Sabermetric Analysis: Wins-Above-Replacement

Karl Hendela

Sabermetric Analysis: Wins-Above-Replacement

Karl Hendela
Seton Hall University

Abstract

Baseball has increasingly turned to analytics in order to evaluate players and teams within the league. The latest statistic to become popular is Wins-Above-Replacement, abbreviated WAR. WAR takes into account all factors of a player's contribution to a team, including pitching, batting, base-running, and fielding, and combines them to produce a single number which approximates how many games the team has won by utilizing this player as compared to a replacement-level or minor league level player at the same position. The WAR statistic has become widely used as a way to compare players, even players playing different positions or in different time periods. However, current versions of WAR all have one serious drawback: there is no publicly available, standardized formula or process to calculate it. Instead, separate organizations have formulated their own versions of the statistic using proprietary data, rendering the statistic both difficult to understand and difficult to replicate. In this project we create a standardized WAR statistic that accurately indicates the value of a player using only publicly available data, and show that our new WAR statistic strongly correlates with WAR statistics which use proprietary data. This approach has the potential to create a standard WAR formula that an average baseball fan can use and understand in the future as they evaluate a baseball player's value. We call it SHU-WAR.

1. Introduction

The game of baseball has been America's pastime for well over a century. The objective of the game—score more runs than you allow and you win—has led to the development of two axioms, "A ball player's purpose in playing baseball is to do those things which create wins for his team, while avoiding those things which create losses for his team," and "Wins result from runs scored while losses result from runs allowed" (James, 1984). Predicting which team will win a baseball game and measuring how players contribute to those wins has become an American pastime in itself with sabermetrics, the search for objective knowledge about baseball (James, 1984). Recording players' statistics have been an integral part of Major League Baseball since the foundation of the game. It started with the simplest statistics: number of hits and walks; batting average; RBIs (runs batted in) and home runs for position players; innings pitched, strikeouts, bases on balls (walks), earned run average (ERA), and win/loss record for pitchers. Statistics such as these have been carefully compiled since the dawn of the game for managers and fans alike to look at and analyze themselves.

While no statistics can tell the full story of the players and the games they played, over time the question has arisen: Is it possible to have a single statistic that measures the quality of a player? This question has led to the development of statistics such as runs-created, win-shares, and most recently wins-above-replacement (WAR). Wins-above-replacement attempts to answer this question by taking into account all aspects of a player's performance and translating it into the most important part of the game: winning. The idea is if you had a team composed of replacement-level players, statistically they would be expected to finish a typical 162-game season with a record of roughly 48 wins and 114 losses. The WAR statistic determines how many wins above or below 48 would result by replacing one of those replacement players with the player whose WAR is being measured (WAR Explained, 2019). In the creation of this number, the unique factors of the game are normalized to be context, league, and park neutral so that every player is valued on an even playing field. Baseball is one of the few sports that has significant differences between fields and leagues as the sport has never standardized outfield or foul territory dimensions. This makes it difficult to compare players, as these differences cause problems in determining how one player would fare on a typical field. WAR takes this all into consideration, as adjustments are made to even the playing field between players, making this a statistic comparable across all leagues, fields, and time periods.

Normally, this is the part of the paper where I would describe how WAR is calculated. Unfortunately, however, this is impossible because WAR is a non-standardized metric: there are in fact multiple versions of the WAR statistic. To compound matters, all of these WAR statistics, published by companies like Baseball Reference, FanGraphs and Baseball Prospectus, are based at least in part on proprietary data not available to the public. Some of this data requires specialized technology and, while it is always nice to gain new angles on parts of the game, this can result in confusion about just what aspects are being measured. There is also little public information available about these statistics, which leads to confusion on how to actually calculate the final value, as it may be constantly adjusted behind closed doors. The typical baseball analyst therefore is left to take the values generated from these methods for granted since there is no way to reproduce the numbers these organizations create themselves. Baseball Reference has named its version bWAR, FanGraphs is fWAR and Baseball Prospectus is WARP, none of which are identical or uniform across all players. When the statistic is typically talked about though, it is referred to as WAR even though there are multiple versions of it.

Given the amount of publicly available baseball statistics, however, and the various mathematical ideas and techniques already available in sabermetrics, it seems reasonable to believe that a WAR statistic can be constructed that would be practical, understandable, and accessible to the general baseball-loving fan. Our objective in this paper is to create that statistic, a new WAR statistic we call SHU-WAR, which can function as a standardized wins-above replacement statistic while avoiding using any proprietary data that is not available on a typical baseball statistical database.

2. The Linear Weights System

The basis of the SHU-WAR statistic is the Linear Weights System developed by John Thorn and Pete Palmer (Thorn & Palmer, 1985). In this system various baseball events were independently evaluated and given a value that expresses each event in terms of runs, specifically runs produced or prevented. The events evaluated in Thorn and Palmer's linear weights system encompass all physical plays, including even non-scoring plays. Examples of batting events, for example, include events like singles, doubles, triples but also non-hit events like walks, bases on balls, and when a player is hit by a pitch (and thus advances to first base automatically). Using a regression analysis, Thorn and Palmer determined the linear weight of each event, as well as adjustments taking into account the league that the player is in. Hitless at-bats are given a negative score, so that the sum of all linear weights of all events within a league sum up to zero, and so the walks, singles, and so on can be measured by the number of runs they produce above an average at-bat. To give an example, while a home run in a baseball game can directly score between one to four runs (depending on how many people were on base) it was determined through Thorn & Palmer's work that a home run produces approximately 1.40 runs greater than the average at-bat (Thorn & Palmer, 1985).

For pitching statistics, the linear weights system is calibrated to the league's earned-run-average (ERA) and the amount of innings pitched. This means a pitcher with the league average ERA would have a presumptive record of 81-81 if they pitched every game of the 162 game season, while any deviation from the league average will either increase or decrease the number of wins for the season (Thorn & Palmer, 1985). This is parallel to batting where the system takes into account positive and negative events in terms of the league averages to determine how many runs the player produced. The same type of analysis was done for base stealing, fielding, and the various other events that are tracked by other baseball statistics. These linear weight system values obtained by Thorn & Palmer help weight the various baseball events that are measured in our calculation of the SHU-WAR statistic.

3. The Formulas

Our SHU-WAR statistic has six components, respectively covering batting, base stealing, fielding, positional adjustment, pitching, and a replacement player adjustment. While the formulas that were used to calculate SHU-WAR are all weighted using the linear weights system developed by Thorn and Palmer, a number of adjustments are also made in order to further strengthen the performance of the final value. Once each of these six components is calculated, the runs credited to the player for each component are summed up and divided by two times the average runs per game per team within the league. This gives the final value of how many wins are accredited to the player above that of a replacement player. An early version of this method for calculating what we now call SHU-WAR first appears in the book Understanding Sabermetrics (Costa, Huber & Saccoman, 2nd ed.).

3.1 Batting Runs

Batting runs (BR) is a statistic that takes the important portions of the offensive player's game and translates each event into runs created. The statistic used is as follows:

$$\begin{aligned} BR = {} & (0.47 \times H) + (0.38 \times 2B) + (0.55 \times 3B) \\ & + (0.93 \times HR) + [0.33 \times (BB + HBP)] \\ & - [ABF * (AB - H)] \end{aligned}$$

The BR formula for an individual player is calculated using the total number of at-bats a player records (AB), as well as the various positive events that can result: hits (H), doubles (2B), triples (3B), home runs (HR), base on balls (BB), or hit by pitch (HBP). A final adjustment is made using a factor based on the league batting average (ABF), which we will describe shortly. The ABF calculation uses only league totals and is multiplied by the number of outs, which is at-bats minus hits.

In order to simplify the formula above, a player's singles, that is hits that only advance the batter to first base, are not distinguished from hits. Since all doubles, triples and home runs are also hits, the weights given each type of hit gives information about the hit quality. A single, for example, is weighted the same as a hit at 0.47, but doubles, triples, and home runs are given additional weight values because these contribute more to the production of runs. To illustrate with an earlier example a home run in this case is given a weight of 1.40, just as by Thorn and Palmer, since it accumulates the coefficient of 0.47 for being a hit and 0.93 for being a home run, and $0.47 + 0.93 = 1.40$. This formula also accounts for walks (BB) and hits by pitch (HBP) with a coefficient of 0.33. These are events that allow a player to get on base, but are weighted less than a hit because a typical baseball statistic database does not differentiate between intentional and unintentional walks. Lastly, a portion is subtracted to take into account the ABF, a league batting term factor. The ABF is calculated using only league total statistics, and is

$$ABF = \frac{\begin{array}{c}[(0.47\times H)+(0.38\times 2B)+(0.55\times 3B)\\ +(0.93\times HR)+(0.33\times(BB+HBP))]\end{array}}{AB-(LGF\times H)}$$

This ABF statistic uses the positive portion of BR as its numerator, but with league inputs now, and scales it by league at-bats minus the league factor times hits. The league factor (LGF) is simply 1 for both National league and American league, and is adjusted when calculating for other leagues. Two examples include the historical Union Association League having an LGF of 0.8 and the Federal League having a 0.9. An LGF of 1 simply makes the denominator the number of outs, but allows for adjustment depending on the league (Costa, Huber, & Saccoman, Understanding Sabermetrics, 2008).

A pitcher's ABF value in their BR value is divided by two, but for pitchers who have played a position other than pitcher, or have been the designated hitter for multiple games, their ABF value is not divided by two. A modern day case is the player Shohei Ohtani who in 2018 was a pitcher for the Los Angeles Angels, but who is known for his hitting ability and was often used as a designated hitter. This means in the calculation of his batting runs, his ABF was not divided by two.

To complete the batting portion of this WAR statistic, there needs to be a park factor adjustment. Ballparks like Coors Field are notoriously known as hitter's parks since the thin air in Denver makes it easier for the ball to travel. Some of the longest home runs in history have been hit there for this reason, so this needs to be considered. The park factor is calculated as

$$PF = \frac{\frac{ARS_H+ARA_H}{GP_H}}{\frac{ARS_A+ARA_A}{GP_A}}$$

and is used to adjust the batting runs statistic as

$$BR\,PF\,Adj = \begin{cases} BR+\frac{BR\times(1-PF)}{2} & \text{if } BR\geq 0\\ BR+\frac{BR\times(PF-1)}{2} & \text{if } BR<0\end{cases}$$

The park factor takes up to a five year average of the amount of runs scored at home (ARS_H) and adds the amount of runs allowed at home (ARA_H), which is all divided by the amount of games played at home (GP_H). This is divided by the same fraction except calculated now in terms of games played away from home. In this way a number greater than one indicates an easier stadium to score runs in as compared to the games played away. If five years' data does not exist, such as teams having recently changed stadiums, then the BR PF Adj is calculated using up to a five year average for these statistics.

3.2 Base Stealing

Base-running is an underestimated portion of the game, with faster players able to create a competitive advantage on the base paths by advancing on the bases without hit support, simply based on their speed. A stolen base is beneficial to a team, but being caught stealing often ruins a team's chances of scoring in that inning. The linear weights system gives that a stolen base is considered to create approximately 0.3 runs per stolen base, and being caught stealing subtracts approximately 0.6 runs per instance. This leaves us with the simple formula of:

$$SBR = 0.3\times SB - 0.6\times CS$$

3.3 Fielding

The next portion of SHU-WAR covers the aspect of fielding, typically the most difficult aspect of the game to measure because each position and situation is unique in baseball. Calculations here will differ somewhat by position, with some of the weights for various events altered depending on its difficulty and importance to the position. To be explicit, catchers, pitchers, and first basemen will have unique formulas, while second basemen share the same formula with third basemen and shortstops, and left, right, and center outfielders will have the same formula as each other.

A players fielding runs (FR) calculation is done in two stages. First a general fielding run calculation is done for each position on a player's team. The individual player's FR value is obtained by then taking these values and making adjustments based on the number of fielding outs that the player obtained at that position.

First we give the fielding run calculations for each position on the team. The formula for pitcher fielding runs (P FR) is as follows:

$$\begin{aligned} P\,FR = {} & (0.1 \times (PO + (2 \times A) - E + DP)) \\ & - P\,APL \times (Total\ PO - Total\ SO) \end{aligned}$$

where PO is put-outs, A assists, E errors committed, DP double plays, and SO strikeouts. Again, all the inputs for this calculation are team statistics for the pitcher position. The P APL factor is a league factor, analogous to the ABF factor used in the batting runs BR calculation. Like the ABF factor previously, league value are used for the inputs.

$$P\,APL = \frac{(0.1 \times (PO + (2 \times A) - E + DP))}{Total\ PO - Total\ SO}$$

In the APL calculation, the inputs for the numerator are position-specific (for example, E counts the total number of errors committed by league pitchers) while the denominator uses the total putouts and strikeouts for the entire league across all positions. The APL values for other positions, which appear below, have a similar form and are calculated in the same way.

The corresponding formulas for catchers is

$$\begin{aligned} C\,FR = {} & (0.2 \times (PO^* + (2 \times A) - E + DP)) \\ & - C\,APL \times (Total\ PO - Total\ SO) \end{aligned}$$

$$C\,APL = \frac{(0.2 \times (PO^* + (2 \times A) - E + DP))}{Total\ PO - Total\ SO}$$

There is an adjustment to putouts (PO*) for catchers, which is why there is an asterisk in the formula. Catchers are credited with putouts for most strikeouts, so PO* are the putouts for catchers minus the putouts from strikeouts.

For first basemen, second basemen and outfielders we use respectively

$$\begin{aligned} 1B\,FR = {} & (0.2 \times ((2 \times A) - E)) \\ & - 1B\,APL \times (Total\ PO - Total\ SO) \end{aligned}$$

$$1B\,APL = \frac{(0.2 \times ((2 \times A) - E))}{Total\ PO - Total\ SO}$$

$$\begin{aligned} 2B\,FR = {} & (0.2 \times (PO + (2 \times A) - E + DP)) \\ & - 2B\,APL \times (Total\ PO - Total\ SO) \end{aligned}$$

$$2B\,APL = \frac{(0.2 \times (PO + (2 \times A) - E + DP))}{Total\ PO - Total\ SO}$$

$$\begin{aligned} LF\ FR = {} & (0.2 \times (PO + (4 \times A) - E - (2 \times DP))) \\ & - LF\ APL \times (Total\ PO - Total\ SO) \end{aligned}$$

$$LF\ APL = \frac{(0.2 \times (PO + (4 \times A) - E + (2 \times DP)))}{Total\ PO - Total\ SO}$$

As mentioned, the previous FR formulas are team values for each position, which are next used to calculate each individual player's actual fielding runs. The final calculation to determine an individual player's fielding runs FR value is as follows:

$$FR = \left(\frac{Player\ Putouts}{Total\ Team\ Putouts} \right) \times FRpos \qquad (1)$$

In the formula above, the team positional value FRpos (for example, P FR for a pitcher) is weighted by how many putouts the individual player has at the position compared to the total putouts for the team at the same position. The figure in parentheses, which is the amount of putouts the player has at a position compared to the total for the team at the position, represents the percentage of the team's fielding runs for that position that the player is awarded. This is then multiplied by the team's fielding runs for the position.

Over the course of a season a player may play more than one position. In this case the final FR calculation is weighted according to games played at each position. For example, if a player played 100 games total, with 90 games as a catcher and 10 games as a third baseman, then their final FR score would be 90% of their FR C score from equation (1) plus 10% of their FR 2B score from (1). It is possible a player has more games played at positions than total games played, as it is possible to play multiple positions within a single game, but this is a rare occurrence and thus there would only be a negligible difference. The current simplicity of the calculation makes it more manageable to calculate with available data.

3.4 Positional Adjustment

Certain positions are clearly more difficult or more involved than others, a factor that is not taken into account in the fielding runs formula. So SHU-WAR also introduces a positional adjustment factor. Each position is assigned a coefficient to account for this factor, with higher numbers indicating more difficult positions and lower and negative numbers easier positions. The coefficients are weighted by the amount of games a player has played at the position compared to the total games in the season. The formula and the corresponding coefficients used in this version of WAR is as follows:

$$Pos\ Adj = \left(\frac{GP\ at\ Pos}{TotGP}\right) \times PosAdjCoeff$$

where GP at Pos stands for games played at a particular position, TotGP stands for total games played, and the positional adjustment coefficient PosAdjCoeff for the various positions is given in the following table.

C	1B	2B	3B	SS
12.5	-12.5	2.5	2.5	7.5

LF	CF	RF	DH	P
-7.5	2.5	-7.5	-17.5	5

Once again, as with the fielding runs FR calculation, a player may play more than one position in a season so the final positional adjustment is the weighted sum of each position for the respective player.

3.5 Pitching

As important as creating runs is, preventing runs is just as important in winning a game. Thorn and Palmer's original formula—take the innings pitched, multiply it by the number of games played times the league ERA, divide by nine and subtract out the player's earned runs—was league-specific and created before the advent of interleague play in Major League Baseball in 1997. The formula used in our calculations is very similar but with an interleague adjustment and an adjustment for park factor:

$$PR = \frac{IP}{9} \times \left(\left(\frac{GP - TotIG}{GP}\right) \times LeagueERA + \left(\frac{TotIG}{GP}\right) \times MLB\ ERA \right) - (Player\ ER)$$

where IP stands for innings pitched by the player, GP for games played, TotIG for total number of interleague games played, LeagueERA is the earned run average for the player's league (American league or National league), and MLB ERA stands for the earned run average for both leagues combined.

Similar to batting runs, pitching is inversely affected by the park so there is a park factor adjustment because any park that is easier to hit in, is therefore more difficult to pitch in.

$$PR\ PF\ Adj = \begin{cases} PR + \frac{PR \times (PF-1)}{2} & \text{if } PR \geq 0 \\ PR + \frac{PR \times (1-PF)}{2} & \text{if } PR < 0 \end{cases}$$

It also needs to be noted that positive values are better than negative values where the final value is in terms of runs prevented instead of runs allowed. A negative value indicates the pitcher allowed more runs per game than the average pitcher while a positive value indicates that the pitcher prevented that amount of runs above the average pitcher.

3.6 Replacement Player Adjustment

The last component of SHU-WAR is the replacement player adjustment which takes into account the amount the player actually played. Here we make a distinction between position players and pitchers, where the pitchers' replacement player adjustment is based on batters faced (BF) and position players' adjustment is based on plate appearances (PA).

$$(P)\ Replacement\ Player\ Adjustment = \frac{BF}{50}$$

$$(Pos)\ Replacement\ Player\ Adjustment = \frac{PA}{30}$$

This aspect of SHU-WAR works as a longevity measure where a player is considerably more valuable to a team if they play more. This is in comparison to that of a replacement player who would not be expected to play as much as a quality starter. It also should be noted that the National League currently has pitchers batting, but a pitcher's replacement player adjustment is based purely on batters faced. This distinction is made if a player has multiple games as a pitcher, as their replacement player adjustment is made using batters faced divided by 50; otherwise the positional player adjustment is used with plate appearances divided by 30. This accounts for cases like Shohei Ohtani in 2018 where his replacement player adjustment is based on batters faced as a pitcher.

3.7 SHU-WAR

The final calculation of SHU-WAR is the summation of each of the preceding six components, scaled by two times the runs per game average for the league. For reference the final formula for SHU-WAR is

$$\frac{(BR\ PF\ Adj + SBR + FR + Pos\ Adj + PR\ PF\ Adj + Repl\ Adj)}{2 \times RPG}$$

where RPG is the average runs per game for the league.

Each of the six components, and the RPG value, can be constructed using freely available box score data, so SHU-WAR can be calculated for any player from any league using these formulas, and the SHU-WAR value can be broken down into the six components so good and bad aspects of a player's SHU-WAR value can be directly attributed to specific parts of it. Two people can calculate the wins-above-replacement for any player and come up with the same value using these formulas. There's undoubtedly room to further perfect the SHU-WAR statistic, which can be the study for future research, but as we will show in the rest of the paper in this current state the formula already works well.

Before we compare SHU-WAR to other WAR statistics, though, we briefly recap. In constructing SHU-WAR, Thorn and Palmer's formulas were primarily used, but a few adjustments were made in an attempt to reinforce the quality of the numbers produced. Four primary adjustments were made to different sections of the formulas that closed the gap between our initial SHU-WAR values and those of proprietary WAR statistics like, for example, Baseball Reference's bWAR. Changes to batting and pitching included the modifications to Thorn & Palmer's one year park factor. Using only a one year park factor presents a problem for those years that are above or below average, where more or less runs may be scored at the home park than normal. Since 2010, Baseball Reference has reportedly used a three-year park factor while FanGraphs has used a five year regressed park factor (WAR Comparison Chart, 2019). We therefore added the adjustment of using up to a five-year park factor adjustment to have a greater sense of the true impact the park has on both batting and pitching. Another adjustment made was the pitcher's ABF value was divided by two. Pitchers' batting seemed to have a major negative impact, specifically for National League pitchers on their final SHU-WAR value, so an adjustment was made to the ABF value by dividing it by two. The ABF is the comparison value to the average hitter in the league and since pitchers are typically not the best hitters, their batting statistics are much worse than the average major league hitter. Also, pitchers' batting performances are often sacrificed in terms of winning a game: pitchers are often pinch-hit for if an important situation comes up in the later innings of a game, as pitchers are regularly expected to either get out quickly in a game or try for a sacrifice play such as bunting a player over to the next base. A pitcher's main purpose is to pitch, so their typical negative batting runs values were adjusted to more accurately represent their actual impact on the game.

Two other adjustments made to the SHU-WAR statistic were adjustments to pitching runs with the introduction of interleague play, and the decision to use a 5.0 coefficient for the pitcher's positional adjustment. Interleague play was introduced into Major League Baseball in 1997 where teams from the American League began playing a few series with teams from the National League. Since pitching runs was originally conceived as a league-specific statistic only dealing with the league's ERA, it was important to take into account the approximately twenty interleague games (depending on the season) that each team played. For any season in which there were no interleague games, the original formula from Thorn and Palmer is simply used as the numerator for interleague games would be zero. The last modification to the list of formulas is the positional adjustment coefficient for pitchers. Initially there was no available coefficient for a pitcher's fielding so it was determined that the difficulty of the position fielding-wise was approximately 5.0. This is between shortstop and second base which respectively have coefficients of 7.5 and 2.5, where infield positions are often more involved and are generally more difficult than other positions. Pitchers have to deal with bunts and hard ground balls within a short distance, so it was deemed at least approximately fair to give a 5.0 positional coefficient to pitchers. Other positions over time have had their coefficient values adjusted as more information is gathered, so this may be necessary for this value as well, but with the current information available a 5.0 seems appropriate.

4. Comparison to Proprietary WAR Statistics

In order to justify the accuracy of our SHU-WAR statistic, it is important to compare our calculations to some of the large selection of proprietary WAR statistics available, such as Baseball Reference's bWAR. The bWAR statistic is one of the most popularly used as it has been developed with advanced technology for a long time, but it uses metrics that are unavailable to the public. For example, in their batting portion of WAR, they use the statistic wOBA (weighted on-base-average). This value takes into account exclusive data, like the fact that from 2003 on, they can discern between infield and outfield singles, which have a 0.06 run difference (WAR Explained, 2019). Also, Baseball Reference primarily uses Defensive Runs Saved (DRS) as their defensive metric in calculating bWAR. In the creation of DRS, they measure the player's fielding range, the catcher's throwing ability, the success of good fielding plays, and the failure of misplays in different situations. Stats for missing the cutoff-man in a relay throw and measurements for a catcher's ability to frame and block are unavailable to the public, but they are all used in Baseball Reference's calculations behind the scenes (Position Player WAR Calculations and Details, 2019). These are just a few examples of the statistics used in formulating their final numbers where there are many pages of information about all of the adjustments and calculations used in order to finally reach a player's wins above replacement.

One of our secondary goals in developing SHU-WAR was to test the hypothesis that many the advanced measures used by organizations like Baseball Reference and FanGraphs that seem useful may in fact be much beyond what is necessary to accurately evaluate a player. A statistic too complicated to state its actual calculations, may be a statistic too complicated to be useful in an accurate analysis. This study therefore challenges the need for these advanced calculations used in proprietary WAR statistics by questioning if there is a considerable difference between that of a statistic that can be understood and calculated by the public to that of an advanced metric from a major organization.

5. Results

To test SHU-WAR versus the proprietary WAR statistics, we calculated SHU-WAR for 234 players from different teams and different eras and compared the results to Baseball Reference's final bWAR values for these players. WAR in general is a non-standardized statistic but if the final results are close to that of a major organization, one using advanced metrics along with proprietary data, then it shows that WAR could become standardized. Competing organizations have minimal agreement on using any of the same methods in their calculations, as the sabermetric community makes it almost a game in itself to come up with better ways to determine the value of each player. However simplicity is often underrated, and we believe it is likely possible the value of a player should be able to be determined with the vast amount of data already available online.

The time periods that were studied ranged from 1927 to 2018 in order to demonstrate that SHU-WAR is consistent across different eras. The full list of teams that were studied are as follows: the 2010 New York Yankees, the 2010 Colorado Rockies, 2018 Los Angeles Angels, the 1968 Detroit Tigers, the 1930 Philadelphia Phillies, and the 1927 New York Yankees. Two teams from the 2010 season were looked at, as the most available information about Baseball Reference's bWAR statistic is available from 2010. The 2010 season was also representative of an average MLB season, in terms of historical norms for runs scored per game. Each team was purposefully chosen to cover different aspects that are accounted for in SHU-WAR; for example, the Colorado Rockies are notoriously known for having a hitter's park. Coors Field is in Denver Colorado where the air is much thinner than the average ballpark so the ball is known to travel further, so it was important to see how a large park adjustment impacted the final values. Along the same lines, the 1930 Phillies team was chosen since 1930 was known as "the year of the hitter," while the championship-winning 1968 Detroit Tigers played in "the year of the pitcher." The 2018 Los Angeles Angels were chosen see how arguably the best player in the present day, Mike Trout, was valued along with the special player Shohei Ohtani, who is the rare pitcher that is also a quality batter. Lastly, the 1927 New York Yankees are famous for the home run battle between Babe Ruth and Lou Gehrig, and we thought it would be valuable to see how legends in the past are valued compared to legends in the making, such as Mike Trout.

We calculated SHU-WAR for all players from these six teams, with positional players (POS) separated categorically from pitchers (P). Pearson correlation values between SHU-WAR and Baseball Reference's bWAR values were then calculated for each team and POS and P segment. The results showed that for each team and segment there was greater than a 0.90 Pearson correlation, a remarkably high correlation value (see Table 1). Moreover, a number of segments had correlation values greater than 0.98, including the pitchers for the 2010 New York Yankees, the pitchers for the 1968 Detroit Tigers, the position players for the 2018 Los Angeles Angels, and the position players for the 1927 New York Yankees. This indicates a strong consistency between SHU-WAR values and Baseball Reference's bWAR values across po-

Year/Team	POS Correlation	P Correlation
2010 NYY	0.91	0.98
2010 COL	0.94	0.94
2018 LAA	0.98	0.92
1968 DET	0.95	0.99
1930 PHI	0.95	0.93
1927 NYY	0.99	0.92

Table 1. Correlations between SHU-WAR and bWAR statistics for position players (POS) and pitchers (P) for the teams considered.

sitions and time periods for entire teams.

Full teams showed significant correlations to that of Baseball Reference's bWAR values, but SHU-WAR is a statistic for individual players. That leads to the question, were there any major differences between a player's SHU-WAR and bWAR? A chart was made comparing the paired differences bWAR minus SHU-WAR and bWAR minus fWAR as a control. That chart appears as Figure 1, with the *x*-axis ordered by bWAR minus SHU-WAR value. When comparing the WAR statistics, we chose to focus on differences of 2 or more, since 2 wins-above-replacement can significantly impact an evaluation of a player; for example, someone with a 0 WAR is considered replacement level, while someone with a 2 WAR is generally considered a good player and players having WAR values above 4 are considered all-stars (Slowinski, 2010). We found that, of the 234 players calculated, 13 differed by more than plus or minus two wins when comparing SHU-WAR and bWAR values, while a total 9 players were outside of this margin between bWAR and fWAR, a fairly small discrepancy. Examples of players valued differently by SHU-WAR and bWAR included Brett Gardner of the 2010 New York Yankees and Phil Collins of the 1930 Philadelphia Phillies. A closer look at both WAR calculations reveals that Brett Gardner's discrepancy can be explained as a discrepancy in defensive value. Advanced calculations in the defensive portion of the bWAR significantly benefited his final value, while his defense in the SHU-WAR calculations did not reward him as much. Phil Collins of the 1930 Philadelphia Phillies was the only player to appear on both the SHU-WAR versus bWAR list and the fWAR versus bWAR list of players that were outside the plus or minus two margin. Collins was calculated to have a SHU-WAR value of 2.9, a bWAR value of 5.5, and an fWAR value of 3.3. This shows that the true value of Phil Collins is generally not known, other than that he was an above average player in 1930. The only WAR value here that can be justified is the SHU-WAR value though, since it is possible to see the calculations done to obtain this statistic. In general, almost all 234 values are within the same reasonable range of valuing this subset of players.

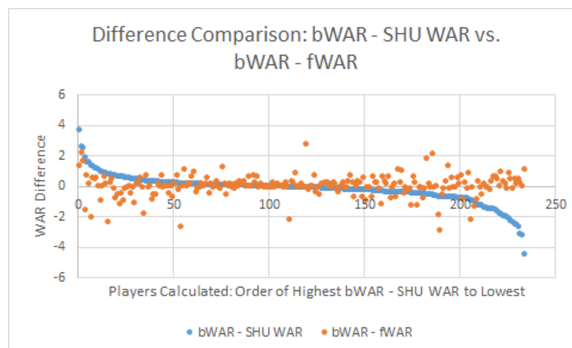

Figure 1. Differences in WAR statistics.

A second set of graphs, shown in Figures 2 and 2, display the same information in a different way. In these graphs values off the $x = y$ line indicate differences of final SHU-WAR values. The final result showed a R^2 value of 0.8801 displaying a strong correlation between SHU-WAR and bWAR. In comparison, there was a R^2 value of 0.9063 between bWAR and fWAR which are two established statistics that use proprietary measures. There is slightly larger variability between SHU-WAR and bWAR, but the difference between fWAR and bWAR clearly shows that there is still inconsistencies between two statistics from major organizations that use advanced technology attempting to measure the same thing. The differences between bWAR and fWAR also remind us that the goal of a standardized WAR statistic is not to produce the exact same values as one of these other wins-above-replacement statistics, but to measure the same thing in a comparable way. And at least for SHU-WAR, a transparent and replicable way.

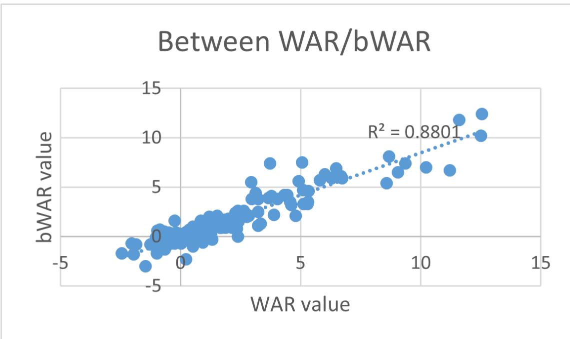

Figure 2. SHU-WAR versus bWAR values.

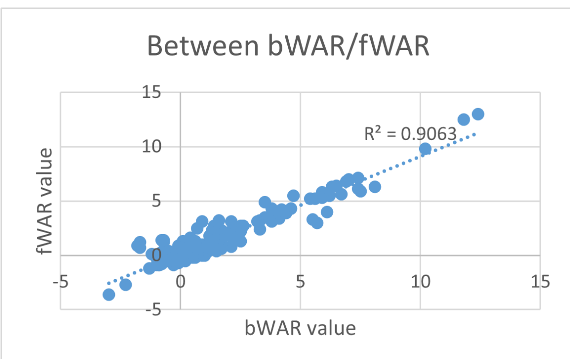

Figure 3. bWAR versus fWAR values.

Lastly, a comparison between SHU-WAR and fWAR was made by looking at data from the 2010 season of FanGraphs. The FanGraphs fWAR statistic for all players in the 2010 season saw 35% of all players fall between WAR values of -1 and 0, 38% of players fall between 0 and 1, 19% between 1 and 4, and 6% above 4 (Slowinski, 2010). Since SHU-WAR values were measured for players from different eras, and for individual teams and not full leagues, it was more important to get a sense of the general shape of the WAR statistic. The replacement level player is indicated as having a 0 WAR, so the majority of players should be

Range	Percent of Players
Less Than -2	0.9%
$-2 \leq x \leq 0$	30.8%
$0 \leq x \leq 2$	44.4%
$2 \leq x \leq 4$	11.5%
$4 \leq x \leq 6$	6.0%
$6 \leq x \leq 8$	2.6%
$8 \leq x \leq 10$	1.7%
Greater Than 10	2.1%

Table 2. Percent of players studied in various SHU-WAR ranges.

above this value. Of the 234 players whose SHU-WAR values were calculated, 68.3% of players were greater than 0 with 12.4% of players over 4. This last number is a fairly large value, but is understandable given the teams we elected to calculate, which included a number of elite players like Babe Ruth, Lou Gehrig, Mike Trout, and Robinson Cano, and also primarily had an above average number of quality players on each team. The scale for SHU-WAR therefore is similar to that of Baseball Reference and Fangraph's scale, where below-average to average players are between -2 and 2, quality players are between 2 and 4, and great players up to hall of famers are above 4.

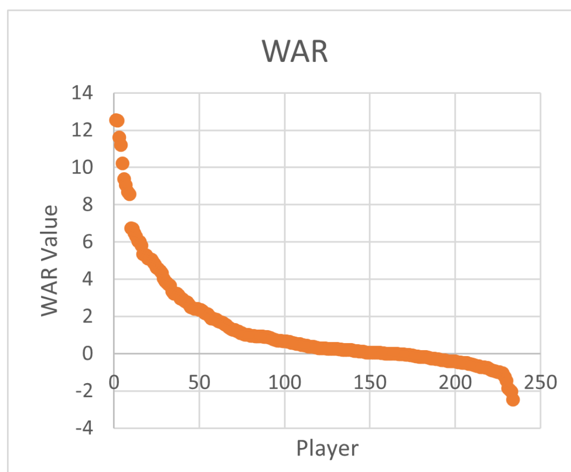

Figure 4. Graph of all 234 SHU-WAR values.

6. Conclusion

Baseball statistics are only as useful as the ability of the user to understand what they are saying. With advances in technology and data gathering methods, there is a strong temptation to say "more is always better" and to try to use all of the data possible to its greatest extent. However, more is not always better, and in sabermetrics there is a certain point where what a number is trying to say can get lost, and one of the ways this can happen is when the value is constructed using techniques beyond anyone's understanding. Wins-above-replacement is one place where the battle of who can use the best and most advanced measures has become a major discussion, to the point currently that the principal reason for the statistic is lost in all of its calculations. Simplifying, standardizing, and making transparent the calculation offers a great benefit to baseball fans and analysts alike, who can then use and appreciate the statistic while staying on the same page.

In the current state of the game, wins-above-replacement is often referred only as WAR, which often obscures the fact that the source of these values, and the values themselves, can vary. As we have seen there is clearly a difference between major organizational WAR values, where the final number of wins for an individual player can even differ by even greater than 2 WAR. This fact is remarkable, especially given the stature of the WAR statistic: as noted by FanGraphs, "we think teams are paying about $8 million per every WAR they add to their roster" (Weinberg, 2016). Take a player like Esmil Rogers in the 2010 season for the Colorado Rockies: Rogers had a -1.7 bWAR value according to Baseball Reference but a 1.2 fWAR value according to FanGraphs. Both of these statistics are often referred to simply just "WAR", so Esmil Rogers would likely be a huge advocate of using FanGraph's WAR over Baseball Reference's WAR, since it would mean he should be worth more than $16,000,000 more. A natural question, of course, is why can there be such a big difference in WAR values? The answer is, there is no way to tell. The calculations are not public information and these numbers must simply be taken at face value for what they are since you cannot determine how these numbers are created.

These principles—simplicity, replicability, and transparency—are why the standardization of the statistic is necessary and why we set about to create a WAR statistic that adhered to them. The steps given in this paper for calculating SHU-WAR are at the very least a proof of concept, and a starting point for a better WAR statistic. Esmil Rogers' SHU-WAR value can be calculated using the methods of this paper to be -0.99, and this calculation can be replicated by anyone using this set of formulas here. This value falls right between the bWAR value of -1.7 and the FanGraphs value of 1.2, but the SHU-WAR value comes with the additional information, for example, that Rogers' low WAR is due to his poor pitching runs value of -15.06 for the 2010 season. All this is done without using any proprietary data that the average person does not have access to. Baseball has always been a game of numbers that fans and analysts have loved to use, and our hope is that with SHU-WAR it may finally be the case that WAR becomes a standardized statistic that everyone can appreciate and understand for themselves.

References

Costa, G. B., Huber, M. R., & Saccoman, J. T. (2008). *Practicing Sabermetrics*. Jefferson: McFarland.

Costa, G. B., Huber, M. R., & Saccoman, J. T. (2008). *Understanding Sabermetrics*. Jefferson: McFarland.

Costa, G. B., Huber, M. R., & Saccoman, J. T. (2012). *Reasoning with Sabermetrics*. Jefferson: McFarland.

James, B. (1984). *The Bill James Baseball Abstract*. New York: Ballantine.

Position Player WAR Calculations and Details. (2019, May 10). Retrieved from Baseball Reference: https://www.baseball-reference.com/about/war_explained_position.shtml

Slowinski, S. (2010, February 15). *What is WAR?* Retrieved from FanGraphs: https://library.fangraphs.com/misc/war/

Thorn, J., & Palmer, P. (1985). *The Hidden Game of Baseball*. Chicago: The University of Chicago Press.

WAR Comparison Chart. (2019, May 10). Retrieved from Baseball Reference: https://www.baseball-reference.com/about/war_explained_comparison.shtml

WAR Explained. (2019, May 6). Retrieved from Baseball Reference: https://www.baseball-reference.com/about/war_explained.shtml

Weinberg, N. (2016, January 14). *Basic Principles of Free Agent Contract Evaluation*. Retrieved from FanGraphs: https://library.fangraphs.com/basic-principles-of-free-agent-contract-evaluation/