

# **The influence of pass blocks, shot blocks, and tackle success on game outcome in the English Premier League.**

Karim A. Aliyev

College of Natural Sciences, Physics Math & Astronomy Building, 2515 Speedway, Austin, TX 78712

The University of Texas at Austin

Correspondance email: ka29688@utexas.edu

## **Abstract**

Elite soccer clubs and fans have historically prioritized players who excel at scoring goals and making assists while paying less attention to their defensive abilities. However, in soccer, all players are required to participate in defense regardless of their position. The question that arises is whether it is important to consider the players' defending skills when evaluating their potential contribution to the team. In this study, we approached this question by investigating whether teams' defensive strength is connected to their victories. The study was conducted on matches played in the English Premier League (EPL), an annual competition that involves twenty elite soccer teams. Here, we quantitatively measured defensive strength using three variables: the proportion of pass blocks, the proportion of shot blocks, and the proportion of tackles won which are the building blocks of any team's defensive tactics. We then analyzed how these variables correlate to match outcomes on a game-to-game basis for 2280 EPL matches played in the past 6 years. Based on this data, winning teams tend to perform better in all three variables, but the magnitude of this difference is quite small, making it hard to predict game outcomes using solely these variables. With more advanced data, the study can be improved to include more defensive variables with the isolation of key defensive moments in games. Using this, the correlation between the team's defensive preparedness to wins can be further evaluated in the EPL and other major soccer leagues.

**Keywords:** soccer defense, soccer player transfers, English Premier League (EPL), shot blocks, pass blocks, tackles

## **Introduction**

Soccer is one of the most popular sports in the world, with some games reaching over a billion live viewers. Just as in any sport, statistical data is calculated every game, with which researchers can analyze teams' tactics and maneuvers. This, in turn, allows them to identify patterns that explain why some teams might be more successful than others (Farias, 2020). The most important rule in a game of soccer is to score more goals than your opponent to win. Unsurprisingly, scoring goals or creating goal-scoring opportunities tends to be more valued than defending, as shown by trends in the transfer market, where players are traded between clubs. It was found that there is a significant positive relationship between a player's ability to score goals and create assists and their transfer price (Carmichael et al., 1999, Barbuscak, 2018). Furthermore, when predicting a player's transfer price, the public also values goal scoring as an important factor for the player's value (Herm et al., 2013). As can be seen, both the club management and the fans give substantial credit to goal-scorers and their contribution to the team.

Despite this, it has been shown that goal-scoring moments, although crucial to the outcome of the game, represent only a small fraction of the team's full dynamics; hence, it is important to focus on other predictors of the teams' success or failure during a particular game or competition (Ricardo et al. 2013). Existing studies

have attempted to identify factors that contribute most to consistent victories. For example, Lima et al. (2021) analyzed the connection of ball possession, passes, and shots on goal to winning matches. Ball possession is the amount of time a team controls the ball during the game. Contrary to popular belief, it was found that teams with less ball possession tend to win more often while shots on target and pass accuracy did not have a significant influence on the game outcome (Lima et al., 2021). There have also been some studies on what actions create the most successful defense (Vogelbein et al., 2014, Forcher et al., 2022). It was shown that increased defensive pressure, defined as the amount of players pressuring the attacker, increases the chances of successful defense overall (Forcher et al., 2022). Another study concluded that quick possession regains also positively contribute to a stronger defensive execution (Vogelbein et al., 2014).

However, there is only a small number of works that attempt to correlate a team's defensive strength to its ability to win games. This topic was touched on by Vogelbein et. al (2014) when studying ball recovery speeds in the 2011 season of the German Bundesliga. Ball recovery speed refers to the time it takes for a team to regain ball possession and is considered a defensive variable. Using 308 games, the authors found that the teams that recover the ball faster typically win more games than others. A study involving more defensive variables conducted by Yudibascara et al. (2023) found that Argentina's victory in the 2023 World Cup was largely due to the team's superior defense tactics every match. However, due to the World Cup being a seven-game-long competition, the size of the data was relatively limited. At this moment, there have been no similar studies conducted for large-scale competitions and we aimed to fill this apparent gap in soccer analytics.

We wanted to find out if there exists a correlation between a team's defensive strength and its ability to seal victories in games. We chose to conduct our study in the English Premier League (EPL), one of the biggest soccer competitions in the world. Twenty elite English clubs compete annually for the championship title playing a total of 380 games each season. This gave us a substantial amount of match data to analyze and make conclusions. For the EPL, we hypothesized that the team that is stronger defensively than its opponent will also be more likely to win the game. To quantify defensive strength, we used the following three variables: the proportion of tackles won, the proportion of shots blocked, and the proportion of passes blocked. These parameters are normally performed by all players on the pitch, regardless of position, and thus demonstrate the defensive strength and coordination of the entire team. These variables are also extensively tracked by game-play technology which allows us to collect substantial data to reach more robust results. Significant evidence in favor of our hypothesis can suggest that defensive preparedness is a crucial contributor to the teams' win rate. This, in turn, can push for prioritization of a player's defensive abilities when evaluating their transfer price.

## **Methods**

### **Data Collection & Cleaning**

We first needed to gather data for matches played in the past 6 seasons of the EPL which occurred in the years 2017 to 2023. A total of 380 matches are played each season, allowing us to study a total of 2280 matches. For each game, extensive statistical data is collected by game-play technology and then reported by public sources online. The particular source we used for match data was Sports Reference LLC which has a strong reputation in reporting many major sports data (Sports Reference, 2023).

We used the 'worldfootballR' package in R (Zivkovic, 2022) to access the match report of each of the 2280 games. For this research, we scraped the "Defense", "Passing", and "Shooting" subsections creating the respective "defense", "passing", and "shooting" datasets. Next, we used the 'tidyverse' package in R (Wickham et. al., 2019) to remove reported match statistics that went beyond the scope of this study. The remaining match data was used to calculate variables for our final data set (Supplementary Data File 1):

- (1) **team** - the name of the team to which the rest of the variables correspond.
- (2) **outcome** - a categorical variable indicating whether the team won (“W”), drew (“D”), or lost (“L”) a particular match. Since two teams play in every match, there are two outcomes recorded per match. Hence, the first row in our data set indicates the outcome for the home team and the second row indicates the outcome for the away team that played in that same match. This process was repeated for subsequent rows until all outcomes for the 2280 matches were represented. To perform this calculation, the columns, home\_team, away\_team, home\_score, and away\_score in the “defense” data set were used
- (3) **shots\_blocked** - the number of shots blocked divided by the total number of shots attempted by the opposition. A shot is considered blocked when a defender stands in the way of the ball after the opposing player attempts to shoot it at goal. The number of blocked shots was in the “defense” dataset and it was divided by the total number of shots in the “shooting” dataset to get the final proportion.
- (4) **passes\_blocked** - the number of passes blocked and intercepted divided by the total number of passes attempted by the opposition. A pass is considered blocked when a defender either stands in the way of the ball or steals the ball as it is traveling from one opponent to another. The number of blocked passes was also in the “defense” dataset and it was divided by the total number of shots in the “passing” dataset to get the final proportion.
- (5) **tackle\_success** - the ratio of tackles won divided by the total number of attempted tackles. A tackle happens when a defender attempts to steal the ball from the opposing player and they could either be successful or not.

## Data Analysis

We first visually investigated whether a team's defensive performance differs based on the outcome of the game. Specifically, we wanted to see if the proportions of passes blocked, shots blocked, and/or tackles won by the team differ on average based on whether they won, lost, or drew. We represented our data in boxplots using the ‘tidyverse’ R package (Wickham et. al., 2019).

Next, only Win and Loss outcomes were considered. We wanted to see if the defense variable proportions for winning teams are significantly larger than those for the losing team. In the sample of 2280 games, 1760 games resulted in one team winning and the other team losing. These games were used to conduct a paired test for the difference in proportions where the pairs were the winners and losers of the game. As mentioned previously, every two consecutive rows represent teams that played in the same game; therefore, game outcomes are not independent of each other. For example, a victory for one team in a game will automatically imply a loss for the other team in the same game. Due to this dependence, we chose to conduct a paired test for significance in proportion differences. Each pair was made up of the proportion of pass blocks/shot blocks/tackle success for the winning team and the losing team from the same game. A total of three tests were conducted, one for each defensive variable:

- A. Test for significance in the difference between the proportion of shots blocked by the winning team and the proportion of shots blocked by the losing team.
- B. Test for significance in the difference between the proportion of passes blocked by the winning team and the proportion of passes blocked by the losing team.
- C. Test for significance in the difference between the proportion of tackles won made by the winning team and the proportion of tackles won made by the losing team

For each test, the following null and alternative hypotheses were tested with the rejection region of 5%.

- Null Hypothesis,  $H_0$ : the difference in proportions is 0
- Alternative Hypothesis,  $H_a$ : the difference in proportions is positive

Before testing, it was necessary to ensure that the distribution of differences was normal for each variable. The Shapiro-Wilk Normality Test was used to determine that the difference in the proportion of shots blocked and passes blocked was not normal while the difference in the proportion of tackles won was normal (Table 1). Hence, we used the paired Wilcoxon test for the proportions of passes and shots blocked and the paired t-test for the proportion of tackles won.

## Results

Our findings show that the proportions of passes blocked, shots blocked, and tackle success change little with game outcome. For all three variables, the median proportion associated with draws and wins was practically the same and the median proportion associated with losses was only slightly smaller. Moreover, the IQRs of associated boxplots look almost identical and the outliers follow a similar pattern (Figure 1). Therefore, regardless of whether the team wins, draws, or loses any given match, we can expect approximately the same proportions for associated defense variables as their opponent.

When performing our hypothesis tests for the proportion differences, it was found that the winning team, on average, had a larger proportion of pass blocks, shot blocks, and tackles won than the losing team. However, these proportion differences, although statistically significant, are less than 0.02 on average (Table 2).

## Discussion

Teams who have been victorious in EPL in the years 2017 to 2023 tended to slightly outperform their opponents in proportions of pass blocks, shot blocks, and tackles (Table 2). These findings support our overarching hypothesis that defensive performance positively contributes to game outcomes. Despite this, the difference in proportions for all three variables was very small (Table 2) and their distributions fall in similar ranges regardless of whether the teams won, drew, or lost (Figure 1).

Our conclusions differ from other authors who studied the contribution of defense to game outcomes in soccer. For example, Ruiz et al. (2017) have used defensive statistics to explain the rise and downfall of a soccer club called Leicester FC. The authors evaluated how successful Leicester was at stopping different types of opponent passes and shots and found out that the club was successful at stopping not all, but certain key passes and shots that gave the team the edge over their opponents in games (Ruiz et. al. 2017). In our study, pass blocks and shot blocks were not categorized by type due to our data source lacking such specifics which could have created the difference in conclusion. While Ruiz et. al. devised multiple complex categories to classify shots and passes, even simple categorization has the potential to affect conclusions about defensive success. To consider, in a game of soccer, teams typically make hundreds of passes but not every pass helps advance the opponent's offense (Fernandez-Navarro et. al. 2016). Similarly, some shots taken by the opponent are not even aimed at the goal and whether they were blocked or not matters little in these cases. Therefore, it might be more insightful to focus on truly influential passes and shots and evaluate how successful the defense is at blocking them. Better defensive play in such crucial moments can probably be a better predictor of game outcome.

In another study, Yudibascara et. al. (2023) evaluated the connection between Argentina's defensive performance and victories in the 2023 World Cup. The authors considered pass blocks and tackling success like in this study but with the inclusion of defensive pressure, which calculates the number of defenders attempting to steal the ball from the opposition. In the study of Argentina's team, defensive pressure proved to be an important predictor of the game outcome (Yudibascara et. al., 2023). There was not sufficient data in our study to factor in defensive pressure, but according to Forcher et al. (2022), its addition could potentially lead to a larger influence of defensive performance on game outcome. It is important to note that both Ruiz et al. (2017)

and Yudibascara et. al. (2023) evaluated the defensive performance and wins of one club. Because each soccer team has a unique style of play, the correlation between defensive performance and wins could potentially vary between clubs. Our study has shown that, in the EPL, the proportions of pass blocks, shot blocks, and tackle success do not largely vary based on outcome. But it is possible that for certain clubs the defensive variables correlate highly with the level of their success in the league.

We must finally mention that a game of soccer involves many intricacies that even complete match data may not always capture. This typically makes it difficult to identify connections between the team's performance and the outcome of the game (Machado et. al., 2010, Casanova et. al., 2012). We believe that certain information may be further lost when select variables are considered which could have potentially happened in this study. Moving forward, including defensive pressure like Yudibascara et. al. (2023) and isolating key defensive moments like Ruiz et al. (2017) will be essential to further understand the relationship between defense and game outcome. It will also be beneficial to conduct this study on other major soccer leagues in Europe like the German Bundesliga, Italian Serie A, French League 1, Spanish La Liga, and others. The leagues are very similar in structure to the EPL so seeing how defensive performance relates to game wins in these competitions will allow us to make more robust conclusions.

## **Acknowledgments**

We would like to express our sincere gratitude to Dr. Julia Clarke and Dr. Lucas Legendre for their invaluable guidance and support throughout the entire process of this research. Their expertise and mentorship played a crucial role in shaping the content of this paper. We are particularly thankful for their assistance in comprehending and implementing the statistical methods employed in this study. Their commitment to fostering academic excellence has been instrumental in the successful completion of this research.

## References

- Barbuscak, Lukas (2018). What Makes a Soccer Player Expensive? Analyzing the Transfer Activity of the Richest Soccer. Augsburg Honors Review: Vol. 11, Article 5.  
[https://idun.augsburg.edu/honors\\_review/vol11/iss1/5](https://idun.augsburg.edu/honors_review/vol11/iss1/5)
- Carmichael, F., Forrest, D., & Simmons, R. (1999). The labour market in association football: who gets transferred and for how much? Bulletin of Economic Research, 51(2), 125–150.  
<https://doi.org/10.1111/1467-8586.00075>.
- Casanova, F., Garganta, J., Oliveira, J., (2012). Representativeness of offensive scenarios to evaluate perceptual-cognitive expertise of soccer players. The Open Sports Sciences Journal, 5, 161–166.  
<http://dx.doi.org/10.2174/1875399X01205010161>
- Farias, V. M., Fernandes, W. B., Bergmann, G. G., & dos Santos Pinheiro, E. (2020). Relationship between ball possession and match outcome in UEFA Champions League. Motricidade, 16(4), 1-7.  
<https://doi.org/10.6063/motricidade.18382>
- Sports Reference LLC (2023). Football Reference, Premier League Stats:  
<https://fbref.com/en/comps/9/Premier-League-Stats> (Accessed October 11, 2023)
- Fernandez-Navarro, J., Fradua, L., Zubillaga, A., Ford, P. R., & McRobert, A. P. (2016). Attacking and defensive styles of play in soccer: analysis of Spanish and English elite teams. Journal of Sports Sciences, 34(24), 2195–2204.  
<https://doi.org/10.1080/02640414.2016.1169309>
- Forcher, L., Altmann, S., Forcher, L., Jekauc, D., & Kempe, M. (2022). The use of player tracking data to analyze defensive play in professional soccer - A scoping review. International Journal of Sports Science & Coaching, 17(6), 1567-1592.  
<https://doi.org/10.1177/17479541221075734>
- Herm, S., Callsen-Bracker, H.-M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: accuracy and evaluation attributes of an online community. Sport Management Review, 17(4), 484–492.  
<https://doi.org/10.1016/j.smr.2013.12.006>
- Machado, J. C., Barreira, D., Teoldo, I., Travassos, B., Júnior, J. B., Santos, J. O. L. D., & Scaglia, A. J. (2019). How does the adjustment of training task difficulty level influence tactical behavior in soccer? Research quarterly for exercise and sport, 90(3), 403-416.  
<https://doi.org/10.1080/02701367.2019.1612511>
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.  
<https://cran.r-project.org/>
- Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., & Richardson, M. J. (2013). Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. Human movement science, 32(4), 555-566.  
<https://doi.org/10.1016/j.humov.2013.01.011>
- Rocha-Lima, E. . M. ., Tertuliano, I. W., & Fischer, C. N. . (2021). The influence of ball possession, passes and shots on target in winning premier league football matches. Research, Society and Development, 10(8).  
<https://doi.org/10.33448/rsd-v10i8.17824>
- RStudio Team (2023). RStudio: Integrated Development for R. Posit, PBC, Boston, MA.  
<https://posit.co/products/open-source/rstudio/>
- Vogelbein, M., Nopp, S., & Hökelmann, A. (2014). Defensive transition in soccer—are prompt possession

regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. Journal of sports sciences, 32(11).

<https://doi.org/10.1080/02640414.2013.879671>

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” Journal of Open Source Software, 4(43), 1686.

<https://doi.org/10.21105/joss.01686>

Yudibaskara, G. P. A., Ria, L., Tomoliyus, & Perdana, G. (2023). Decoding Argentina’s path to World Cup glory: Unveiling the winning factors in match result analysis. Journal Sport Area, 8(2), 217-229.

[https://doi.org/10.25299/sportarea.2023.vol8\(2\).12759](https://doi.org/10.25299/sportarea.2023.vol8(2).12759)

Zivkovic J (2022). worldfootballR: Extract and Clean World Football (Soccer) Data. R package version 0.6.2.

<https://CRAN.R-project.org/package=worldfootballR>.

## Tables:

**Table 1. Results of the test for normality of differences in proportions of defensive variables.** For each game that ended in a Win for one team (hence a Loss for the other), the defensive variable proportions were calculated for the winning team and the losing team. The total number of games in the sample was 1760. The Shapiro-Wilk Test was used to find out if, for each variable, the paired differences were normal. It was found that differences in passes and shots blocked are not normally distributed but the differences in tackles won are normally distributed.

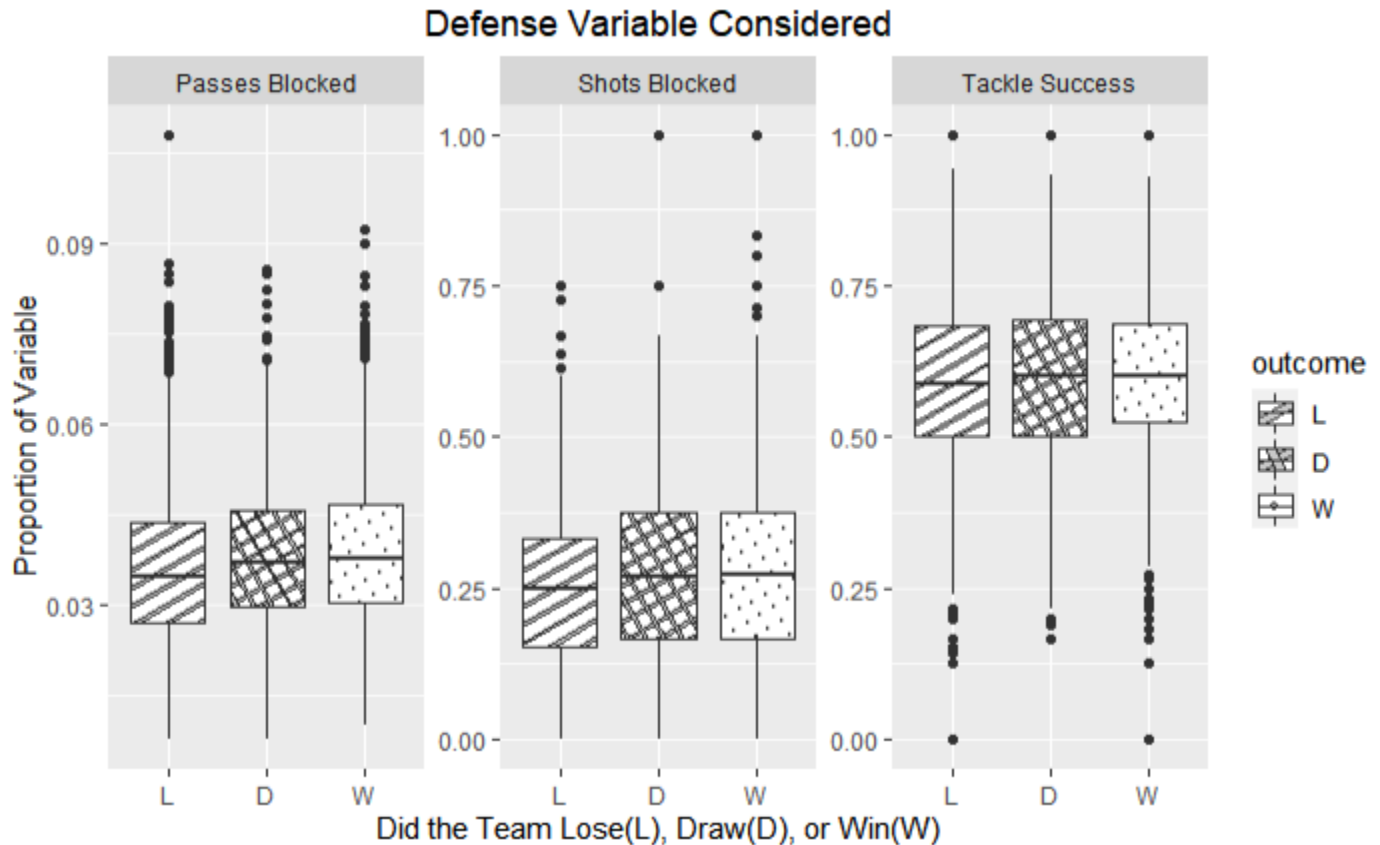
Defense Variable	p-value	Conclusion
passes blocked	0.001306	Not Normal
shots blocked	0.006402	Not Normal
successful tackles	0.1731	Normal

**Table 2. Result for statistical test for the difference in defensive variables and game outcome.**  $p_1$  stands for the defensive variable proportion of the team that won (W) and  $p_2$  stands for the defensive variable proportion of the team that lost (L). The test was conducted on 1760 games that ended up in one team winning and another one losing. In all three cases, the null ( $H_0$ ) was rejected at the significance level of  $p < 0.05$  in favor of the alternative ( $H_a$ ) indicating that teams that win tend to have higher defensive variable proportions than the losing teams.

Defense Variable	$H_0$	$H_a$	Test	mean difference	p-value	Conclusion
passes blocked	$p_1 = p_2$	$p_1 > p_2$	Wilcoxon	2.55E-03	5.76E-05	$p_1 > p_2$
shots blocked	$p_1 = p_2$	$p_1 > p_2$	Wilcoxon	2.10E-02	2.21E-10	$p_1 > p_2$
successful tackles	$p_1 = p_2$	$p_1 > p_2$	t-test	0.0121247	0.002485	$p_1 > p_2$



## Figures:



**Figure 1. Distribution of proportions of each defensive variable for Wins (W), Draws (D), and Losses (L).**

The dataset evaluated consists of 4560 teams. 1760 teams won (W), 1760 lost (L), and 1040 drew (D). The figure is divided into three sections, each representing one of the studied defense variables: pass blocks, shot blocks, and tackle success. For each variable, the three boxplots represent the proportion distribution of that variable for the teams that won (W), drew (D), or lost (L). In each section, it can be observed that the medians are similar between W, D, and L, and the IQR almost completely overlap. This suggests that regardless of the outcome of the game, the proportions of defensive variables do not tend to vary.