Analyzing the Significance of Defensive Stats on the Quality of Defending

Ansh Singhal

1. Introduction

In football, it is often very difficult to isolate the contributions made by an individual due to the interdependence of various on-pitch factors. This applies to defensive contributions as well. Sir Alex Ferguson (The most decorated manager in the history of football) once said, "*Attack wins you games, defence wins you titles*". A good defensive line-up can often be the determining factor in a team's success, and there is thus a lot of scrutiny about what a defender brings to a team on an individual level. However, when it comes to defenders, a lot of the basic statistics that are brought to the table are often misleading. I thus wanted to analyze the direct relation between some of the defensive statistics that people quote in debates involving the comparison between defenders, and the number of goals a team concedes.

Listed below are the three superficial statistics that are a direct measure of a defender's actions and are often brought up while comparing defenders. These three measures are really the three main ways in which a defender impedes the opponent team.

Interceptions[3] - an interception impedes a successful pass - therefore Interceptions are counted as a specific defensive activity that impedes a successful pass.

Tackles[3] - this is a defensive activity that strips the ball from an opponent.

Blocks[3] - this is a defensive activity that blocks a shot taken by a player on the opposing team.

Often the arguments made in support of or against the quality of defending are baseless, and it is futile to compare defenders based on defensive efforts when the true value of these defensive contributions is not fully understood. The results of this paper will be significant in the following ways. For one, they will give an empirical relation between the defensive stats mentioned and the defensive success of a team. This will allow for a more direct measure of both a team's defensive quality based on defensive stats, and how much a defender is truly contributing to the team's defensive success. If it is found that certain defensive contributions have a tight positive relation with the team's defensive success (reducing the number of goals conceded), then those defensive contributions, and the success percentage in those defensive actions should be valued higher while comparing defenders. This will have implications when it comes to analyzing defenders during talent scouting, deciding transfers in defensive positions and analyzing defender performances. Additionally, a better insight into defensive contributions will allow for tactically maximizing opportunities for positive defensive contributions and minimizing the need for defensive contributions that lead to more goals.

I am going to attempt to answer this question by cumulating defensive contributions for defenders from a team over a season and measuring their average defensive contributions against the team's defensive record at the end of the season. In Michael Stöckl's paper[1] on – Using a Graph Convolutional Network to Understand Defensive Performance in Soccer – he hypothesizes, 'the art of good defending is to prevent something from happening before it has

even happened'. He further questions the usage of mere 'tackles and interceptions', in any argument pertaining to defensive performance. His paper postulates that proactive actions made by defenders work towards making an attack more predictable and summed up his model's analysis with a toolkit to analyze a defense's influence (on both a team and individual level) on limiting an attack's threat. I would like to dig further into the claim that basic defensive contributions such as 'tackles, blocks and interceptions' don't give a full picture and look at how they affect defensive success. In an American Soccer Analysis[3] by Drew Olsen, he looks at defensive efforts in the attacking third of the pitch (*Fig 1*) and concludes a strong correlation between defensive efforts in the final third and the win% of a team. I thus also want to look at the comparison between significance of tackles depending on the zone they're made in.

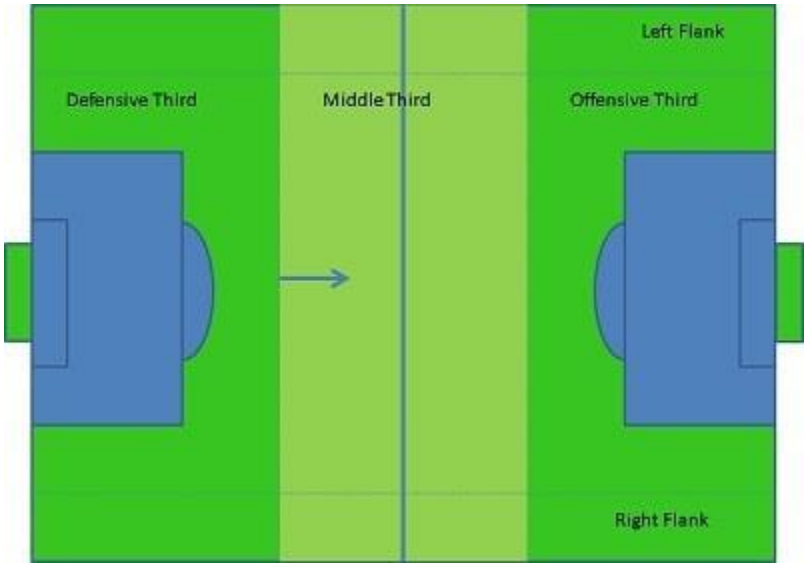

Fig 1: *Dividing a soccer pitch by defensive, midfield and attacking zones*

In this paper, I will look to compare tackles in each third of the pitch. Additionally, I want to analyze the significance of tackles, blocks and interceptions. Showing a lack of significance or a negative impact to defensive performance between some defensive contributions will further Michael Stöckl's claim that those statistics don't give a full picture, and will exemplify the need for model's based on player tracking and xPected (xG – expected goals, xA – expected assists, etc.) Metrics[2] (Metrics based on models trained to score player actions based on the situation on the pitch, *see appendix*).

2. Data

I got player data from the top 5 soccer leagues (English, French, Italian, German, Spanish leagues) by season from a dataset on *Kaggle* (https://www.kaggle.com/datasets/kriegsmaschine/soccer-players-values-and-their-statistics), for the 2018-19 and 2019-20 seasons (most players had two entries, one from their 2018-19 season, and the other from their 2019-20 season). Next, I filtered the data to only include players classified as defenders. I then aggregated all the defensive contributions made by defenders on a **team-season level** (eg. Real Madrid 2018-19 and Real Madrid 2019-20 represent two different

entries) and divided that figure by the number of games each team played over the respective season. Thus, the aggregate statistics for a team, say Real Madrid 2018-19, represents the per game average for all Real Madrid defenders over the 2018-19 season, and that represents one row. Real Madrid 2019-20 will represent a separate row of data.

To measure defensive performance, I am using the number of goals conceded per game by a team over a season. I got this variable by merging the aggregate team data with team data from the two seasons in question using the standings tables provided by *Fbref* (https://fbref.com/en/comps/Big5/2018-2019/2018-2019-Big-5-European-Leagues-Stats).

The variables I am using include average tackles per game, average blocks per game and average interceptions per game, *(see appendix)* made by a team defense over a season. In addition to these variables, I am also using the percentage of tackles by zone (defensive, midfield and attacking thirds) for a team's defenders over a season. I will attempt to study the correlation between the average per game defensive contributions and the number of goals a team conceded per 90 minutes (1 game of football) over a season. Additionally, I will also look at the impact of the distribution of a team's tackles by zone on the number of goals conceded.

Table 1 shows the descriptive statistics of each variable, over the 184 team-season observations I was able to extract from the top 5 leagues over 2 seasons. As seen, on average a team's defense made 16.08 tackles per game over a season. The split of tackles into each third adds up to the mean value in the first row, and as seen, defenders make more tackles in their own defensive third than the midfield and attacking thirds. On average, half of the tackles a team's defense makes over a season occur in their own defensive third.

Teams can employ vastly different tactical setups and differ on a wide spectrum when it comes to the quality of players, thus there are large variations between the minimum and maximum values of these per game defensive contributions, as well as the percentage of tackles made in the defensive third. While Mainz 05's defense attempted only 9.53 tackles per game in 2018-19, Metz' defense attempted a staggering 22.96 tackles per game in the 2019-20 season! A team like 'Bayern Munich', which controls large amount of possession, and maintains a very high defensive line, only attempted 35.5% of tackles in their defensive third during the 2018-19, and only intercepted 4.88 of their opponents' passes in the 2019-20 season. On the other hand, 'Newcastle United' were forced to make 61.3% of their tackles in their own third in the 2019-20 season and intercepted 10.9 opponent passes per game in the 2018-19 season.

Team Season Averages per90	Mean	Median	St Deviation	Min	Max
tackles	16.08	16.13	2.49	9.53	22.96
pct_tackles_def_3rd	0.50	0.50	0.05	0.36	0.61
pct_tackles_mid_3rd	0.38	0.39	0.03	0.30	0.47
pct_tackles_att_3rd	0.12	0.12	0.02	0.07	0.20
blocks	14.62	14.47	2.28	9.26	19.85
interceptions	7.32	7.26	1.15	4.88	10.90
goals_per_90	1.37	1.34	0.34	0.58	2.24

Table 1: Descriptive Statistics

3. Empirical Results

I estimate three different specifications. The dependent variable in each specification is defensive performance, as measured by the number of goals scored against a team per game over a season. Tackles per game, percentage of tackles made in the defensive third, blocks and interceptions per game, for a team's defense over a season are the independent variables. Table 2 shows the results.

In the first specification, I regress goals conceded over the tackles attempted by a team per game. No correlation was observed with just tackles being compared.

Dependent Variable: goals_per_90 (Conceded by team over a season)			
	1	**2**	**3**
Intercept	1.31 (0.16)***	-0.18 (0.27)	-0.25 (0.26)
tackles	0.00 (0.01)	0.00 (0.01)	-0.04 (0.01)**
pct_tackles_def_3rd		3.17 (0.48)***	2.51 (0.48)***
blocks			0.06 (0.01)***
interceptions			0.02 (0.02)
R-squared	0.34	0.31	0.29
Adjusted R-squared	0.00	0.18	0.28
Number of observations is 182			
Standard errors are in parentheses			
*** significant at 0.1%, ** significant at 1%, * significant at 5%			

Table 2: Regression Results

In the second specification, I regressed the goals per 90 against the tackles per game, and percentage of tackles in the defensive third for each team in a season. The percentage of tackles made in the defensive third is very tightly related to the goals conceded. And while the coefficient itself doesn't hold all that much meaning, since the percentage of tackles is a ratio and is bound by one, it is clear that a higher percentage of tackles in the defensive third leads to more goals being conceded.

In the third specification, the coefficient for interceptions is statistically insignificant, while the coefficient for blocks, tackles and ratio of defensive third tackles is significant. This means that holding the number of tackles and interceptions made constant, an increase of 1 in the average number of blocks per game for a team over a season is associated with an additional 0.06 goals conceded per 90 by a team. On the other hand, increasing a team's tackles per game by 1 results in 0.04 less goals conceded per 90.

In general, the interceptions made by a defender seem to be insignificant when it comes to explaining a team's defensive performance, however the number of blocks they are forced to

make is very tightly related to the number of goals conceded. On the other hand, the zone by which average per90 tackles are measured, is very key in determining defensive performance.

4. Conclusion

Based on these results, I would advocate to discredit the value of blocks and interceptions as a measure of a defender's defensive capabilities. One would even argue that, whether by virtue of a team's system or an individual's defensive style, limiting the number of blocks needed to be made is a better determinant of good defending. The results indicate that a more proactive style of defending, involving regaining possession through tackles further up field, is an indicator of goal limiting defensive contributions. Now it is not the case that making blocks is inherently a negative action, in fact they are in most cases defensive actions that impede a shot on goal being conceded. However, defenders that play in teams that sit back and soak up pressure are more likely to get opportunities to block shots made by opponents, simply because of facing more shots. There is simply too much reverse causality when it comes to the kind of system a team plays, and the amount of possession they maintain that influences the number of opportunities for blocks and interceptions a defender can make. It is clear however, that as a team, keeping the game away from their own defensive third should be a key objective in strategizing a good defense. Defenders who can act proactively by committing tackles, especially further up the field provide great value in that sense.

The statistical insignificance of interceptions, and the strong negative correlation between blocks and defensive performance confirm Michael Stöckl[1]'s claim that these statistics do not give a full picture and are too dependent on factors such as a team's quality in other positions, their possession and tactical set-up. Further, the strong relation between the zone in which a defensive contribution is made and defensive performance asserts that tactical set-ups in modern football should strive for maintaining a high line that allows for more tackles in the midfield third than the defensive third. Thus, more complex models, such as the one in Stöckl's paper, are a better indicator of defensive performance, based on defensive presence, positioning and limiting an opposition's threat[1].

The regressions I ran do not consider the tactical set up of a team, which is something I would like to incorporate in a second edition of the same test. Including a team's possession and average defensive line position might paint a more accurate picture of the significance of defensive contributions. As the result of my paper indicates, working on models based on player tracking that take into account more advanced metrics, should be treated with more focus in arguments and decisions involving defensive quality.

5. References

[1] Stöckl, Michael, et al. "Making offensive play predictable-using a graph convolutional network to understand defensive performance in soccer." *Proceedings of the 15th MIT Sloan Sports Analytics Conference*. Vol. 2022. 2021.

[2] Liu, Guiliang, et al. "Deep soccer analytics: learning an action-value function for evaluating soccer players." *Data Mining and Knowledge Discovery* 34.5 (2020): 1531-1559.

[3] Drew Olsen, *Defense Metrics, Possession Stats Analysis, Possession With Purpose, Team Analysis*, https://www.americansocceranalysis.com/home/category/Defense+Metrics, 13 November 2022.

6. Appendix

Curated Dataframe – Aggregated per90 Defensive Data for teams (2018-19,2019-20) https://docs.google.com/spreadsheets/d/1n6WvI0fwxd5FdKpp0rPCJB9aRxskif0P-FLwmtdf0v4/edit?usp=sharing

Jupyter Notebook – Notebook used to modify and merge relevant dataframes https://drive.google.com/file/d/1xIWwIqz2ZQah4qzaeJzEdVJ739aLSa7C/view?usp=share_link

xG – expected goal is the value given to a shooting action in soccer based on how likely it is that the shot would be a goal. The likeliness is measured based on a model trained on player tracking that takes into account defenders' position, distance and angle to goal while taking the shot

xA – expected assist is the value given to a pass leading to a shot worth a certain amount of **xG**.

Rstudio Code:

```
> def_data<-read.csv('team_def_data_1820_final.csv')

> tackles_model <- lm(goals_per_90 ~ tackles, def_data)
> summary(tackles_model)

> tackles_model <- lm(goals_per_90 ~ tackles + tackles_def_3rd, def_data)
> summary(tackles_model)

> tackles_model <- lm(goals_per_90 ~ tackles + tackles_def_3rd +
blocks+interceptions, def_data)
> summary(tackles_model)
```