Statistics in Sports: What Makes a Good Sports Statistic?

ZACHARY BINNEY, PHD MPH

OXFORD COLLEGE OF EMORY UNIVERSITY

FALL 2023

Roadmap

- What Makes a Good (Sports) Statistic?
  - Does it match your question?
  - Does it measure success in the sport? (**Validity**)
  - Is it noisy or repeatable/stable? (**Reliability**)
  - Is it useful?
  - Is it better than an alternative?

What Makes a Good (Sports) Statistic?

Does Your Stat Match Your Question?

- A statistic starts with a QUESTION
- It should be designed to answer that QUESTION
- **A statistic may be good for one QUESTION but not another**
  - Be *transparent* and *explicit* about what your QUESTION is and what the statistic is designed to do – and, critically, *not* do!

Source: https://twitter.com/bernybelvedere/status/1171435503586926593

Three Types of Questions

- **1. Descriptive**
  - Describing the world as it *is*
- **2. Predictive**
  - Describing the world as it *will be*
- **3. Causal (Counterfactual Prediction)**
  - Describing the world as it *could be*

Does Your Stat Match Your Question?

- Player-level stat or team-level stat?
  - Differences?
  - In which do you want to isolate the contributions of an individual player?

Does Your Stat Match Your Question?

- How good is this running back? – RYOE
- How *accurate* is this quarterback? – CPOE
- How *good* is this quarterback? – EPA per play QB is involved in
- How *valuable* is this quarterback? – Total EPA, WAR

Does Your Stat Match Your Question?

- Teams: How much should I pay this player? – WAR, and make sure it has *predictive* not just *descriptive* value

- Fantasy value of wide receiver? – Target share

Does Your Stat Match Your Question?

- Who's the GOAT? – Need era-adjusted statistic
  - Some stats aren't, because <u>they're meant for another question!!</u>

Source: https://www.tvovermind.com/is-the-new-match-game-doing-justice-to-the-original/

Does It Measure Success in the Sport?

- Does the statistic correlate with some gold standard measure of "**success**?"

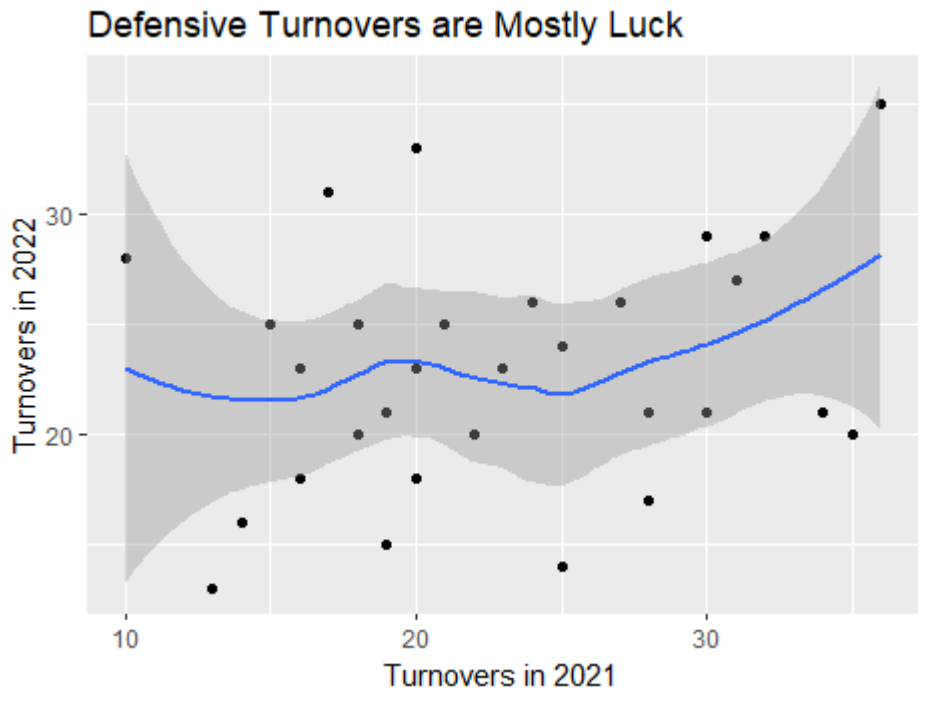
- Step back: what is the goal/purpose of a sports team? What defines its "success?"

Does It Measure Success in the Sport?

- “Success” depends on your question, but common options:
  - Wins for holistic stats
  - Points/goals/runs for offensive stats
  - Points/goals/runs against for defensive stats

Is The Stat Repeatable/Stable?

- Is it measuring **signal** or **noise**? “True talent” or luck?
- Look at year-to-year (or split half season) correlations



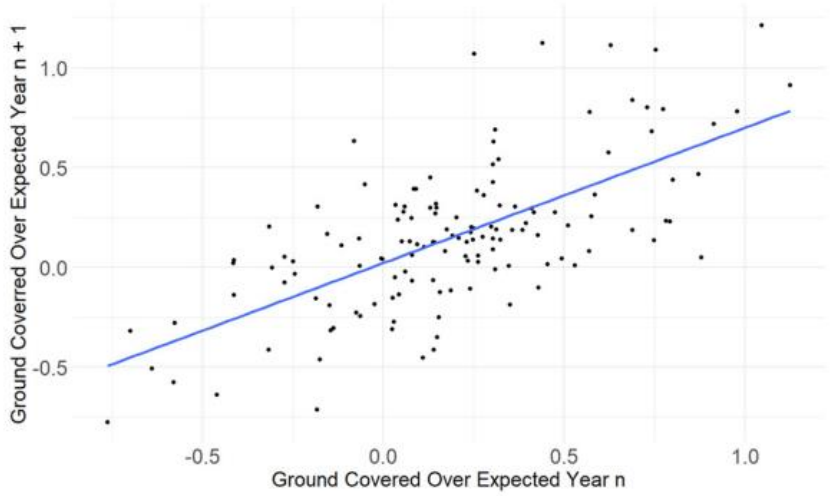
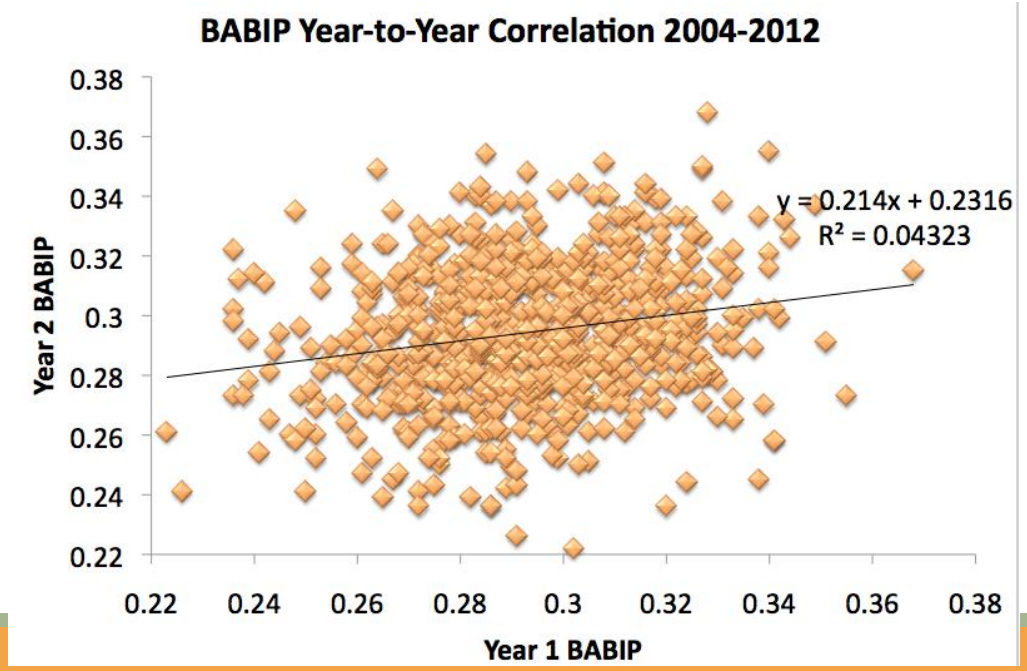


Figure 10: Year-to-year relationship between ground covered over expected (GCOE) and itself. Minimum 200 run-defense snaps in a box position.

Source: Eager and Seth, JQAS 2023

Is The Stat Repeatable/Stable?

- Voros McCracken article: batting average on balls in play not repeatable and depend on defense → not a characteristic of a pitcher. Noise.
- Remove it to get better idea of "true" pitcher skill – focus only on what they control (walks, strikeouts, home runs, hit batters)



Source: https://www.baseballprospectus.com/news/article/878/pitching-and-defense-how-much-control-do-hurlers-have/, https://www.beyondtheboxscore.com/2013/1/4/3827868/projecting-babip-regression-toward-the-mean-baseball-math-sabermetrics

Is It Useful?

- Can you use it to **predict** or **intervene** on something?
  - Predict: Will it help me better predict player performance over the duration of a contract?
  - Intervene: Does it give me new knowledge about the game I can use to change strategies, player development, etc.?
  - Not *necessary* for stat to be good (?), but increases value a lot

Is It Better Than an Alternative?

- Does it correlate more with “success in the sport” than the stat you propose it replace?
  - How does that balance against “complexity” cost?
    - Cost to gather/analyze data
    - Cost of complexity/explanation/adoption

Thanks!

- Questions? zbinney@emory.edu, @binney_z on Twitter

Appendix – Baseball Stats and Quality Metrics

Batting

- DISCLAIMERS: In no way even a comprehensive list of baseball batting player valuation stats
  - Not using Statcast data (e.g. out-of-zone swing %s)

- Many different “flavors” of some stats with different formulas – we use Fangraphs

Batting

- Runs (R) and Runs Batted In (RBIs)
  - Problems/Limitations? **(For this and all subsequent slides, don't look forward until I say so.)**

R and RBI

Batting

- Batting Average (BA) = $\frac{Hits\ (H)}{At\text{–}Bats\ (AB)}$

  - **Rate stat** vs. **counting stat** – solves denominator problem
  - Also improves dependence on other players
  - Problems/limitations?

R and RBI → BA

Batting

- On-Base Percentage (OBP) = $\frac{H + Walks\ (BB) + Hit\ by\ Pitch\ (HBP)}{AB + BB + HBP + Sacrifice\ Flies\ (SF)}$

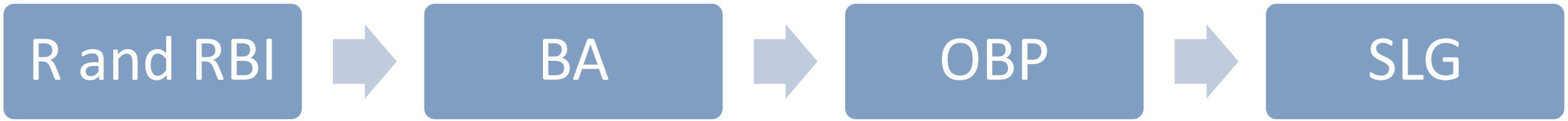
  - Solves undervaluing of walks, HBP
  - Problems/limitations?

≈ Plate Appearances (PA)

R and RBI → BA → OBP

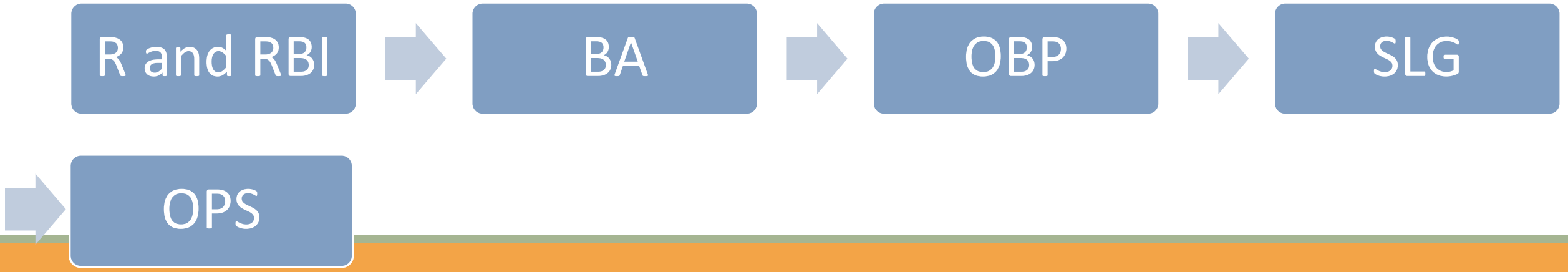
Batting

- Slugging Percentage (SLG) = $\frac{Total\ Bases\ (TB)}{AB}$
  - Accounts for power, XBH and homers
  - Problems/limitations?



Batting

- On Base Plus Slugging (OPS) = $OBP + SLG$
  - Closer to all-encompassing stat for batting skill
  - Problems/limitations?



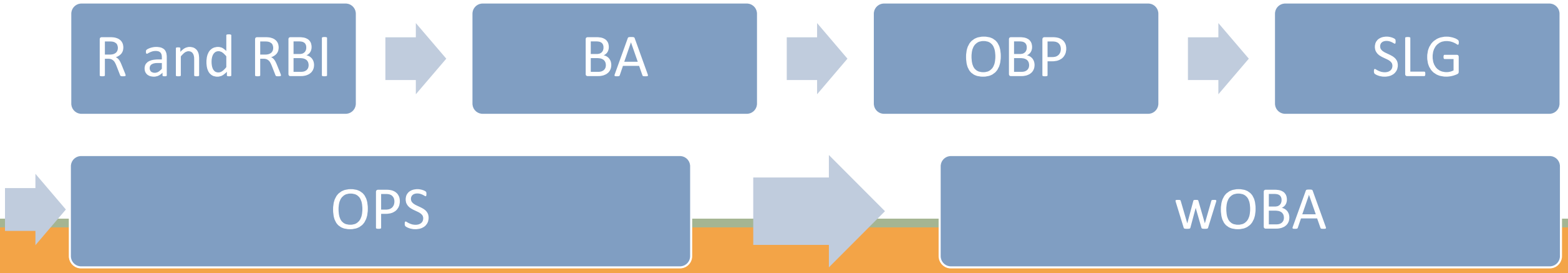
Batting

Weights vary slightly by season based on season OBP

- Weighted On Base Average (wOBA) =

$$\frac{0.69*uBB + 0.72*HBP + 0.89*Singles\ (1B) + 1.28*2B + 1.64*3B + 2.14*HR}{AB + BB - Intentional\ Walks\ (iBB) + SF + HBP}$$

  - More accurate weights than SLG based on **run expectancy**
  - Scaled to OBP, interpreted similarly
  - Problems/limitations?



Run Expectancy

- **Run Expectancy (RE):** every PA moves you from one state to another
- Ex.: leadoff single goes from 0.481 to 0.859 → worth 0.378 expected runs
- Average over all singles → RE for singles, which is (part of) the singles weight for wOBA

Run Expectancy Matrix 2010-2015			
Runners	0 outs	1 outs	2 outs
_ _ _	0.481	0.254	0.098
1B _ _	0.859	0.509	0.224
_ 2B _	1.100	0.664	0.319
1B 2B _	1.437	0.884	0.429
_ _ 3B	1.350	0.950	0.353
1B _ 3B	1.784	1.130	0.478
_ 2B 3B	1.964	1.376	0.580
1B 2B 3B	2.292	1.541	0.752
SOURCE: Tom Tango			

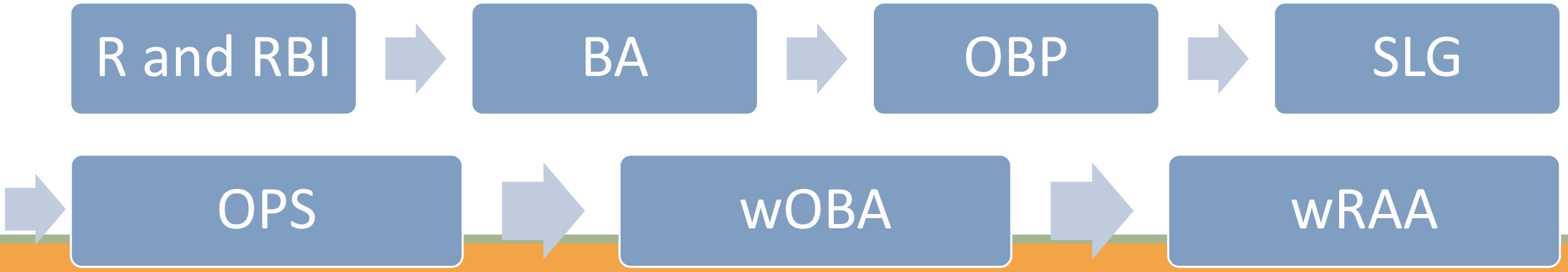
Source: https://library.fangraphs.com/the-beginners-guide-to-deriving-woba/

Batting

- Weighted Runs Above Avg (wRAA) = $\frac{wOBA - League\ avg.wOBA}{wOBA\ Scale} * PA$

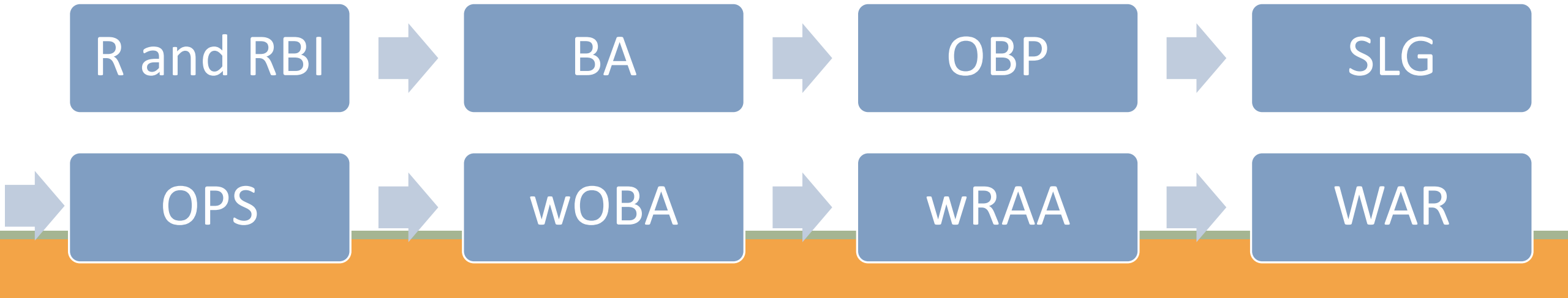
Undoes scaling so wOBA represents actual expected runs

  - Translates wOBA to actual runs above *average* player
  - Problems/limitations?



Batting

- (Batting) Wins Above Replacement (WAR) = *Complicated*
  - Adjusts for park, league, position; compares to **replacement player**
  - Then translates runs to wins (1 win ≈ 9-10 runs)
  - **Ultimate value. How much $ is a win worth x (total) WAR = contract value**



Pitching

- Wins (or Win-Loss Record, W-L)

Indians

In pitching, the only thing that really matters is wins: Paul Hoynes' Rant of the Week

Updated Jan 12, 2019; Posted Sep 12, 2010

- Problems/limitations?

W-L

Source: https://www.cleveland.com/tribe/2010/09/in_pitching_all_that_matters_a.html; https://nesn.com/2014/04/mets-pitcher-bartolo-colon-swings-hard-misses-pitch-loses-helmet-gif/

Pitching

- Earned Run Average (ERA) = $\frac{Runs\ Not\ the\ Result\ of\ an\ Error}{Innings\ Pitched\ (IP)} * 9$

  - Accounts for *errors* by defense, eliminates offensive performance
  - Problems/limitations?

W-L → ERA

Breaking Down Pitching Performance

- Fundamental problem:

$$Pitching\ Performance = Pitcher\ Skill + Defense\ Skill + Luck$$

- How do we isolate *just* pitcher skill?

- Enter **Defense-Independent Pitching Statistics (DIPS)** – despite name, try to deal with defensive skill *and* luck

Source: https://i.gifer.com/OOxX.gif

Breaking Down Pitching Performance

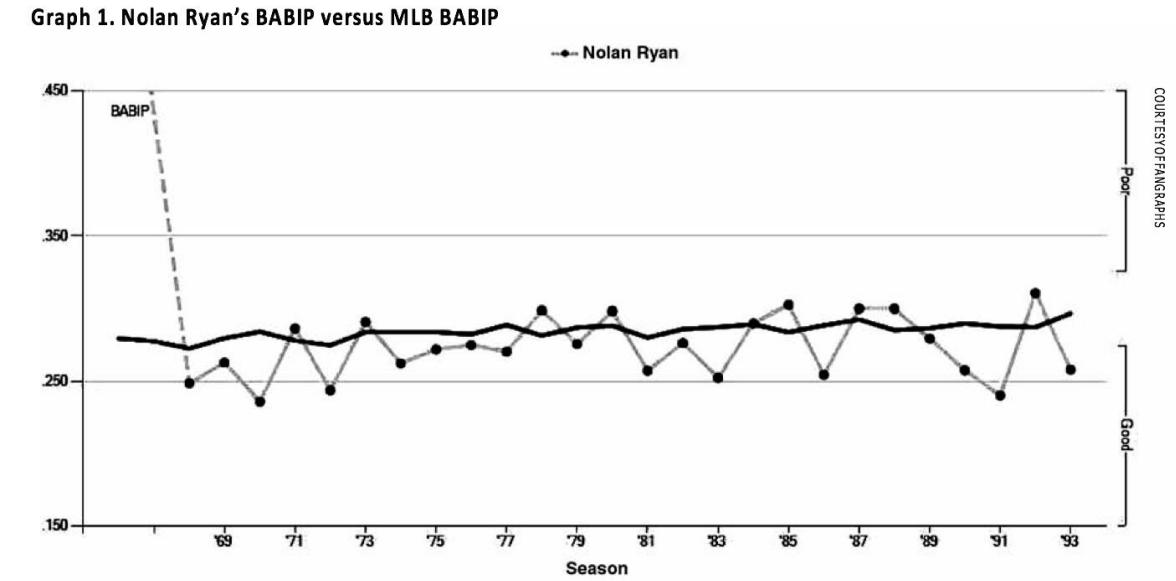
- $Pitching\ Performance = Pitcher\ Skill + Defense\ Skill + Luck$

- In any plate appearance, the ball can end up **in-play** (must be fielded) or **not-in-play** (strikeout (K), BB, HBP, HR)

  - Voros McCracken's key insights:
    - 1) We should split outcomes into these categories
    - 2) Pitchers have complete control over **not-in-play** outcomes, but are at the mercy of the defense for **in-play** balls

Batting Average on Balls in Play (BABIP)

- $Pitching\ Performance = Pitcher\ Skill + Defense\ Skill + Luck$
- Strong year-to-year correlations among pitchers for strikeouts and walks (≈0.7-0.8 or higher), pretty strong for HRs (≈0.4-0.5)
- Much weaker for BABIP (≈0.15-0.25)



Source: http://sabr.org/journal/article/the-many-flavors-of-dips-a-history-and-an-overview/

Pitching

To make FIP read like ERA

- Fielding Independent Pitching (FIP) = $\frac{13*HR+3*(BB+HBP)-2*K}{IP} + Constant$
  - Only measures things a pitcher has or appears to have control over ("defense-independent statistics" that are consistent year-to-year); excludes all BIP
  - Problems/limitations?

W-L → ERA → DIPS (FIP)

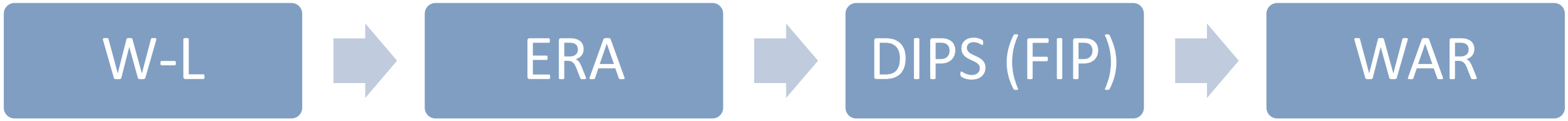
Ground and Fly Balls

- Of course it's not really that simple
- Pitchers don't control *overall* BABIP, but do control:
  - % ground balls (GB)
  - % fly balls (FB)
  - Ratio GB/FB
  - Etc.
- Stats like tRA, QERA, SIERA take this batted ball data into account
- In-season vs. next season projections

Source: https://medium.com/@mxoliverleigh/really-it-should-not-be-this-complicated-fd8a13056747

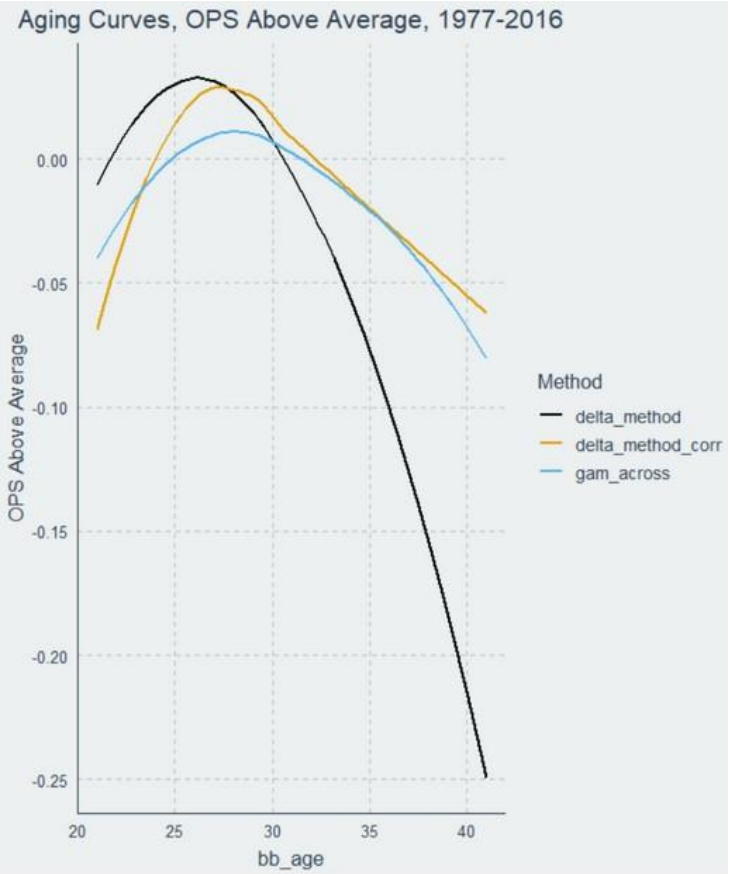
Pitching

- (Pitching) WAR = *Complicated, again*
  - Adds infield flies to FIP, translates to runs, adjusts for park and league, then translates to wins (but here every pitcher has a different runs per win metric), compares to replacement level
  - Remember this is Fangraphs WAR; other sources may use something other than FIP as basis
  - **Ultimate value. How much $ is a win worth x (total) WAR = contract value**



Predicting Future Value

- Different from **cross-sectional** stats describing what an athlete *is*, we need to know what an athlete *will be*
- Two challenges: 1) Try to identify a player's **true talent**, and then 2) project that forward
- **Repeatability:** look for stats that are either stable year-to-year, or at least predictive of future performance
  - Not luck-based
  - Adjust for things you know will happen (e.g. **aging curves**)



Source: https://www.baseballprospectus.com/news/article/59972/the-delta-method-revisited/, https://arxiv.org/abs/1802.00998

The Scientific Process

- Is there a better alternative to measure what you're trying to measure?
  - **Science is built on accumulating small answers that lead to big insights – "get a little less wrong each day"**
  - **No single analysis answers everything – understand and accept that**
  - BUT sometimes an analysis is worse than no analysis at all; judgement call