR, then out = just allowed 3 runs.
* W/ 2 outs 🡪 allow a HR, single, single, then out = just allowed 1 run
* *even though the 4 events were identical.*
* Essentially, FIP = an attempt to measure how well a pitcher ACTUALLY performed, *independent of factors outside his control that contribute to RA-based statistics.*
* FIP = not perfect + certain pitchers have skills to allow fewer RA than their FIP suggests, but they are reasonably rare + FIP remains highly accurate + extremely simple at the same time.
* How To Use FIP:
* Using FIP = extremely easy b/c it’s designed to look exactly like ERA 🡺 can read + use FIP exactly like ERA
* 3.15 FIP = like saying 3.15 ERA, as far as making comparisons among players is concerned 🡪 don’t have to learn a new scale to interpret a player’s FIP.
* *OTOH*, FIP requires a bit of caution 🡪 is best to think of it as a *starting place* for analysis of pitcher performance, especially if interested in determining how pitcher is likely to perform in future
* In long run, majority of pitchers will have ERAs + FIPs = very close together, but over course of a season could vary greatly
* Typically, people attribute difference between the 2 to luck on BIP, but there are other factors that can lead to a difference.
* Ex: Pitchers w/ ability to limit running game or generate fly balls @ expense of line drives or ground balls = more likely to beat their FIP than the average player.
* Doesn’t mean every lefty fly ball pitcher will do so, but simply that holding runners + generating a type of batted ball that falls for hits less frequently = legitimate skills that might allow limiting of RA
* FIP = better bet vs. ERA, but also tells you about a subset of a pitcher’s results = means it is possible it’s missing something important about that pitcher’s profile that allows them to run consistently high/low BABIPs.
* **\*\*\*FIP = terrific stat, but as always, looking into components of the stat + other measures of pitcher performance will help understand how a pitcher is truly performing\*\*\***
* Context:
* Estimate Chart (league-average FIP varies on a year-by-year basis so that it’s always the same as league-average ERA)



* [1] McCracken’s research was called **Defense Independent Pitching Theory (DIPS Theory) =** building block of much of today’s pitching analysis.
* can be tricky to understand + counter-intuitive for most baseball fans.
* [2] FIP = better job of predicting future than measuring present, as there can be a lot of fluctuation in small samples.
* less effective in describing single game performance + is more appropriate in a season’s worth of innings.
* **Doesn’t mean it isn’t a retrospective statistic,** simply it requires more than a handful of innings to be a reliable indicator of performance, just like any statistic.
* [3] FIP is NOT league or park adjusted 🡺 pitchers in good pitcher’s parks = consistently lower FIPs + pitchers who pitch during eras of lower run scoring = consistently lower FIPs.
* To control for both factors, use offers **FIP- =** park + league adjusted version of the statistic.
* [4] FanGraphs’ **WAR** for pitchers = based on FIP, but there’s also **RA9-WAR** stat if 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)

* **Fielding Independent Pitching (FIP**) **=** devised by Tom Tango based on research by Voros McCracken
* Simplest version 🡺 https://web.archive.org/web/20100106005516im_/http:/upload.wikimedia.org/math/6/4/c/64cce9b9812b7fbc1dabe63152818531.png
* Usually a **constant** is added to scale FIP to league average, so the formula winds up looking like:

https://web.archive.org/web/20100106005516im_/http:/upload.wikimedia.org/math/6/4/c/64cce9b9812b7fbc1dabe63152818531.png**+ 3.20**

* BB can also be represented as **BB + HBP – IBB** = forgives pitcher for walks a manager ordered but penalized him)
* There is also a stat = **xFIP** (**expected FIP**) = utilizes many *more* pitching components in an attempt to normalize for luck in HR
* FIP = a stat that's supposed to look like ERA, but isn't 🡪 shocking in simplicity, a result of McCracken's revolutionary finding that **pitchers only have control over HRs, walks, SO + nearly every other outcome (from H to ERR to runs from on anything other than HR) = essentially a product of random chance.**
* “There is little if any difference among ML pitchers in ability to prevent hits on balls hit into field of play”
* Other things ARE pitcher's fault = hit batters, balks, wild pitches, etc., + it's possible to create more nuanced formulas that account for exactly how much of each possible outcome = pitcher's fault; this is the simplest form.)
* Hardball Times: "FIP helps understand how well a pitcher pitched, regardless of how well his fielders fielded."
* Put another way, FIP = how well he should have done, not how well he *ACTUALLY* did.
* Coefficients came from Tango’s matrix w/ run values for each play outcome + they attempt to adjust for how much each HR + walk contribute to other team's RS + how much each SO contributes to preventing other team's RS
* Fundamental difference between old-school + new-school stats = old-school stats measure what happened *at the surface level* (BA, ERA, W) vs. new-school stats = try to measure **each player's** **contribution** to surface stats, while **filtering out contributions of teammates + random fluctuations of chance**.
* B/c of this, newer stats tend to predict future better 🡪 FIP = much better predictor of future performance than ERA, wOBA = much better predictor of future performance than BA, + so forth.
* However 🡺 **much better @ predicting future than measuring present.**
* On a year-to-year basis, due to luck or defense, player may vastly over/underperform his FIP.
* 1 thing FIP can't tell you = *how many runs the other team scored.*
* FIP = nice tool for player evals, but the 1 thing it *doesn't* measure = the 1 thing teams cares about most == whether runs were actually prevented.
* Cannot live by FIP alone = ultimately, baseball is about runs, RS, + runs prevented.
* FIP tells you something about how much pitcher contributed to team's cause, but obviously still need to look at old reliable’s = ERA, RS/Runs Allowed, + so on.
* Still, FIP = great tool to have = version of ERA that tells you how well a pitcher *actually* pitched, rather than how well the *team* did.
* There are always more complicated measures of a pitcher's *exact* contributions, 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)

* **Wins Above Replacement (WAR) =** wins a player contributed to team's win total above + beyond what they would’ve gotten from a **replacement value** **player** (picked up off the scrap heap for next to nothing)
* WAR = probably most popular **total value stat** today = single # to attempt to quantify **player's worth** by looking @ his **offense, defense (or pitching), defensive position, + context of the year + league**
* Bill James' **Win Shares** = another total value stat, but has been eclipsed in popularity by WAR
* Also 1 of the hardest to calculate = basically aggregates a lot of methods + insights from other advanced stats (**[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)**, UZR**, etc.) 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**."
* **Replacement player =** someone who is *below average* + should be easily obtainable (AAA, on waiver wire, or acquire for a **PTBNL** = warm body who hurts team the more he plays)
* W/ WAR, anyone on 25-man roster should have an implicit value just by keeping those AAAA guys down on the farm + off the team.
* After calculating the run measures, a **positional adjustment** is added to account for relative importance of different positions + a **league adjustment** to account for relative strength of the AL (higher run environment thanks to DH, higher salaries thanks to the presence of Red Sox + Yankees, + currently, a greater concentration of talent) compared to the NL
* For a position player, then add their **wRAA**, **UZR**, **positional adjustment**, + **replacement adjustment** to get their **Runs Above Replacement**, then scale all those contributions to be expressed in terms of **total team wins**.
* Usual scale 🡺 10 runs = 1 team win, so **WAR = RAR/10**
* WAR = much more complicated for pitcher 🡪 to derive an equivalent of RAR from FIP, FIP must be rescaled to RA rather than ERA (b/c *unearned* runs count for purposes of RAR), then adjusted to the pitcher's IP, + *then* *THAT* result must be adjusted for replacement level, for the **run environment** the pitcher is creating, + for **park effect**.
* Dave Cameron = 7-part series explaining the methodology + showed the formula in action = doesn't fully explain ALL the constants in the final formula.
* Another way to derive it from **Pitcher Winning Percentage**, a variant on Bill James's **Pythagorean Win Expectation formula**.
* **WAR =** basically 1 attempt at a **Grand Unified Stat = a single #** that *usefully* expresses a player's worth + can be compared to all other players.
* Dave Cameron argues Utley = more valuable than Pujols b/c of the **comparative value** of their defensive positions + team contracts.
* B/c WAR takes all this into account = it's a great starting point for argument.
* WAR = also an interesting way to look @ player salaries via a basic assumption = there is a [**linear relationship**](https://web.archive.org/web/20100106115433/http:/baseballanalysts.com/archives/2009/11/war_salary_and.php) between WAR + salary, + multiplying a player's WAR by average value of a win that year can = an approximate figure of how much $ his performance that year was truly worth
* Conceptually, WAR = simple 🡺 sum of the **win value** of a player's offense, defense, pitching, adjusted for that player's defensive position, playing time (thus keeping replacement level players off the field) + year, park, + league context.
* Heavy lifting occurs in individual calculations of the values + constants
* Fuzziest of all = concept of "**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’d pay the *min.* salary on open market, or for which you can obtain *at minimal cost* in a trade."
* On the other hand, all seen teams struggle w/ players below replacement level,
* In the majors last year, there were 8 players w/ negative WAR, so "replacement level" = more theoretical conception rather than a concrete reality.
* Positional adjustments = *fixed*: catchers, center fielders, + shortstops = significantly more valuable than DHs + 1st basemen, so if you look at 2 players w/ similar offensive + defensive production @ their positions, a catcher w/ those skills = worth ~2 more wins than a 1st baseman.
* Generally speaking, very biggest stars in the game, both pitchers + hitters, post a WAR >= 7 + anything > 9 = extremely rare.
* Since 2001, 1st year we have all data available for WAR calculations, only 2 players have ever posted a WAR > 10: [Adrian Beltre](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/6039/) w/ hitting .334 w/ 48 HR w/ world-class defense @ 3rd + [Barry Bonds](https://web.archive.org/web/20100106115433/http:/sports.yahoo.com/mlb/players/3918/) from 2002-2004
* Very good players = between 4-6 WAR; players between 1-3 = useful, + WAR < 1 = by definition, easily replaceable.
* An awful lot of fixed constants are used in the calculation of WAR, both in individual run wOBA + FIP calculations + then in the positional, replacement, + league adjustments.
* **\*\*\*Every model relies on assumptions, + WAR is no exception\*\*\***.
* Simplicity comes @ price of the ability for the model to predict all the variance we see.
* Fortunately, sabermetricians have continued to go back + rejigger constants every year, as more games are played + more data is added, so assumptions don't become outdated.
* Still, caveat remains 🡺 there’re a lot of different ways of calculating player's worth, + a lot of different ways of choosing which on-field events to include in a model (intentional walks? double plays?) + which to ignore (temperature? wind?).
* In addition, UZR = useful defensive stat, but far from perfect 🡺 frequently contradicts findings of +/, from John Dewan's Fielding Bible, which results from video analysis of every play made by every defensive player, + there's no easy way to reconcile the contradiction.
* **UZR** can show some major fluctuations year to year, resulting in multiple-win swings in defensive WAR
* Ultimately, WAR = terrific **shorthand** for a player's worth, but by no means final word on a player.
* Care about WAR for 2
* 1) Whatever its flaws, it's lets you express just about everything you can say about a player.
* 2) B/c 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.
* WAR = not the perfect stat (no *single* measure ever will be) but is impressive in its scope, + awfully fun to use