

Introducing Grid WAR: Rethinking WAR for Starting Pitchers

Ryan Brill and Abraham Wyner

University of Pennsylvania

SABR Analytics Conference, March 2024

Max Scherzer, June/July 2014

game	1	2	3	4	5	total
earned runs	0	1	2	1	1	5
innings pitched	9	6	7	8	7	37

1.2 ERA in
37 innings
≈2 WAR

Max Scherzer, June/July 2014

game	1	2	3	4	5	total	
earned runs	0	1	2	1	1	5	1.2 ERA in 37 innings
innings pitched	9	6	7	8	7	37	≈2 WAR

game	1	2	3	4	5	6	total	
earned runs	0	1	2	1	1	10	15	3.3 ERA in 41 innings
innings pitched	9	6	7	8	7	4	41	≈0.9 WAR

Which pitcher would you rather have?

Pitcher *A*:

game	1	2	3	4	5	6	...	30	avg.
earned runs	5	5	5	5	5	5	...		5
innings pitched	9	9	9	9	9	9	...		9

Pitcher *B*:

game	1	2	3	4	5	6	...	30	avg.
earned runs	1	1	8	1	1	8	...		5
innings pitched	8	8	2	8	8	2	...		9

Which pitcher would you rather have?

Pitcher *A*:

game	1	2	3	4	5	6	...	30	avg.
earned runs	5	5	5	5	5	5	...		5
innings pitched	9	9	9	9	9	9	...		9

Pitcher *B*:

game	1	2	3	4	5	6	...	30	avg.
earned runs	1	1	8	1	1	8	...		5
innings pitched	8	8	2	8	8	2	...		9

Both pitchers average 5 runs/9 innings → 0.0 standard WAR

Which pitcher would you rather have?

Pitcher *A*:

game	1	2	3	4	5	6	...	30	avg.
earned runs	5	5	5	5	5	5	...		5
innings pitched	9	9	9	9	9	9	...		9

Pitcher *B*:

game	1	2	3	4	5	6	...	30	avg.
earned runs	1	1	8	1	1	8	...		5
innings pitched	8	8	2	8	8	2	...		9

Both pitchers average 5 runs/9 innings → 0.0 standard WAR

Pitcher *B* will win \approx 20 of the 30 starts → ≈ 5 “real” WAR

Averaging pitcher performance across games is *wrong*.

- Traditional estimates of WAR for pitchers map pitcher performance *averaged over the entire season* to wins: $\text{WAR}(\mathbb{E}[R])$
 - FanGraphs WAR: map FIP (per inning) or RA (per nine) and IP to wins
 - Baseball Reference WAR: same, with xRA

Averaging pitcher performance across games is *wrong*.

- Traditional estimates of WAR for pitchers map pitcher performance *averaged over the entire season* to wins: $\text{WAR}(\mathbb{E}[R])$
 - FanGraphs WAR: map FIP (per inning) or RA (per nine) and IP to wins
 - Baseball Reference WAR: same, with xRA
- But WAR is a **convex** function of runs allowed:
 - $\text{WAR}(R + 1) - \text{WAR}(R)$ decreases as R increases
 - *If it weren't convex, you could lose a game more than once*

Averaging pitcher performance across games is *wrong*.

- By convexity, Jensen's inequality: $\text{WAR}(\mathbb{E}[R]) \leq \mathbb{E}[\text{WAR}(R)]$
- The consequence of this math:
 - Terrible games cause less damage
 - Extraordinary games become more valuable

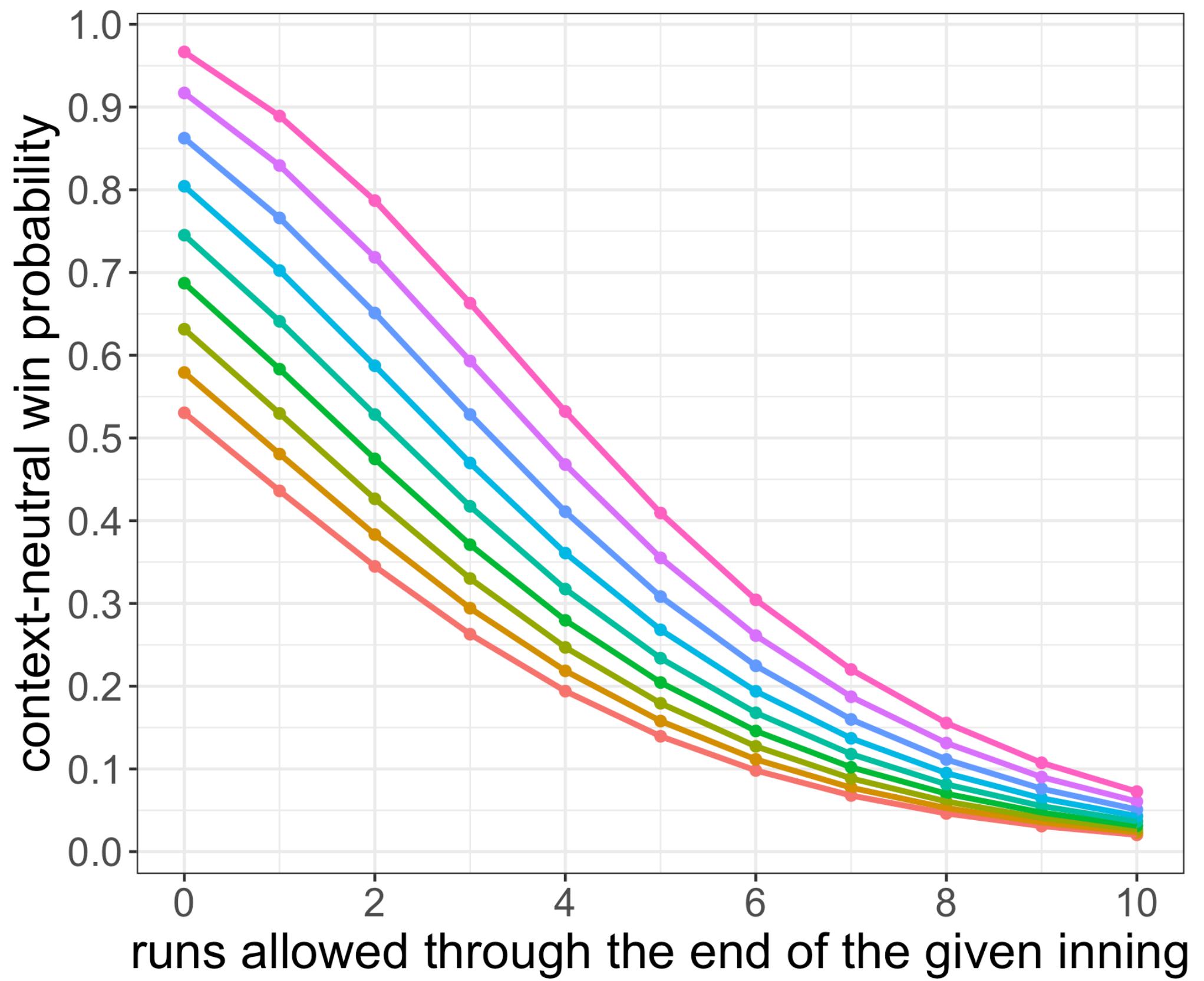
The solution: Game WAR

- In estimating a starter's seasonal WAR, we should not average his performance across games
- Instead, define a starter's WAR in each game
- Seasonal WAR is the sum of the WAR of his individual games

Defining Grid WAR for Starting Pitchers

The grid f

- $f(I, R)$ = the (context-neutral) probability a team wins the game after giving up R runs through I complete innings



inning

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

- *Empirical grid (bin and average)* – overfits massively
- *Machine learning (monotonic XGBoost)* – overfits at the tails (large R)
- **Parametric mathematical (Poisson) model** – a powerful approximation technique, yielding reasonable grids for our application which vary across each league, season, and ballpark without overfitting

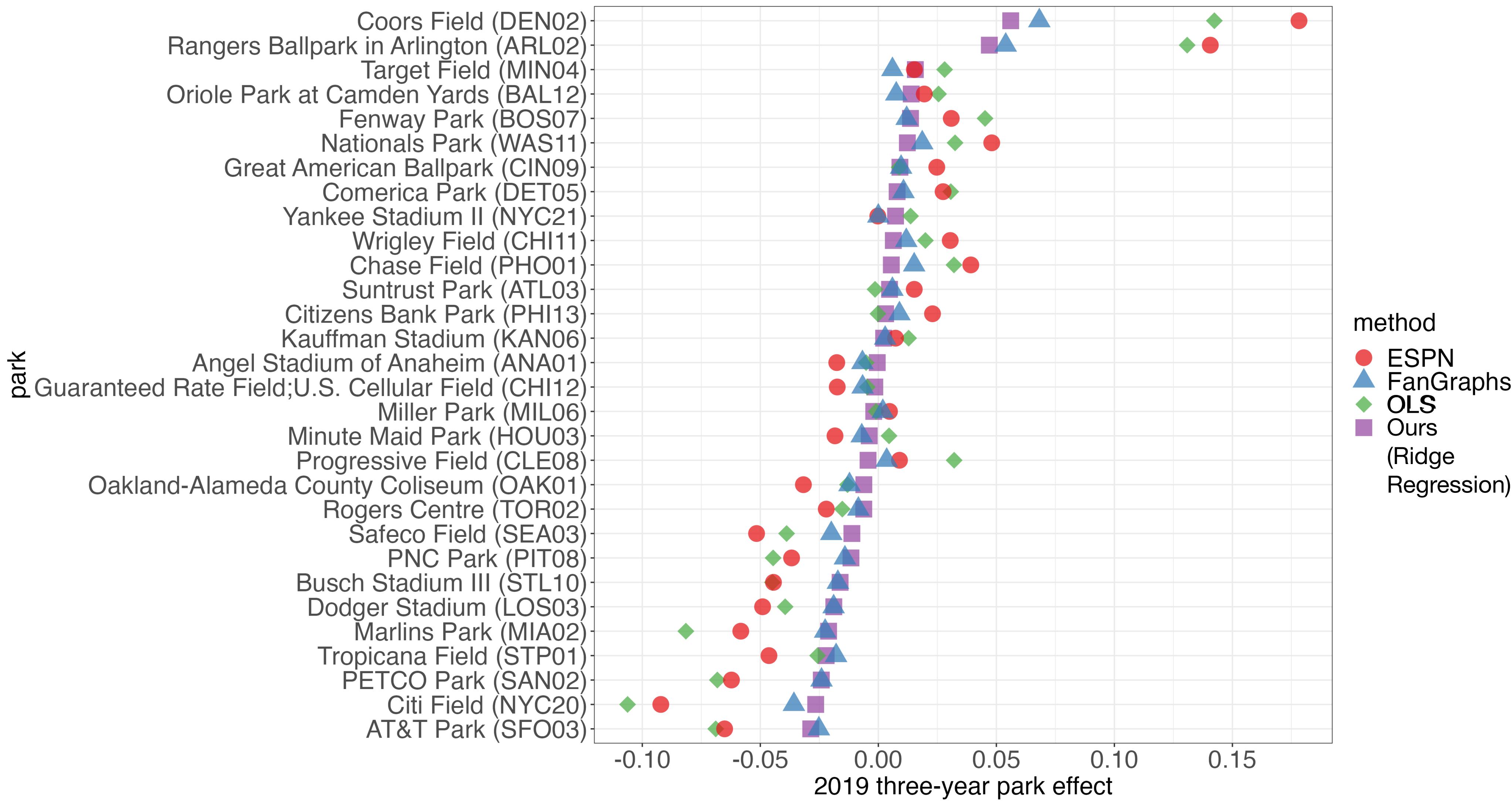
Grid WAR

- A starter's Grid WAR for a game in which he allows R runs through I complete innings is $f(I, R) - w_{rep}$.
- $f(I, R)$ = the probability a team wins the game after giving up R runs through I complete innings (under some assumptions)
- w_{rep} = the probability a team wins a game with a replacement-level starter

Adjusting for park, league, and season

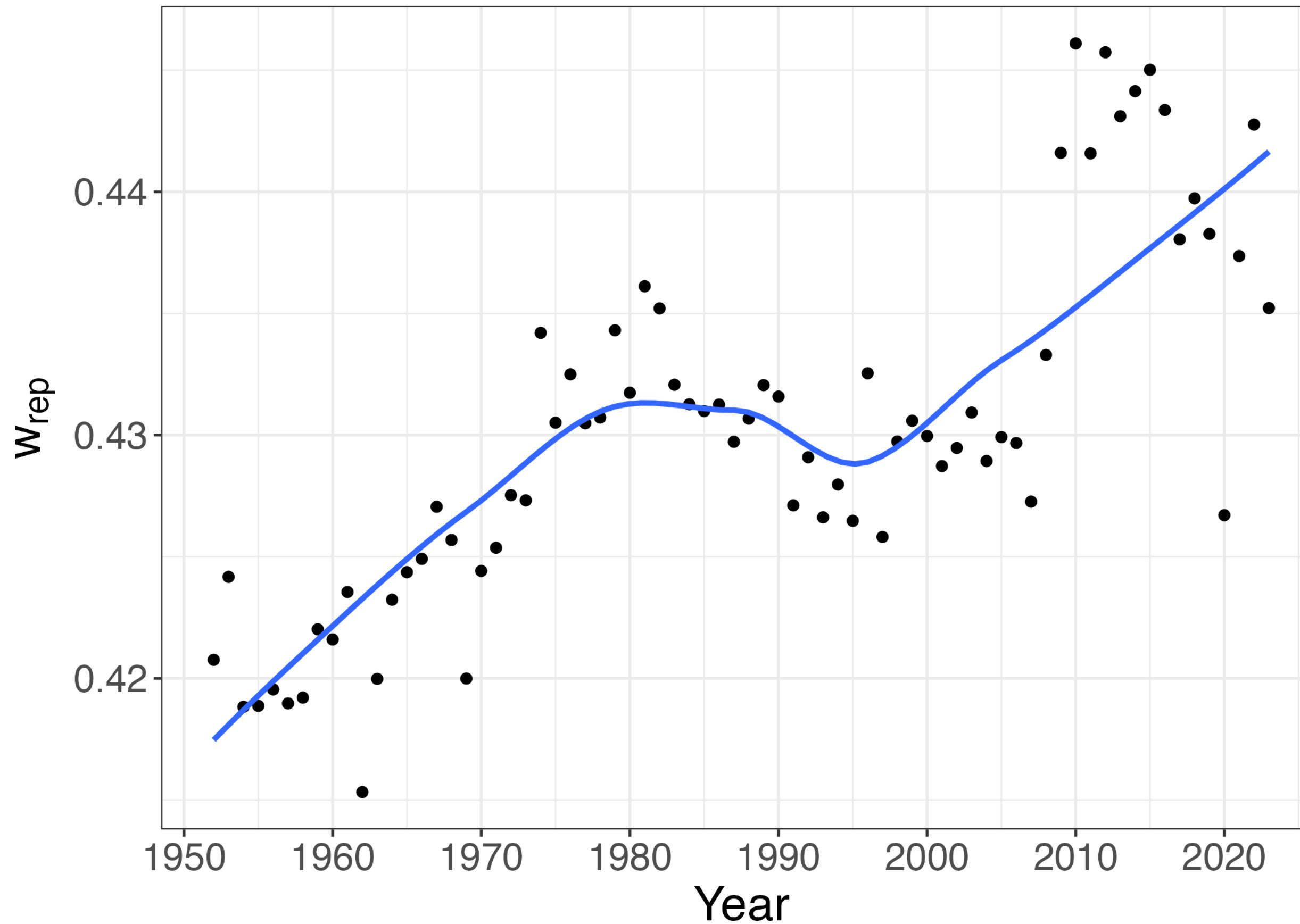
- f is a function of λ = the mean runs scored in a half-inning
- Estimate λ separately for each league-season
- *Park effect* α of a ballpark = the expected runs allowed in a half-inning at that park above that of an average park
- ***Park adjustment: recompute f but replacing λ with $\lambda + \alpha$***

Estimating the park effects α



Estimating w_{rep}

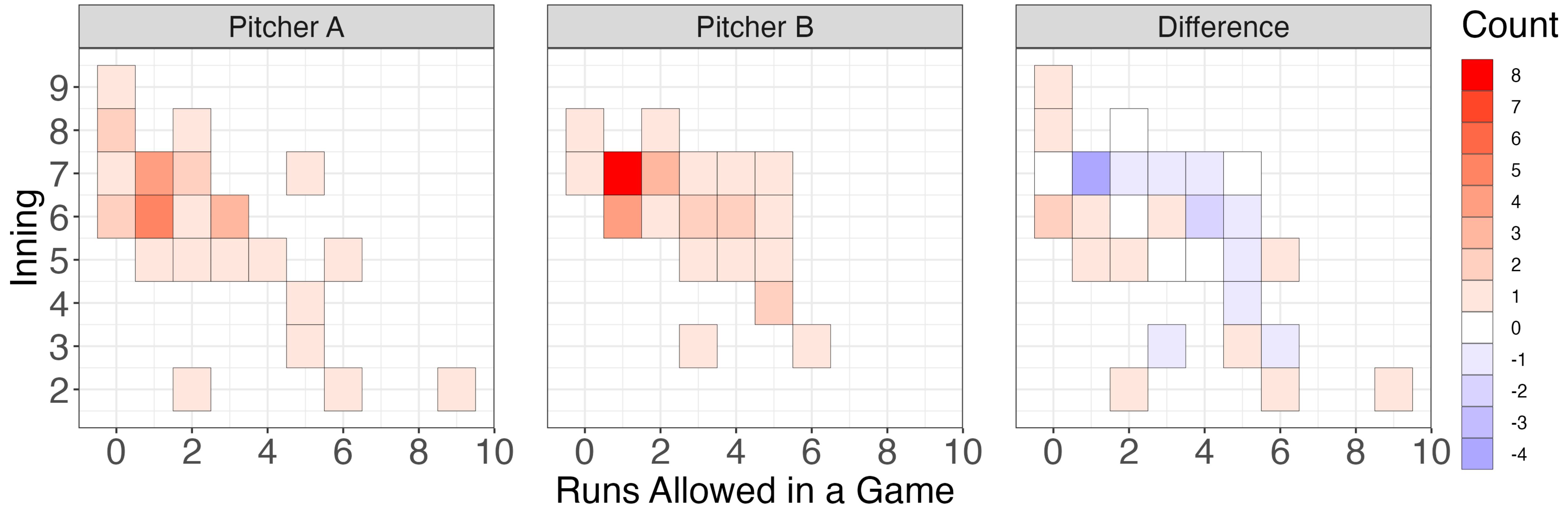
- w_{rep} = the (context neutral) probability a team wins a game with a replacement-level starter



- We estimate w_{rep} to match FanGraphs' definition of replacement-level
- For each season, find w_{rep} such that the sum of all starters' Grid WAR equals the sum of FanGraphs WAR
- Starters have become less valuable over time because they pitch fewer innings per game and relievers have gotten better

Results

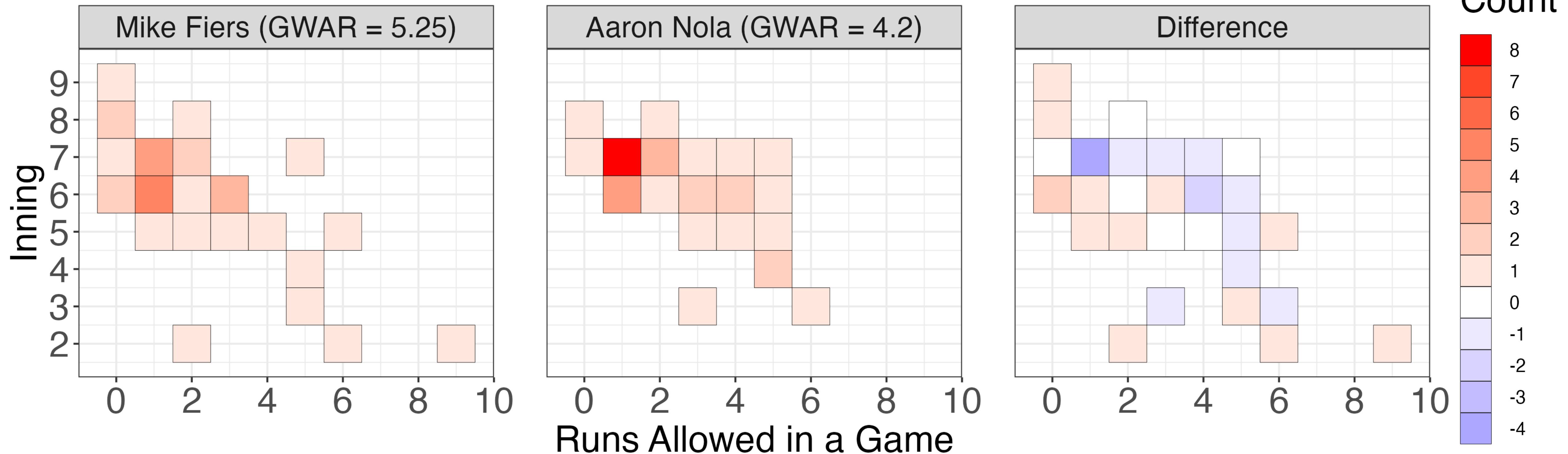
Which pitcher would you rather have?



The color indicates the number of games thrown by Pitcher *A* (left), Pitcher *B* (middle), and the Difference (right) in 2019 with x Runs Allowed through y Innings.

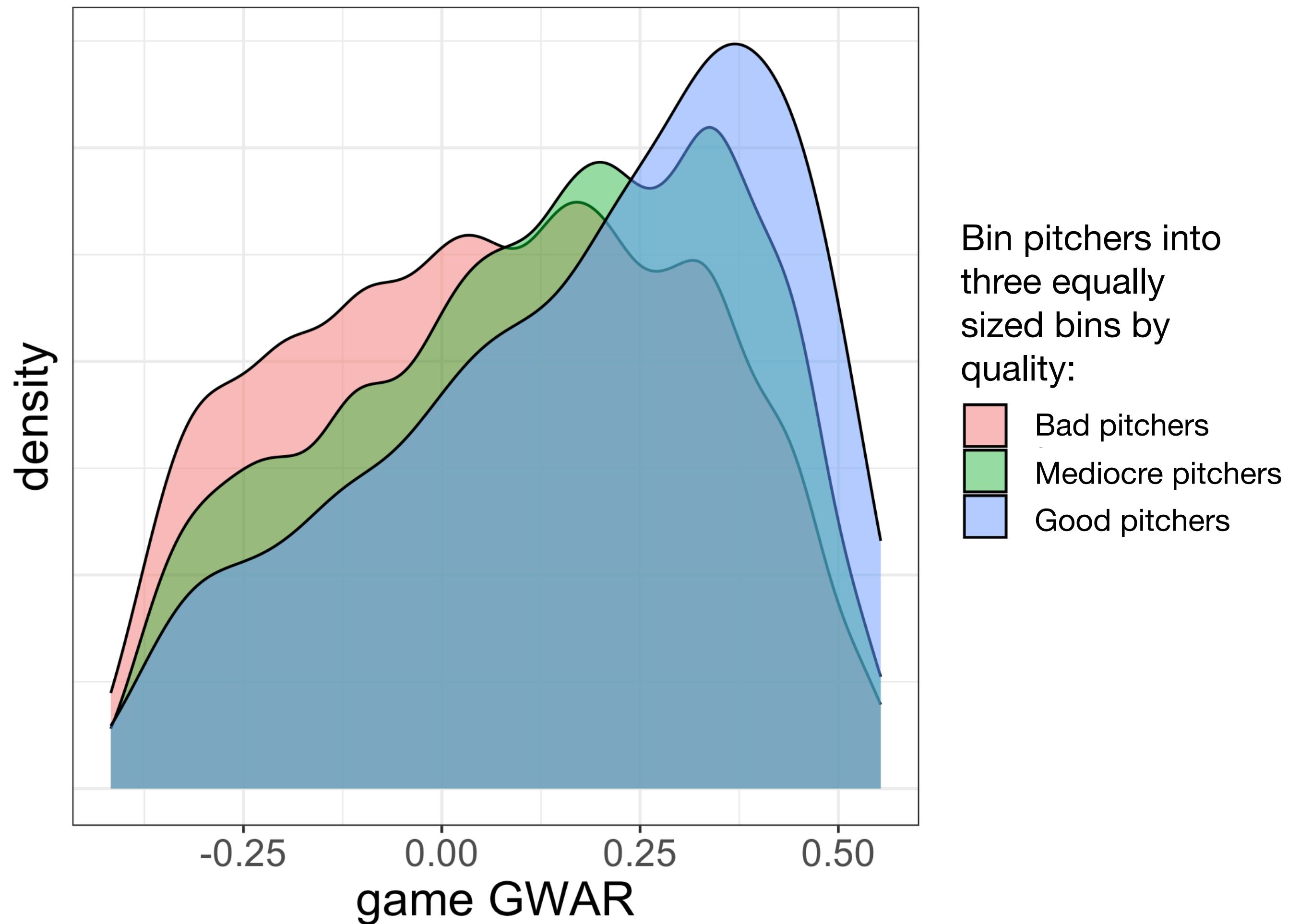
Averaging allows worse games to dilute better games

FWAR (RA/9) = 4.1



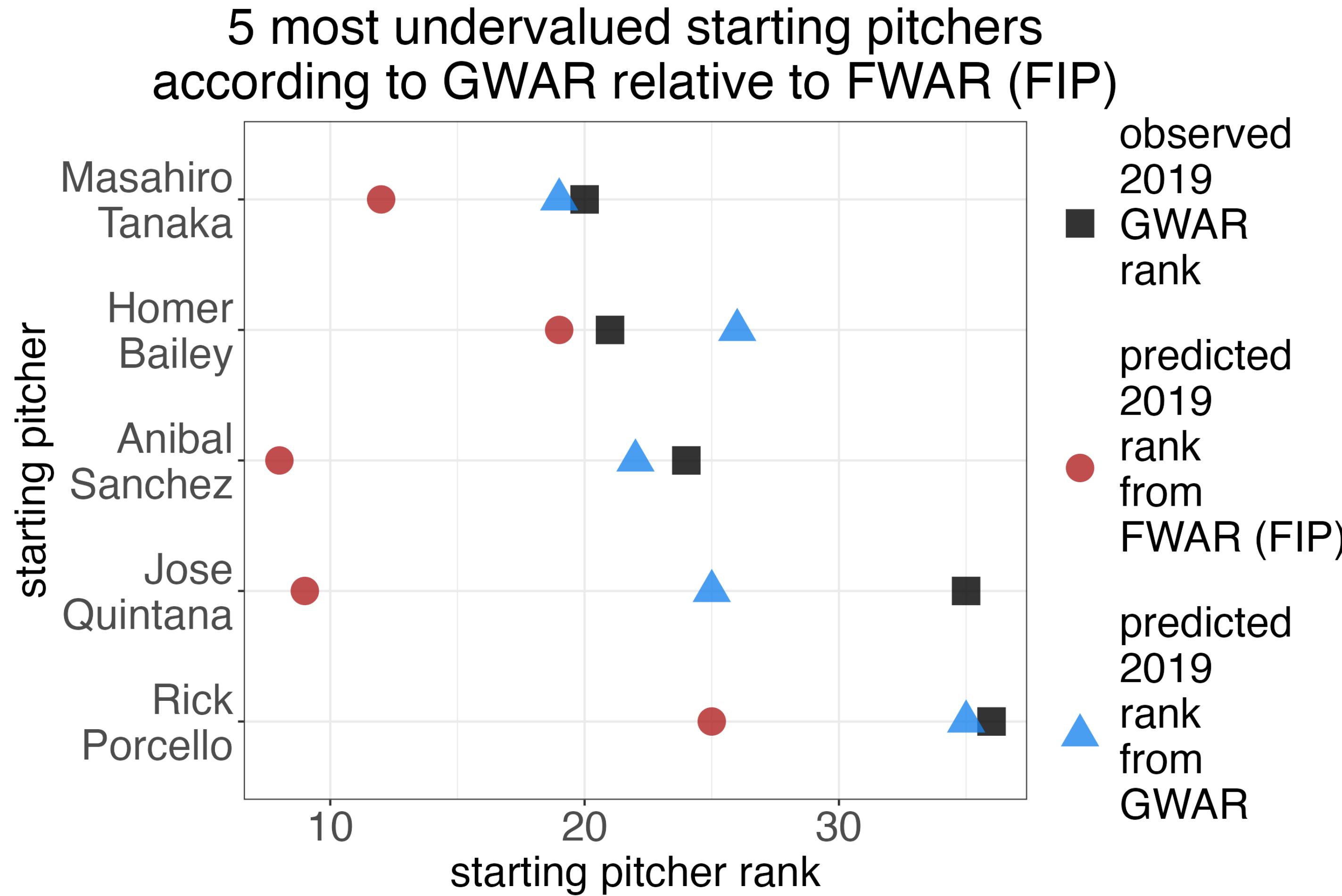
The color indicates the number of games thrown by Fiers (left), Nola (middle), and the Difference (right) in 2019 with x Runs Allowed through y Innings.

Averaging undervalues variance



- All pitchers have great games, but good pitchers (blue) have few terrible games.
- By convexity, Grid WAR caps the impact of these terrible games.
- Thus, averaging generally undervalues high variance pitchers.

Grid WAR is historical, but also predictive!



- Past Grid WAR is more predictive than past FanGraphs WAR of future Grid WAR.
- This suggests game-by-game variance is a fundamental trait of starting pitchers, rather than just purely noise.

Extreme disagreements between GWAR and FWAR

- For many starters, Grid WAR is similar to FanGraphs WAR.
- Grid WAR, however, looks much more favorably upon the careers of some starters with intrinsic game-by-game variance (that is, the occasional tendency to pitch an awful game).

Whitey Ford

	Career WAR (≥ 1952)	Starting pitcher rank (≥ 1952)
FanGraphs	53	49
Grid WAR	78	19
difference	25	30

- ***Ford is an extreme high variance pitcher.***
- 1961 Cy Young winner (7.2 GWAR, 3.7 BBRef WAR, 3.43 RA9):
 - started 39 games
 - 6 complete game shutouts, 18 great games (>0.25 GWAR)
 - 7 blow-up games (< -0.1 GWAR).

The best starting pitcher-seasons

- Which starting pitcher-season (after 1951) has
 - the highest total Grid WAR+?
 - the highest Grid WAR+ per game (min. 25 starts)?
 - *Note: Grid WAR+ adjusts for opponent quality*

The best starting pitcher-seasons (post 1951)

Top 10 starting pitcher-seasons by total Grid WAR+

#	Year	Pitcher	Team	GWAR+	GWAR+ per start	Num Games
1	1966	Sandy Koufax	LAN	11.402	0.278	41
2	1997	Roger Clemens	TOR	11.028	0.324	34
3	1972	Steve Carlton	PHI	10.706	0.261	41
4	1953	Robin Roberts	PHI	10.649	0.260	41
5	1968	Bob Gibson	SLN	10.610	0.312	34
6	1985	Dwight Gooden	NYN	10.387	0.297	35
7	2000	Pedro Martinez	BOS	10.233	0.353	29
8	1963	Sandy Koufax	LAN	10.138	0.253	40
9	1978	Ron Guidry	NYA	9.977	0.285	35
10	1964	Dean Chance	LAA	9.858	0.282	35

Top 10 starting pitcher-seasons by Grid WAR+ per game

#	Year	Pitcher	Team	GWAR+	GWAR+ per start	Num Games
1	2000	Pedro Martinez	BOS	10.233	0.353	29
2	1997	Roger Clemens	TOR	11.028	0.324	34
3	1968	Bob Gibson	SLN	10.610	0.312	34
4	1994	Greg Maddux	ATL	7.668	0.307	25
5	1995	Greg Maddux	ATL	8.574	0.306	28
6	1999	Pedro Martinez	BOS	8.784	0.303	29
7	1953	Warren Spahn	MLN	9.016	0.301	30
8	1985	Dwight Gooden	NYN	10.387	0.297	35
9	1978	Ron Guidry	NYA	9.977	0.285	35
10	1964	Dean Chance	LAA	9.858	0.282	35

- 1966 Koufax: 8 complete game shutouts, 9 one-run complete games, 41 total games, but 3 blow-up games (< -0.1 GWAR); it is just the 6th best season according to FWAR (RA/9) and the 20th best season by FWAR (FIP)
- 2000 Pedro: had zero blow-up games (< -0.1 GWAR), just three negative GWAR games (worst -0.071 GWAR)

The best starting pitcher careers

- Which starting pitcher (after 1951) has
 - the highest career Grid WAR+?
 - the highest career Grid WAR+ per game (min. 100 starts)?
 - *Note: Grid WAR+ adjusts for opponent quality*

The best starting pitcher careers (post 1951)

Top 10 starting pitchers by total career Grid WAR+

#	Pitcher	GWAR+	GWAR+ per start	Seasons
1	Roger Clemens	133.721	0.189	1984-2007
2	Tom Seaver	101.591	0.157	1967-1986
3	Greg Maddux	101.518	0.137	1986-2008
4	Randy Johnson	97.909	0.162	1988-2009
5	Nolan Ryan	96.039	0.124	1966-1993
6	Gaylord Perry	93.902	0.136	1962-1983
7	Don Sutton	92.399	0.122	1966-1988
8	Bert Blyleven	89.995	0.131	1970-1992
9	Jim Palmer	85.752	0.165	1965-1984
10	Steve Carlton	82.867	0.117	1965-1988

Top 10 starting pitchers by career Grid WAR+ per game
minimum 100 starts

#	Pitcher	GWAR+	GWAR+ per start	Seasons
1	Roger Clemens	133.721	0.189	1984-2007
2	Pedro Martinez	76.670	0.187	1992-2009
3	Clayton Kershaw	77.024	0.183	2008-2023
4	Sandy Koufax	57.031	0.182	1955-1966
5	Jacob deGrom	38.870	0.181	2014-2023
6	Whitey Ford	76.547	0.180	1953-1967
7	Johan Santana	47.758	0.168	2000-2012
8	Roy Halladay	64.666	0.166	1998-2013
9	Jim Palmer	85.752	0.165	1965-1984
10	Warren Spahn	74.266	0.164	1952-1965

- Just *three* of the top 25 starters by career Grid WAR began their career after 2000.
- Top starters back then pitched more games per season and pitched more innings per game.

The starting pitcher Hall of Fame

- Who is the best starting pitcher who *didn't make the hall of fame*?
 - excluding active players, recently active players, players mired by steroid controversy, and players mired by other controversy

The starting pitcher Hall of Fame

#	Pitcher	4 Yr Peak GWAR+	Career GWAR+	HOF	Career Rank	Peak Rank	Geo Mean Rank
1	Greg Maddux	30.783	101.730	BBWAA	2	6	3.464
2	Randy Johnson	33.287	98.099	BBWAA	4	3	3.464
3	Roger Clemens	29.726	133.721		1	12	3.464
4	Tom Seaver	31.313	101.591	BBWAA	3	5	3.873
5	Pedro Martinez	33.849	77.132	BBWAA	18	2	6.000
6	Jim Palmer	31.569	85.752	BBWAA	10	4	6.325
7	Sandy Koufax	38.108	57.031	BBWAA	48	1	6.928
8	Gaylord Perry	28.500	93.902	BBWAA	6	16	9.798
9	Justin Verlander	29.730	88.904		9	11	9.950
10	Bob Gibson	30.418	79.473	BBWAA	15	7	10.247
11	Clayton Kershaw	29.036	82.838		12	13	12.490
12	Don Sutton	25.195	92.399	BBWAA	7	27	13.748
13	Bert Blyleven	26.237	89.995	BBWAA	8	24	13.856
14	Nolan Ryan	22.771	96.039	BBWAA	5	51	15.969
15	Jim Bunning	29.978	65.817	Veterans	30	9	16.432

#	Pitcher	4 Yr Peak GWAR+	Career GWAR+	HOF	Career Rank	Peak Rank	Geo Mean Rank
16	Juan Marichal	29.758	66.578	BBWAA	28	10	16.733
17	Robin Roberts	30.271	60.864	BBWAA	35	8	16.733
18	Steve Carlton	25.491	82.867	BBWAA	11	26	16.912
19	Kevin Brown	28.559	75.702		20	15	17.321
20	Whitey Ford	28.263	76.547	BBWAA	19	18	18.493
21	Roy Halladay	28.786	66.574	BBWAA	29	14	20.149
22	Warren Spahn	27.246	74.266	BBWAA	21	20	20.494
23	Don Drysdale	27.588	69.048	BBWAA	26	19	22.226
24	Fergie Jenkins	26.362	72.405	BBWAA	22	23	22.494
25	Max Scherzer	24.855	70.241		23	29	25.826
26	Mike Mussina	22.852	82.012	BBWAA	14	48	25.923
27	CC Sabathia	25.840	68.231		27	25	25.981
28	Tom Glavine	23.380	79.425	BBWAA	16	46	27.129
29	Dave Stieb	28.462	57.560		45	17	27.659
30	Curt Schilling	23.998	65.537		32	36	33.941

Hall of Fame Score = GeomMean(Career GWAR Rank, Peak GWAR Rank)

$$\text{GeomMean}(a, b) = \sqrt{a \cdot b}$$

Introducing Grid WAR: Rethinking WAR for Starting Pitchers

- WAR for starting pitchers should be estimated separately for each game.
- Averaging pitcher performance across his games allows his terrible game performances to dilute his great ones.

- **Thank you!**
- Twitter: [@RyanBrill_](https://twitter.com/RyanBrill_) and [@adiwyner](https://twitter.com/adiwyner)
- Email: ryguy123@sas.upenn.edu
- Grid WAR website: gridwar.xyz
- Paper available online

- Thanks to those who contributed to earlier versions of this project:
 - Eric Babitz
 - Sam Bauman
 - Justin Lipitz, Emma Segerman, Ezra Troy