

Wins Above Replacement or How I Learned to Stop Worrying and Love deGrom

The War Talk to End All War Talks

By Sean Forman and Hans Van Slooten, Sports Reference LLC

Follow along at:
gadel.me/2018-sabr-analytics

Hello, I'm sean forman. Originally I vowed to include no WAR-related puns, but as you can see I failed miserably in that task.

Sean



Hans



Hans, who is pictured there, runs baseball-reference on a day-to-day basis.

Sean Smith



Tom Tango



We also relied on some outside experts when developing our WAR framework. Sean Smith developed the original set of equations that we started with in 2009. He originally published under the name rallymonkey (which is why our WAR is sometimes called rWAR). We did a major revamping in 2012 and Tom Tango answered maybe two dozen tedious emails from me during the process.

Ways to Measure Value

- Wins and Losses
- Runs Scored and Runs Allowed
- Win Probability Added
- Component Measures, WOBA, FIP, DRS, etc.
- Launch Angle & Velocity, Catch Probability, Tunneling, Framing, Spin Rate, etc => “Statcast”
- Do you care more about:
 - Did it directly lead to winning outcome?
 - How likely are they to do this again? or
 - What is the context-neutral value of what they did?

Differing views on what matters leads to many of the arguments over WAR. To some degree, I'm not willing to argue these points. We've made our choice and implemented a system based on that. You can make your choice and implement your system based on that.

Ways to Measure Value



Wins & Losses

Skills & Statcast

Where you are on the continuum guides your implementation details.

BR's Approach:

If we were the GM of a AAAA-quality team at the start of the YYYY season, which player's performance would we most want to add to our team?

Our approach is attempts to answer the question, what was the value of what the player did to a typical team?

If we changed our implementation it would change the question we are answering.

Lots of Approaches

- Pete Palmer, Total Player Rating
- Bill James, Win Shares
- Keith Woolner, VORP
- Baseball Prospectus, WARP
- Fangraphs, fWAR
- Baseball-Reference, bWAR or rWAR
- Ben Baumer, openWAR

These are “Uber-Stats”, all-in-one

GDP vs WAR, 100 page guidelines how to construct it

Data can be hard to come by

Can be restated after the fact

Different ways of measuring can tell you different things

If you like tinkering or want a starting point for your system you should check out openWAR

Lots of Approaches

Reggie Jackson

bWAR: 30th

fWAR: 25th

win shares: 13th

WARP: 18th

Fergie Jenkins

bWAR: 11th

fWAR: 13th

win shares: 11th

WARP: 12th

1954-present

The approaches are pretty different in some cases, but they are in the same general ballpark when evaluating players.

How we make WAR

- Compare the player to average across various roles in the game.
- Adjust based on various contexts.
- How much would they help an average team? WAA.
- Then compare an average player with that playing time to a replacement player. $WAR = WAA + \text{Replacement Level}$.
- Details and Resources here: gadel.me/war-explained.

One of the things that hope you'll see as we go through this is that we really try very hard to account for lots and lots of factors. Many times people on twitter or in our email will say, you really should account for x, y and z. And typically in the 10 years we've been doing this, we've considered and studied all of those things at some point.

I realize this may sound like I'm stifling debate and making an appeal to authority, but I just want to assure you there is a lot of theory and data underlying our assumptions.

Details at bbref.com/about/

Wins Above Replacement

[WAR-Wins Above Replacement Explanation and Links to Further Resources](#)

[WAR-wRAA explanation](#)

[WAR-Position Players](#)

[WAR-Pitchers](#)

[WAR-Runs to Wins](#)

[Competing WARs Comparison Chart](#)

WAR daily data, full history: [Batters](#), [P](#)

Backup of version 2.1: [Batters](#), [Pitcher](#)

Backup of pre-2012 version: [Batters](#), [P](#)

[Directory listing with daily archive](#)

WAR Further Discussion

- How Baseball-Reference.com calculates WAR and how others do:

[WAR-Wins Above Replacement Explanation](#)

[WAR-wRAA explanation](#)

[WAR-Position Players](#)

[WAR-Pitchers](#)

[WAR-Runs to Wins](#) (this is especially relevant)

[Competing WARs Comparison Chart](#)

- My Answer to "I Don't Like How Complicated WAR Is and How it is Constantly Changing." "WAR is Like GDP for Baseball"
- Response to Critiques by Bill James and Joe Posnanski
- Bill James: [Judge and Altuve](#), and then
Bill James: [MVP Followup](#)
Joe Posnanski: [More on WAR](#), following up on Judge and Altuve.
- Dave Cameron's reply to James: [Putting WAR in Context: A Response to Bill James](#)
- Jonathan Judge at Baseball Prospectus wrote an excellent piece on variability in WAR: [Prospectus Feature: Bill James vs. The Noise](#)
- [Bill James and I discuss on twitter what I think is the crux of the matter](#)
- [Esay Jazayeri on how WAR should handle luck](#)

First Compare the Players to Average

- Summing all of these components across a league will give you zero for each value.
- Later, we'll discuss replacement level.

Position Players – Batting

- We use a park-adjusted WOBA based system.
- Linear Weights based on the run scoring environment.
- Filter out pitcher hitting for the league baselines.
- Include some play-by-play based elements, infield hits, reached on errors.
- Different factors for strikeouts and non-SO outs, to account for changes in average GIDP rates.

Don't publish WOBA because our WOBA would differ from FG's WOBA since we do some different things.

Deleted for brevity: We assume that opposing managers are rational with SH and IBB's. SH & IBB don't change batters rate-value, but we treat them as an average value result for the batter.

Position Players – Batting

Rk	Player	Rbat	Ft
1	Babe Ruth	1335.4	1
2	Barry Bonds	1128.5	1
3	Ted Williams	1069.3	1
4	Ty Cobb	995.8	1
5	Lou Gehrig	971.4	1
6	Casey Stengel	895.4	1

Rk	Player	Rbat	Ft
1	Joey Votto	59.2	2
2	Aaron Judge	57.6	2
3	Giancarlo Stanton	54.0	2
4	Mike Trout	53.0	2
5	Jose Altuve	51.9	2

Rk	Player	Rbat	Ft
1	Tommy Corcoran	-315.2	1
2	Larry Bowa	-304.6	1
3	Ozzie Guillen	-296.2	1
4	Tim Lincecum	-291.6	1
5	Alfredo Griffin	-290.2	1

Rk	Player	Rbat	Ft
1	Rougned Odor	-32.9	20
2	Billy Hamilton	-28.3	20
3	Alcides Escobar	-27.9	20
4	Alex Gordon	-25.8	20
5	Adam Engel	-24.2	20

Larry Bowa .260/.300/.320 line.

Would need to hit 186 consecutive home runs to reach a .400 slugging percentage

Batters— Baserunning

- Stolen bases and caught stealing. Make adjustments for eras where CS is not known.
- Track base advancement. First to third, second to home, first to home, outs on bases.
- Other advances, wild pitches, passed balls. Scoring on groundouts.
- Double-play avoidance. Counts times they hit GB's in DP situations vs # of DP's grounded into. Runs_dp
- All tracked via play-by-play.

For double plays, we don't penalize players for grounding into DP's at a league average rate given opportunities. Your teammates put you into that situation

Position Players – Baserunning

Rk	Player	
1	Rickey Henderson	147.9
2	Willie Wilson	145.5
3	Johnny Damon	126.0
4	Tim Lincecum	122.7
5	Ichiro Suzuki	117.5

Rk ▲	Player	
1	Paul Konerko	-76.2
2	Victor Martinez	-57.3
3	Joe Torre	-56.8
4	Miguel Cabrera	-56.7
5	Mike Piazza	-56.1

Rk	Player		Fr
1	Byron Buxton	11.4	20
2	Mookie Betts	10.1	20
3	Dce Gordon	10.0	20
4	Billy Hamilton	8.9	20
5	Christian Yelich	6.6	20

Rk	Player		I
1	Josh Bell	-7.8	
2	Matt Kemp	-7.1	
3	Kendrys Morales	-7.0	
4	Joey Votto	-6.8	
5	Eugenio Suarez	-6.1	

Paul Konerko went first to third on singles 15% of the time. Tim Raines 36% of the time. 500+ opportunities. More than 100 bases Raines added that Konerko did not.

Position Players – Fielding

- 2003-present use Sports Info Solutions (née Baseball Info Solutions) Defensive Runs Saved, note SIS (BIS) sometimes reevaluates their data.
- 1980s-2003, use total zone rating includes who fielded hits and outs, often has hit location data.
- pre-1980s, use regression based methods including putouts, assists, pitcher/batter handedness, GB/FB estimates, etc.
- No pitcher defense included (goes into their pitching WAR).
- Todo: Catcher Framing? Assess current measures with Statcast data. Please? Pretty, pretty please?

Position Players – Fielding

Rk	Player	Rfield
1	Brooks Robinson	293.1
2	Mark Belanger	240.7
3	Ozzie Smith	238.7
4	Andruw Jones	235.7
5	Adrian Beltre	229.7

Rk ▲	Player	Rfield
1	Derek Jeter	-246.3
2	Gary Sheffield	-195.3
3	Adam Dunn	-173.5
4	Michael Young	-154.0
5	Bernie Williams	-139.1

Rk	Player	Rfield	F
1	Andrelton Simmons	32.0	:
2	Mookie Betts	31.0	:
3	Byron Buxton	24.0	:
4	Kevin Kiermaier	22.0	:
5	Tucker Barnhart	21.0	:

Rk	Player	Rfield	Fr
1	Denard Span	-27.0	2
2	Jose Reyes	-26.0	2
3	Nicholas Castellanos	-21.0	2
4	Melky Cabrera	-18.0	2
5	Dexter Fowler	-18.0	2

Note first of all that these are only sort of position correlated. Adrian Beltre can rack up positive #'s because managers play Gary Sheffield at 3B. Points to Jeter *chin stroking*, anyways moving on

Position Players – Position

Current values (per 1350 (150*9) innings played) are:

- . C: +9 runs
- . SS: +7 runs
- . 2B: +3 runs
- . CF: +2.5 runs
- . 3B: +2 runs
- . RF: -7 runs
- . LF: -7 runs
- . 1B: -9.5 runs
- . DH: -15 runs
- $\text{RunsPosition_pitchers} = - (\text{RunsBatting_p} + \text{RunsBR_p})$
- Todo: Incorporate Jeff Zimmerman's work. gadel.me/zimmerman-defense & DH value?

We are currently using Sean Smith's estimates from 2008.

Players are given these bonuses no matter how well they played the position.

Second Basemen and Third Basemen are easier to find than Shortstops

Corner outfielders vs Centerfielders

Catchers vs everyone else

Designated Hitters vs everyone else

As a group pitcher offensive contributions sum to zero, since no pitcher is chosen for his offensive contributions.

Position Players – Position

```
on DESC runs_pos;
+-----+-----+
| name_common | runs_position |
+-----+-----+
| Rabbit Maranville | 160.26 |
| Ozzie Smith | 154.19 |
| Luis Aparicio | 152.70 |
| Oran Vizquel | 151.76 |
| Cal Ripken | 149.97 |
+-----+-----+
```

```
on ASC runs_pos;
+-----+-----+
| name_common | runs_position |
+-----+-----+
| David Ortiz | -209.40 |
| Harold Baines | -196.08 |
| Eddie Murray | -189.95 |
| Frank Thomas | -181.45 |
| Carl Yastrzemski | -175.95 |
+-----+-----+
```

Wins Above Average Barry Bonds 1993

- Bonds, 159 Games, +68.3 on offense, +3 on defense.
- Average team, scores 4.93/G, allows 4.50/G.
- Plug into an advanced pythagorean W-L calculator (PythagenPat) to get a W-L% Against Average for the player.
- Winning percentage, .5492.
- $159 * (.5492 - .500) = 7.8$ WAA.

Pitchers – Start with RA

- We use Run Average/9 rather than Earned Run Average/9.
- We have better defensive measures than fielding percentage.

ERA is fielding percentage adjusted pitching.

Compare to a League Average Pitcher

- All pitchers are compared to a league-average pitcher, but one transferred into their context
 - Park context
 - Opponent quality context
 - Defensive support context

Park and opponent quality are two different things.

The Yankees are going to be tough to face at home or on the road.

Coors Field helps both teams dramatically.

Pitchers — Quality of Opponent

- Adjust the league average pitcher by the avg runs/ game of their opponents.
- Also account for games with and without the DH. Some NL pitchers have multiple AL road games and some have none.

If we didn't account for the DH we'd overvalue some players relative to their league and opponents.

Facing an AL team can change a great deal in an NL park. We account for that.

Pitchers — Quality of Opponent

- Rich Hill: 2xARI,ATL,CLE,MIL,3xSDP,2xSTL 4.48 R/G.
- H. Ryu: CHC,CIN,3xCOL,MIN,PIT,WSN 4.64 R/G.

Pitchers - Customize Park Factor

- Park Adjusted, but by the parks they actually pitch in.
- Weighted by batters faced in each park.

Think of a Padres starting pitcher who manages to avoid Coors, and Chase Field and has road games at Citi Field and Marlins Park. Or maybe they were hurt during a road trip or long homestand.

Pitchers with their small sample sizes 34 games vs 162 don't have everything even out in the end, but we could go back and do this for players. It's probably worth studying.

Pitchers - Customize Park Factor

- Ryu, 99.9 PPF
- Hill, 95.4 PPF

Pitchers — Quality of Defense

- ERA is a defense adjusted measure, but based on Fielding percentage.
- We know how many balls in play allowed by pitcher.
- We know the full season results for the team fielders.
- Proportionally assign runs saved or cost to the staff.
- Todo: add catcher framing, adjust SS/3B vs 1B/2B, OF vs IF.

Pitcher — Quality of Defense

name_common	team_ID	IPouts	xRA_def_pitcher
Rick Porcello	BOS	610	10.456
Alex Cobb	TBR	538	8.070
Clayton Kershaw	LAD	525	8.045
Drew Pomeranz	BOS	521	7.929
Chris Sale	BOS	643	7.910

name_common	team_ID	IPouts	xRA_def_pitcher
Jacob deGrom	NYM	604	-7.466
Jordan Zimmermann	DET	480	-6.973
Michael Fulmer	DET	494	-6.412
Jeff Samardzija	SFG	623	-6.392
Ty Blach	SFG	491	-6.087

Pitcher - Combine these Factors

```
desc limit 10;
```

name_common	team_TD	RA_avg_pitcher
Matt Boyd	DET	5.74
Jordan Zimmermann	DET	5.69
Andrew Cashner	TEX	5.66
Kyle Farnsworth	COL	5.64
Jon Gray	COL	5.55
Antonio Senzobela	COL	5.54
Martin Perez	TEX	5.54
Michael Fulmer	DET	5.53
German Marquez	COL	5.52
Justin Verlander	DET	5.51

name_common	team_TD	RA_avg_pitcher
Clayton Kershaw	LAD	4.06
Rich Hill	LAD	4.12
Alex Wood	LAD	4.13
Kenta Maeda	LAD	4.15
Jake Odorizzi	TBR	4.23
Hyun-Jin Ryu	LAD	4.29
Chris Archer	TBR	4.32
Charlie Morton	HOU	4.36
Alex Cobb	TBR	4.38
Mike Piers	HOU	4.44

xRA for an average pitcher with this pitcher's set of defense, parks and opponents.

Matt Boyd had a 5.60 ERA. 87 ERA+, but was by our measure a bit better than average for the year all things considered. Jake Odorizzi had a 4.14 ERA, 100 ERA+, but was well below zero in WAA. Lots of UER, good defense, parks

Pitcher – Runs to WAA

- Add up all of the adjustments and we have the pitcher vs an average pitcher per game.
- Runs to WAA conversion for pitchers is the same as for hitters.
 - Adjust a league average team by the Runs/Game this player is better or worse than average and plug into PythagPat to get a W-L% against average.
 - Multiply by the number of games pitched.

Pitcher — Role Adjustment

- Relievers are typically low-rated by WAR.
- We add a leverage adjustment, so that closers and top setup pitchers get a small bonus.
- Multiplying their WAA by: $(\text{leverage index} + 1.0)/2.0$.
- This also knocks SP's down very slightly as it's zero-sum.

Top reliever last year Archie Bradley 24th with 3.7 WAR, behind Jason Vargas. Added half a win.

whispers relievers are definitely overrated by HOF voters.

Why use a Replacement Level?

- A. Soriano: 1975GP, 412 HR, .819 OPS, **1.0 WAA.**
- T. Neel: 230 GP, 37 HR, .837 OPS, **0.7 WAA.**
- Neel 3.7 WAR, \$350k, Soriano 27.4 WAR, \$159M.

- **If two players have the same playing time in the same year/league, where we set replacement level is irrelevant.**
- Only matters if in different contexts or with different amounts of playing time.

Being average has a lot of value to a team.

So for MVP races what replacement level is really doesn't matter.

Replacement Level for Leagues

- Replacement level is the difference between league average and a “freely available” player.
 - Players who go through waivers, Minor-League Free Agents, Rule 5 Draftees.
 - Non-prospects who play, but aren’t on OD roster.
- AL is more difficult league. We can see this for players who move from AL and NL and vice-versa. Inter-league Play.
- Current total is set at 1,000 (.294 w-l%, 48 wins/team) WAR per season, same as Fangraphs.

Two things we look at here.

Talent level of freely-available players.

Talent differences between the leagues.

Different years can have different values (Integration, WW2, Expansion)

Replacement Level to Players

- WAR = WAA
+ Win below Average
for a replacement player w/same playing time.
- Compute the value of the replacement player relative to average just as you would a player.

Recent Tweaks

- SIS recalculates some of their fielding data using new measures.
- Verify/Update park factors using years N , $N - 1$, $N + 1$.
- Ohtani adjustment. Pitchers who play the field. More accurately split position and replacement level between their pitching and non-pitching playing time. Red Ruffing, Red Lucas, George Uhle.
- No changes bigger than ± 2.5 WAR for an entire career.

Future Study

- Work to adjust replacement level on an ongoing basis.
- Work on adjusting position run values on an ongoing basis.
- Error Bars.
- Defensive metrics.
- Park Factors.
- Position players strength of opponent.
- More data: retrosheet adding PBP every year.

Sports Reference Crew

- Hans Van Slooten (@cantpitch), Mike Kania, David Corby, Mike Lynch, Adam Wodon, Jonah Gardner, Jaclyn Mahoney, Jay Hutchinson, Alex Bonilla, Jay Virshbo
- @sean_forman
- talk slides: gadel.me/2018-sabr-analytics