OpenWAR:
An Open Source System for Overall Player Performance in MLB

Ben Baumer\textsuperscript{1}  Shane Jensen\textsuperscript{2}  Gregory Matthews\textsuperscript{3}

\textsuperscript{1} Smith College
\textsuperscript{2} The Wharton School
University of Pennsylvania
\textsuperscript{3} University of Massachusetts

Joint Mathematical Meetings
Baltimore, MD
January 17th, 2014
WAR - What is it good for?

- **Wins Above Replacement**
- Question: How large is the contribution that each player makes towards winning?
- Four Components:
  1. Batting
  2. Baserunning
  3. Fielding
  4. Pitching
- Replacement Player: Hypothetical 4A journeyman
  - Much worse than an average player
Units and Scaling

- In terms of absolute runs:

- In terms of Runs Above Replacement (RAR):

- In terms of Wins Above Replacement (WAR):
Example: 2012 WAR leaders

<table>
<thead>
<tr>
<th>FanGraphs</th>
<th>$fWAR$</th>
<th>BB-Ref</th>
<th>$rWAR$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike Trout</td>
<td>10.0</td>
<td>Mike Trout</td>
<td>10.9</td>
</tr>
<tr>
<td>Robinson Cano</td>
<td>7.8</td>
<td>Robinson Cano</td>
<td>8.5</td>
</tr>
<tr>
<td>Buster Posey</td>
<td>7.7</td>
<td>Buster Posey</td>
<td>7.4</td>
</tr>
<tr>
<td>Ryan Braun</td>
<td>7.6</td>
<td>Miguel Cabrera</td>
<td>7.3</td>
</tr>
<tr>
<td>David Wright</td>
<td>7.4</td>
<td>Andrew McCutchen</td>
<td>7.2</td>
</tr>
<tr>
<td>Chase Headley</td>
<td>7.2</td>
<td>Adrian Beltre</td>
<td>7.0</td>
</tr>
<tr>
<td>Miguel Cabrera</td>
<td>6.8</td>
<td>Ryan Braun</td>
<td>7.0</td>
</tr>
<tr>
<td>Andrew McCutchen</td>
<td>6.8</td>
<td>Yadier Molina</td>
<td>6.9</td>
</tr>
</tbody>
</table>

Table: 2012 WAR Leaders

- Baseball Prospectus also publishes **WARP**
- There is no ONE formula for WAR!
WAR is the Answer
What’s Wrong with WAR?

- **Not Reproducible**
  - WAR is an unknown hypothetical quantity – not a statistic
  - No reference implementation of WAR
  - No open data set
  - No open source code

- **No unified methodology**
  - Each component of WAR is viewed as a separate problem – not a piece of the same problem
  - Ad hoc definitions: what is replacement level?

- **No error estimates**
  - Only reported as point estimates
  - Only hand-wavy estimates of variability or margin or error

- **Bug or Feature?: Competing black-box implementations**
Our Contribution: openWAR

- **openWAR**: a reproducible reference implementation of WAR
  - Principled *estimate* of WAR
  - Fully open-source R package (free as in freedom)
  - Partially open data (free as in beer)

- Unified Methodology:
  - Conservation of Runs
  - Each component is estimated as a piece of the larger problem

- Error estimates:
  - Use resampling methods to report WAR interval estimates

- Version 0.1: Emphasis at this stage on reproducibility
openWAR

- R package to be submitted to CRAN
- Currently available for download on GitHub
  https://github.com/beanumber/openWAR
- Scrapes XML files from MLBAM GameDay server
- Processes using XSLT and compiles detailed play-by-play info into a data frame
- Computes \textit{openWAR}
- Diagnostic and visualization tools
- Paper (currently under review)
  http://arxiv.org/abs/1312.7158
Conservation of Runs

- $\rho(baseCode, outs)$: expected number of runs scored in remainder of inning, from the state $(baseCode, outs)$
- Empirically estimate $\hat{\rho}$
- **Conservation of Runs:**
  - Every run gained by the offense is a run lost by the defense
- $\delta_i$: Change in expected runs occurring on the $i^{th}$ play:
  \[
  \delta_i = \rho(b_{i+1}, o_{i+1}) - \rho(b_i, o_i) + runsOnPlay_i
  \]
Sample Play

- 5/08/2013: 2 outs, Nick Markakis on 2B, Adam Jones on 1B
- Matt Wieters doubles to right center and both runs score
  \[ \hat{\delta}_i = \hat{\rho}(2, 2) - \hat{\rho}(3, 2) + 2 = 0.31 - 0.41 + 2 = 1.90 \]

https://cvmdo.bamnetworks.com/mlbam/2013/05/08/347228/coaching_video/cv_26934817_4500K.mp4

- How to allocate responsibility among the offensive and defensive players?
openWAR accounting

- $\delta = 1.90$ runs
- $\delta_{br} = 0.32$ runs, after controlling for ballpark and platoon advantage
  - The runner on first (Jones) gets 91% of the baserunning credit
  - The runner on second (Markakis) gets 9% of the baserunning credit
- $\delta_{bat} = 1.58$ runs goes to the batter (Wieters)
  - Remains 1.58 runs after controlling for the fact that Wieters is a catcher
- $\delta_{field} = -0.70$ runs (37% of the blame) go to the fielders
  - 68% of that blame ($-0.47$ runs) goes to the CF
  - 32% of that blame ($-0.22$ runs) goes to the RF
  - Negligible amounts go to the other fielders
- $\delta_{pitch} = -1.20$ runs (63% of the blame) goes to the pitcher
Cumulative Fielding Model
Defining Replacement Level

- Scarcity: Only $30 \cdot 25 = 750$ roster spots
  - Take the 750 players who played the most
  - All other players are by definition “replacements”

- Replacement players have an average RAA per plate appearance

- Each player is assigned a replacement-level shadow based on their playing time (shown in gray in next slide)
Defining Replacement Level - openWAR 2012

Number of Players = 1284, Number of Replacement Level Players = 534
Playing Time (plate appearances plus batters faced)

openWAR Runs Above Average

Total RAA = 0
Total WAR = 1166.3

MLB Player Replacement Player

Baumer (Smith) openWAR
Defining Replacement Level - rWAR 2012

Total RAA = 298.5
Total WAR = 1008.7

Playing Time, 2012 (plate appearances plus batters faced)
Defining Replacement Level - openWAR 2012 normalized

Number of Players = 1284, Number of Replacement Level Players = 603

Playing Time (plate appearances plus batters faced)

openWAR Runs Above Average

Total RAA = 0
Total WAR = 1005.2

MLB Player Replacement Player

Baumer (Smith) openWAR
Modeling Uncertainty

- Both $r\text{WAR}$ and $f\text{WAR}$ are published as point estimates, not interval estimates.
- $\text{openWAR}$ models player sampling error.
- Recall: each player at each position is assigned an $\text{RAA}$ value for each play.
- Idea: resample these values many times!
- Quantitative assessment of uncertainty enables nuanced conclusions.
Cabrera vs. Trout, 2012

Results

Cabrera better 31.6 %

Trout better 68.4 %
## Uncertainty (Sampling)

<table>
<thead>
<tr>
<th>Name</th>
<th>q0</th>
<th>q2.5</th>
<th>q25</th>
<th>q50</th>
<th>q75</th>
<th>q97.5</th>
<th>q100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trout</td>
<td>3.96</td>
<td>5.90</td>
<td>7.66</td>
<td>8.58</td>
<td>9.50</td>
<td>11.26</td>
<td>13.34</td>
</tr>
<tr>
<td>Cano</td>
<td>2.43</td>
<td>4.77</td>
<td>6.82</td>
<td>7.94</td>
<td>9.06</td>
<td>11.17</td>
<td>13.85</td>
</tr>
<tr>
<td>Headley</td>
<td>2.53</td>
<td>4.56</td>
<td>6.39</td>
<td>7.46</td>
<td>8.47</td>
<td>10.42</td>
<td>13.00</td>
</tr>
<tr>
<td>Encarnacion</td>
<td>2.45</td>
<td>4.47</td>
<td>6.33</td>
<td>7.32</td>
<td>8.27</td>
<td>10.09</td>
<td>13.08</td>
</tr>
<tr>
<td>McCutchen, A</td>
<td>2.33</td>
<td>4.41</td>
<td>6.25</td>
<td>7.27</td>
<td>8.26</td>
<td>10.27</td>
<td>12.21</td>
</tr>
<tr>
<td>Votto</td>
<td>2.68</td>
<td>4.80</td>
<td>6.23</td>
<td>7.01</td>
<td>7.81</td>
<td>9.28</td>
<td>11.40</td>
</tr>
<tr>
<td>Fielder</td>
<td>2.11</td>
<td>4.12</td>
<td>5.99</td>
<td>6.98</td>
<td>7.95</td>
<td>9.82</td>
<td>12.22</td>
</tr>
<tr>
<td>Posey</td>
<td>2.09</td>
<td>4.14</td>
<td>5.80</td>
<td>6.75</td>
<td>7.67</td>
<td>9.61</td>
<td>11.68</td>
</tr>
<tr>
<td>Mauer</td>
<td>2.63</td>
<td>4.30</td>
<td>5.88</td>
<td>6.74</td>
<td>7.60</td>
<td>9.27</td>
<td>10.91</td>
</tr>
</tbody>
</table>

**Table**: Interval estimates for 2012 openWAR Leaders: 3500 simulations
Limitations

- Data integrity
- Stolen bases and wild pitches not properly accounted
- Cannot distribute data with the R package
- Can’t distinguish between batted ball trajectories or speeds
- Defense measures only range – not sure-handedness, throwing, etc.
THANK YOU!!