Example of the Research Process: Rethinking WAR for Starting Pitchers

Suppose I want to start a sports analytics research project.

But, I don't have any ideas right now.

A fantastic way to get started is to simply read about something you're interested about.

Perhaps you were listening to a podcast on which someone mentioned that Roger Clemens has the most career Wins Above Replacement (WAR) of all time, 133.7, according to FanGraphs.

You may have also heard that Pedro Martinez in 1999 has the highest single season WAR of all time, 11.6, according to FanGraphs.

You may think that WAR is a really cool concept, and it makes some intuitive sense

over why it seems like a nice way to evaluate pitchers, and more generally, all players.

Wins Above Replacement — replace a player with a replacement-level player (e.g. the best guy you could get on waivers), how many fewer wins would the team have, assuming average teammates and opponents?

Implementation take a player's observed performance, ignoring/adjusting for things that he is not responsible for, and map that to wins

Say you don't know the math behind WAR, although you are curious to learn. So, you **Read**.

The most widely used/accepted public WAR implementations are from FanGraphs and Baseball Reference.

FanGraphs WAR for pitchers:
https://library.fangraphs.com/war/calculating-war-pitchers/

Baseball Reference WAR for pitchers:
https://www.baseball-reference.com/about/war_explained_pitch.shtml

When you Read about WAR for pitchers, a few things catch your eye:

* WAR involves mapping a pitcher's performance (e.g., FIP for FanGraphs, xRA for Baseball Reference) to Wins

$$ifFIP = \frac{13 \cdot HR + 3 \cdot (BB + HPB) - 2 \cdot (K + IFFB)}{IP} + C$$

Fielding Independent Pitching (with Infield Flies!)

The first thing you need to do to calculate a pitcher's WAR is to calculate their FIP. Unfortunately for those of you playing along at home, you can't simply take the pitcher's FIP from their player page because we treat **infield fly balls (IFFB) as strikeouts for the purposes of WAR** but not for the general FIP calculation found on the player's page. We'll call this ifFIP to avoid confusion. Here is the formula:

ifFIP = *((13*HR)+(3*(BB+HBP))-(2*(K+IFFB)))/IP + ifFIP constant*

This is the traditional FIP formula, but with IFFB added in as strikeouts. However, keep in mind that you also need to calculate a special ifFIP constant and can't just grab "cFIP" from our guts page.

*ifFIP Constant = lgERA – (((13*lgHR)+(3*(lgBB+lgHBP))-(2*(lgK+lgIFFB)))/lgIP)*

xRA = expected runs allowed
= ignoring the ordering e.g. 1B, out, out, 1B, HR, out
v.s. HR, 1B, 1B, out, out, out
and just using the events 1 HR, 2 1B, 3 out,
what is the expected runs allowed of the inning?

* there are a series of convoluted adjustments on top of the base performance metric
(e.g., league adjustment
team defense adjustment)

* WAR involves mapping a pitcher's performance averaged over the entire season into wins
ifFIP: divides by IP
xRA: cumulative seasonal xRA

Thoughts: averaging pitcher performance over the course of a season seems weird

Let's explore some implications of this modeling assumption.

Ex

game	1	2	3	4	5	6	total
earned runs	0	10	1	2	1	1	15
innings pitched	9	4	6	7	8	7	41

Table 1: Max Scherzer's performance over six games prior to the 2014 All Star break.

4 dominant performances $\longrightarrow$ ≥ 4 wins

$\frac{15 \text{ Runs}}{41 \text{ IP}} \times 9 \frac{\text{innings}}{\text{game}} = \frac{3.66 \text{ Runs}}{\text{Complete game}}$

$3.66 \frac{\text{Runs}}{\text{Complete game}} \approx 0.55 \frac{\text{Win probability}}{\text{Complete game}}$

$\longrightarrow \approx 3.30$ wins over 6 games

Big difference b/t ≥ 4 and 3.3 wins!

Ex Would you rather have pitcher A or pitcher B?

A: 5 Runs in each game

B: Alternates b/t 10 and 0 Runs in each complete game

All else the same, existing WAR methodologies value these 2 pitchers the exact same.

Would rather have pitcher B though...

Ex Pitcher A: alternates between allowing 7 and 0 Runs per complete game

Pitcher B: alternates between allowing 14 and 0 Runs per complete game

Existing WAR: A ~ 3.5 runs/game
B ~ 7 runs/game
A >> B

"Real" WAR: Both A and B win half of their games.
A ≈ B

"You can only lose a game once"

"Not all Runs have the same value"

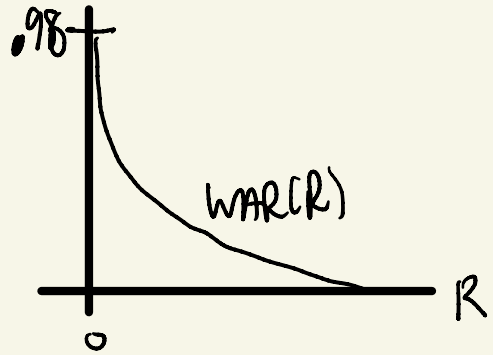
→ the 8th Run allowed in a game is "worth" less than the 1st

→ the marginal difference in Win probability between allowing the 7th vs. 8th Run is less than the marginal difference in game WAR between allow

the 1st and 2nd Run, since you're essentially already lost the game

→ if $R \mapsto WAR(R)$ is game WAR as a function of Runs allowed, then $WAR(R) - WAR(R-1)$ gets smaller as R gets bigger

→ $R \mapsto WAR(R)$ should be convex



Problem: Averaging pitcher performance over the course of the season is clearly wrong

- it ignores the game-by-game variance in pitcher performance
- $R \mapsto WAR(R)$ should be convex (i.e., not all Runs should have the same value)
- a win is the fundamental result of a game, not a season

Goal: Fix this problem.
Make one incremental improvement [Research].

→ Calculate historical WAR in each individual game
seasonal WAR is the sum of game WAR.

Task Game WAR for starting pitchers

How to do this? English → Math™

- **WAR** WAR = wins above Replacement
- **Wins** W = How many wins did Scherzer contribute in this game?
- **Above Replacement** W_{Rep} = How many wins a replacement-level pitcher would've contributed

$$WAR = W - W_{Rep}$$

One step at a time Begin with **Wins** W

Wins How many wins did Scherzer contribute in this game?

Math: win probability

Pitcher valuation: we only want to judge Scherzer using things he's responsible for

Scherzer's game performance: Runs Allowed R → because winning a game is defined by Runs
exit inning I
exit base-state S
exit outs O

Confounders, e.g. variables that affect his performance:

park
opposing team's batting quality
his team's fielding quality

contextual variables that affect the win probability

league (NL vs. AL), season

variables that don't affect his performance and so we shouldn't judge him with:

his team's batters/opposing team's defense

Start Simple Begin with the easiest version of the task. Then, iterate on top of that.

* Begin just with Scherzer's observed performance. Adjust for confounders later.

Task given Scherzer's performance, what's his team's win probability when he exits the game?

- Runs Allowed R
- exit inning I
- ~~exit base-state S~~
- ~~exit outs O~~

context-neutral: assume league-average offenses, defenses, ignore his own team's Runs scored

Start simple: assume he finishes the inning, so ignore (S, O)

Model the function

$$f = f(I, R) =$$

assuming both teams have league-average offenses, compute the probability a team wins a game after giving up R runs through I complete innings

Since $f(I,R)$ can be visualized as a 2D grid, we name our WAR **Grid WAR.**

This is the simplest version of the question, and it is nontrivial.

Takeaway: 2 great ways to do Research in applied statistics (esp. sports)

---

1. "Read First": Read a paper/article/blogpost about sports statistics.
   Replicate it.
   Check what else has been done and replicate the state of the art.
   Find one thing you don't like
   Make one incremental improvement.

   ↳ e.g. GRid WAR

2. "Think First": Think of a cool idea.
   Read relevant literature and replicate state of the art, if any.
   Solve the problem.

   ↳ e.g. My NCAA brackets paper