

Gus Schmidt, Isaiah Holquist

CMSC 12200

1/28/2020

## **Creating a Wins Above Replacement Metric for Soccer**

### **Introduction**

Sports statistics have existed as long as sports themselves. However, since the mid-20th Century, baseball has perhaps been the sport most impacted by the analysis of these statistics. Because baseball is such a discrete game, where every at-bat, hit, and pitch can be parsed and broken down into clear data points, it naturally lends itself to quantitative analysis. However, since the Society for American Baseball Research was founded in 1971, sports analytics has exploded in popularity, use, and research. While today, virtually every sport has developed its own set of metrics and quantifiers to measure how well certain players perform, baseball unsurprisingly presents one of the sports most robust metrics: wins above replacement (WAR). WAR attempts to measure a given player's value across all facets of the game by determining how many more wins he's worth than a replacement-level player at his same position. A player with a +1.0 WAR is expected to grant a team 1 more win over the course of a season than the league average replacement player of his position. Likewise, a player with a -1.0 WAR is expected to cost a team 1 win over the course of a season. While WAR is far from a perfect metric, it is frequently referenced and used by both fans and professionals as a quick way of summarizing the value that an individual player provides to their team.

Soccer is undeniably the most popular sport in the world, but it is one of the least quantified. The game lacks discrete plays that can be easily analyzed, there are infrequent substitutions for the purpose of comparison, teams often play across different leagues, players in different positions have entirely different purposes, and there are relatively few scoring opportunities to break down. Today, there is no universally understood metric to explicitly quantify the contribution of every player on the field. Our project will be to develop a WAR-like

measurement for soccer, using data from the English Premier League. We will develop an algorithm for WAR that is specific to each position: goalkeeper, defender, midfielder, and attacker.

## **Data Sources**

Since we are looking at developing a player specific measure, we will need player specific data. Luckily, there are several sources that have key metrics (goals, assists, tackles, passes, clean sheets, fouls committed, cards, etc.) broken down by player and year:

- <https://fbref.com/en/comps/9/stats/Premier-League-Stats>
- <https://www.premierleague.com/stats>
- <https://www.whoscored.com/Regions/252/Tournaments/2/Seasons/7811/Stages/17590/PlayerStatistics/England-Premier-League-2019-2020>
- <https://fantasyfootball.telegraph.co.uk/premier-league/statscentre/>

## **Timeline**

Weeks 4-6: Scrape and clean all relevant data

- Scrape all data from websites/sources
- Clean all the data and fix any issues
- Sort data into a database to easily query individual data points on every player

Week 7: Develop procedure and calculate WAR

- Design specific algorithm to calculate WAR for each position
- Calculate and assign WAR values to players based on these calculations

Weeks 8-10: Build front end display and presentation of findings

- Build a site that displays best teams by WAR every year/lets users pick and build teams
- Create final project presentation