„Win or defeat? What decides a football match?"
A statistical analysis of success factors in professional football

Zur Erlangung des akademischen Grades eines

DOKTORS DER PHILOSOPHIE (Dr. phil.)

von der KIT-Fakultät für Geistes- und Sozialwissenschaften

des Karlsruher Instituts für Technologie (KIT)

angenommene

DISSERTATION

von

Hannes Lepschy

KIT-Dekan: Prof. Dr. Michael Schefczyk

30 Gutachter: Prof. Dr. Alexander Woll
40 Gutachter: PD Dr. Hagen Wäsche

Tag der mündlichen Prüfung: 27. April 2022

Karlsruher Institut für Technologie

Contents

Contents

Acknowledgements

This thesis would have never been written without the support and guidance of many people and I am very thankful to everyone.

I want to thank my "Doktorvater" Prof. Dr. Alexander Woll who has given me the possibility to write this thesis and who made sure I had everything I need to succeed. He put a high level of trust in this project and me. His advice and direction were exceptional at all times!

I also want to thank my mentor and second reviewer PD Dr. Hagen Wäsche. He was part of my journey from the beginning some years ago. This thesis would not have been possible without him! He not only gave my advice or new insights, he also helped to overcome every difficulty on the way. His optimism and support gave me the confidence pursuing the ideas in this thesis.

I am thankful for the helpful discussions with Peter Kappeler and Claudia Fichtel who helped so much to improve this thesis.

Last but not least I want to thank my family for the overwhelming support in writing this thesis, especially my wife Carolin. They always believed in me and this project and their constant support and encouragement made this possible.

Acknowledgements

Summary

Football[1] is the most popular sport in the world. The knowledge about factors that distinguish between winners and losers represents important information for the interested parties. To contribute to this knowledge, this dissertation investigates the relationship between various physical and contextual factors and success in football.

The first goal of this dissertation is a systematic review of the current research regarding success factors in football (first study). Despite a growing scientific interest in football and the underlying success factors there are insufficiencies in the selected variables as well as the samples used. The review also revealed a lack of research regarding the German Bundesliga in the use of predictive designs and the control of important variables. Consequently, the second and third studies analyze the German Bundesliga and World Cup tournaments, respectively, using a broad selection of important variables.

Overall, the dissertation contains five main chapters. The first and last chapter serve as the frame for the published articles. Chapter 1 provides a general introduction. It places the thesis into the current research regarding success factors in football and introduces the theoretical and methodological background. The notational analysis provides the groundwork of data collecting and performance analysis in football. Additionally, the history of research in football and the overall methodology are described.

Chapter 2 presents a systematic review of the existing literature concerning success factors in football. An initial keyword search of published studies in 2016 or before revealed 19,161 articles. Of those, 68 studies were included in the review and clustered according to comparative studies, predictive studies and studies of home advantage. The review revealed effects of a broad variety of variables. The most influential variables appear to be goal efficiency (number of goals divided by the number of shots), number of shots, ball possession, pass accuracy/successful passes along with quality of opponent and match location. The review also disclosed a deficit in predictive studies especially about the German Bundesliga as well as methodological shortcomings, in particular, a small sample size and a lack of clear operational definitions.

---

[1] The term „football" will be used in this dissertation and it is equivalent to "soccer" or "association football"

Summary

Chapter 3 is a study about the success factors in the German Bundesliga of three consecutive seasons (2014/2015 until 2016/2017). The study included 918 matches and investigated the effect of 29 variables concerning success in football. It incorporated also market value as a contextual variable to examine a possible link between success and market value. This was the first study to use market value in a predictive design. The data were analyzed through a generalized ordered logit regression to account for the need for more predictive data and the violation of the assumption of proportional odds. Marginal effects (command *margins* in STATA) were used to interpret the result. To facilitate a more precise analysis of success factors a new approach considering only close matches was utilized as well. The model was also split into a home and an away team approach and revealed a difference between playing at home or away as well as the predictive power of this variable. Duel success is only significant for away teams and a higher market value seems to have a more positive impact for away teams. Defensive errors, goal efficiency, shots from counter attacks, shots on target, and total shots had the greatest impact on winning or losing. Additionally, crosses (negative effect) and market value (positive) were significantly related to success. The quality of the opponent and home advantage emerged as significant contextual effects, confirming results of previous research.

In Chapter 4 the FIFA World Cup 2014 in Brazil and the FIFA World Cup 2018 in Russia are studied. The study comprised of 128 matches and investigated 29 variables utilizing a generalized ordered logit approach. Only close matches were analyzed. They were also analyzed twice since the home team (first mentioned team) on the schedule is not playing at home ground except for twelve matches of the respective host team. This was also the first time market value was included in a study of success factors of a tournament of national teams. The results showed that defensive errors, goal efficiency, duel success, tackles success, shots from counter attacks, clearances, and crosses had a significant influence on winning a match during those tournaments. However, the full model could only account for about one third of the variance in the results, reflecting the multifaceted structure of success factors in football.

Chapter 5 integrates the previously separate discussions and provides a general discussion. It provides a holistic view on the previous chapters and allows for a broader understanding of success factors in football. It underlines the similarities between league competition on a club level and tournaments on the national team level. Many of the success factors are equally appearing in both conditions (e.g. goal efficiency, defensive errors,

shots from counter attack, crosses) but there are also noteworthy differences (e.g. market value, tackles, and shots from inside penalty area) which need to be addresses in future research. Nevertheless, the conclusion that efficiency factors are more important than frequencies was valid in both studies.

Summary

Zusammenfassung

Fußball[2] ist die populärste Sportart der Welt. Das Wissen um Faktoren, die zwischen Gewinnern und Verlieren unterscheiden, ist eine wichtige Information für alle Beteiligten. Daher untersucht die vorliegende Dissertation die Beziehung zwischen technischen sowie kontextuellen Faktoren im Hinblick auf Erfolg im Fußball.

Das erste Ziel dieser Dissertation war ein systematischer Literaturreview über die aktuelle Forschung zu Erfolgsfaktoren im Fußball (erste Studie). Trotz des wachsenden wissenschaftlichen Interesses am Fußball und den zugrundeliegenden Erfolgsfaktoren gibt es Mängel bei den ausgewählten Variablen sowie den gewählten Stichproben. Die Übersichtsstudie ergab auch einen Mangel an Forschung in Bezug auf die deutsche Bundesliga in der Verwendung von prädiktiven Designs und der Kontrolle wichtiger Variablen. Die zweite und dritte Studie analysierten daher die deutsche Bundesliga bzw. WM-Turniere anhand einer breiten Auswahl wichtiger Variablen.

Die Dissertation besteht insgesamt aus fünf Hauptkapiteln. Das erste und letzte Kapitel fungieren dabei als Manteltext für die veröffentlichten Artikel. Kapitel 1 bietet eine allgemeine Einführung. Es platziert die Arbeit in die aktuelle Forschung zu Erfolgsfaktoren im Fußball und führt in den theoretischen und methodischen Hintergrund ein. Die Sportspielanalyse ist die Grundlage für die Datenerfassung und Leistungsanalyse im Fußball. Außerdem werden die Geschichte der Forschung im Fußball und die generelle Methodik beschrieben.

Kapitel 2 präsentiert eine systematische Übersicht über die existierende Literatur zu Erfolgsfaktoren im Fußball. Eine erste Stichwortsuche zu veröffentlichten Studien bis 2016 ergab 19.161 Artikel. Davon wurden 68 Studien ausgewählt und in Vergleichsstudien, Vorhersagestudien und Studien zum Heimvorteil gruppiert. Die Übersichtsstudie ergab eine Vielzahl von untersuchten Variablen. Die einflussreichsten Variablen scheinen Tor-Effizienz (Anzahl der Tore geteilt durch die Anzahl der Schüsse), Schüsse, Ballbesitz, Passgenauigkeit / erfolgreiche Pässe sowie die Qualität des Gegners und der Spielort zu sein. Die Übersichtsstudie ergab ebenfalls ein Defizit an Vorhersagestudien insbesondere zur deutschen Bundesliga sowie methodische Defizite; hier insbesondere eine kleine Stichprobengröße und das Fehlen klarer Operationalisierungen der Variablen.

---

[2] Der Begriff „football" wird in der vorliegenden Dissertation gleichbedeutend mit den Begriffen "soccer" oder "association football" verwendet

Zusammenfassung

Kapitel 3 ist eine Studie zu den Erfolgsfaktoren der deutschen Bundesliga für drei aufeinander folgenden Spielzeiten (2014/2015 bis 2016/2017). Die Studie umfasste 918 Spiele und untersuchte den Einfluss von 29 Variablen auf den Erfolg im Fußball. Es berücksichtigte auch den Marktwert als Kontextvariable, um erstmals einen möglichen Zusammenhang von Erfolg und Marktwert zu untersuchen. Die Daten wurden durch eine generalized-ordered Logit-Regression analysiert, um den Bedarf an mehr prädiktiven Daten und die Verletzung der Annahme von proportional Odds zu berücksichtigen. Marginal Effects (Befehl *margins* in STATA) wurden verwendet, um das Ergebnis zu interpretieren. Um eine genauere Analyse der Erfolgsfaktoren zu ermöglichen wurde ein neuer Ansatz verwendet, bei dem nur enge Spiele berücksichtigt wurden (Closeness-of-the-game-Approach). Das Modell wurde in eine Heim- und eine Auswärtsperspektive aufgegliedert, und zeigten einen Unterschied zwischen Heim- und Auswärtsspielen, sowie dem Ausmaß, in dem diese Variablen das Gewinnen oder Verlieren beeinflussen. Die Zweikampfquote ist nur für Auswärtsteams signifikant, und ein höherer Marktwert scheint sich für Gastmannschaften positiver auszuwirken. Fehler in der Defensive, Toreffizienz, Schüsse aus Kontern, Schüsse auf das Tor und Schüsse gesamt hatten den größten Einfluss auf den Spielausgang. Zusätzlich hatten die Häufigkeit von Flanken (negativer Effekt) und Marktwert (positiv) einen signifikanten Zusammenhang zum Erfolg. Mit Qualität des Gegners und Heimvorteil zeigten zwei kontextbezogene Einflussfaktoren, wie in früheren Untersuchungen, signifikante Effekte.

In Kapitel 4 werden die FIFA Fußball-Weltmeisterschaft 2014 in Brasilien und die FIFA Fußball-Weltmeisterschaft 2018 in Russland untersucht. Die Studie umfasste 128 Spiele und untersuchte 29 Variablen unter Verwendung eines generalized-ordered Logit-Regression. Es wurden nur enge Spiele analysiert. Diese wurden zweimal untersucht, da die Heimmannschaft (zuerst genannte Mannschaft) auf dem Spielplan, mit Ausnahme von zwölf Spielen der jeweiligen Gastgebernation, nicht zu Hause spielte. Dies war auch die erste Studie, die Marktwert in die Untersuchung der Erfolgsfaktoren eines Turniers der Nationalmannschaften einbezog. Die Ergebnisse zeigten, dass Fehler in der Defensive, Toreffizienz, Zweikampfquote, erfolgreiche Tackles, Schüsse nach Kontern, Clearances und Flanken einen signifikanten Einfluss auf das Gewinnen eines Spieles während dieser Turniere hatten. Das Gesamtmodell konnte jedoch nur etwa ein Drittel der Varianz der Ergebnisse erklären, was die vielfältige Struktur von Erfolgsfaktoren im Fußball zeigt.

Kapitel 5 integriert die zuvor getrennten Diskussionen und führt diese in einer allgemeinen Diskussion zusammen. Es bietet einen ganzheitlichen Überblick über die vorhergehenden Kapitel und ermöglicht ein umfassenderes Verständnis der Erfolgsfaktoren im Fußball. Es unterstreicht die Ähnlichkeiten zwischen Vereinswettbewerben und Turnieren der Nationalmannschaft. Viele der Erfolgsfaktoren zeigten sich gleichermaßen unter beiden Bedingungen (bspw. Toreffizienz, defensive Fehler, Torschüsse nach Kontern, Flanken), jedoch gibt es auch bedeutsame Unterschiede (bspw. Marktwert, Tackles, Torschüsse innerhalb des Strafraumes), die in der zukünftigen Forschung berücksichtigt werden müssen. Dennoch galt die Folgerung, dass Effizienzfaktoren wichtiger sind als Häufigkeiten in beiden Studien.

Zusammenfassung

List of figures

List of figures

List of tables

List of tables

1. General introduction

1.1. Preface

Football[3] is the most popular sport in the world (Dvorak et al., 2004). Every week, millions of fans are visiting stadiums around the world. For example, in the German Bundesliga, on average more than 40,000 spectators are watching a game live in the stadium. This number has doubled in the past 50 years but has remained constant for about 10 years (Deutscher Fußball-Bund, 2018). Besides direct spectators, additional millions of people are also watching football on TV. For example, the FIFA World Cup 2018 in Russia had the biggest TV crowd in history. Half of the world population, an estimated 3.572 billion people, watched some broadcast of it (FIFA, 2018b). In the English Premier League, average views per game are still near one million, despite recent drops in viewership (Kuper, 2018). The engagement in football on social media increased by even 917% in two years (Kuper, 2018). In addition, the social media engagement during the Russia World Cup was the most engaging FIFA World Cup with more than 7.5 billion engagements across all digital platforms (FIFA, 2018c). The interest also comes with considerable revenue. The 20 most successful clubs worldwide generated 8.3 billion Euros in revenue during the season 2017/18 alone. The sporting success of the clubs is a significant part of their economic success (Deloitte Sports Business Group, 2019).

The increasing economic impact and interest went along with an increase in scientific attention, including match analysis (Sarmento et al., 2014). Part of the match analysis is to investigate determinants of successful performance. Those actions can be defined as performance indicators or success factors e.g., passes or shots (Hughes & Bartlett, 2002). Although the performance in ball games is more difficult to evaluate than it is in individual sports, the knowledge of those performance indicators is vital for the understanding of the nature of success in football (Carling et al., 2005). Nevertheless, there is still uncertainty about the influence of certain factors; for example, how ball possession is related to success (Collet, 2013; Lago et al., 2010). In addition, some interesting variables, like market value, have been hardly studied, or only a few variables are considered in the calculations overlooking the complex nature of success in football. Similarly, there is a research focus on the English Premier League and international tournaments. Other leagues, like the German Bundesliga, have been rarely studied in terms of success factors.

---

[3] The term „football" will be used in this dissertation and it is equivalent to "soccer" or "association football"

Subsequently, this thesis explores the success factors in football that increase the probability of winning and losing, respectively. In particular, the success factors of the German Bundesliga of three consecutive seasons as well as a broad variety of success factors during the Word Cups 2014 in Brazil and 2018 in Russia are the focus of this thesis.

1.2. Outline of this thesis

The thesis consists of five main chapters. In the first chapter, the history of football and science is briefly discussed. Moreover, the general methodology is introduced, and the terminology is established. The chapter closes with the aims and scope of this thesis.

The following three chapters contain three research articles that were published in international peer-reviewed journals: The first chapter provides an overview of the state of research as a systematic literature review. The two subsequent articles deal with the research questions mentioned above.

Chapter 2:

Lepschy, H., Wäsche, H., Woll, A. (2018). How to be Successful in Football: A Systematic Review. The Open Sports Sciences Journal, 11(1).
https://doi.org/10.2174/1875399X01811010003

Summary: A systematic literature review analyzing existing studies about success factors in football was commenced. Lastly, 68 articles were included in the review. The studies were grouped regarding comparative analyses, predictive analyses and analyses of home advantage. Altogether, 76 different variables were investigated in the reviewed papers. It seemed that the most significant variables are efficiency, shots on goal, ball possession, pass accuracy/successful passes as well as quality of opponent and match location. Furthermore, new statistical methods were used to reveal interactions among these variables such as discriminant analysis, factor analysis and regression analysis. The studies showed methodological deficits such as clear operational definitions of investigated variables and small sample sizes. The review allows a comprehensive identification of critical success factors in football and sheds light on utilized methodological approaches.

Chapter 3:

Lepschy, H., Wäsche, H., & Woll, A. (2020). Success factors in football: An Analysis of the German Bundesliga. International Journal of Performance Analysis in Sport.
https://doi.org/10.1080/24748668.2020.1726157

Summary: In the second article, three seasons of the German Bundesliga (2014/2015 until 2016/2017) with a total of 918 matches were investigated. To facilitate a more precise analysis of success factors only close matches were included and the home and away team perspective was analyzed separately. Consequently, 29 variables were included in a generalized ordered logit approach. The results showed that, defensive errors, market value, goal efficiency, shots from counter attacks, shots on target, and total shots have the greatest impact. Furthermore, crosses showed a negative relationship with success. In addition, the opponent and home advantage were important contextual effects. Duel success was only significant for away teams and a higher market value seems to have a more positive impact for them. This study provides novel data and contributes to prior results from other European leagues.

Chapter 4:

Lepschy, H., Woll, A., Wäsche, H., (under review). Success factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil.

Summary: The third article studies the success factors during the World Cup 2018 in Russia and the World Cup 2014 in Brazil. In total, 128 matches were analyzed using a generalized order logit approach. 29 variables were identified from previous research. The results showed that defensive errors, goal efficiency, duel success, tackles success, shots from counter attacks, clearances, and crosses have a significant influence on winning a match during those tournaments. Ball possession, distance and market value of the teams had no significant effect on success. In general, most of the critical success factors and those with the highest impact on winning close games were defensive actions. Besides, the results suggest that direct play and pressing were more effective than ball possession play. The study contributes to a better understanding of success factors.

Finally, Chapter five offers a general discussion and conclusions. The results of the research studies are discussed cohesively and areas for future research are identified. Therefore, chapter one and chapter five act as a frame for the published articles. Thereby, the aim and scope of this thesis are framed, the overall methodology is described, and the individual results are discussed on an integrative level to broaden the understanding of success factors in football.

1.3. Brief history of football and science

Despite the long history of football, similar ball games were already played more than 2000 years ago (FIFA, 2007), the science behind the game has been around for only about 50 years (Drust, 2019). Understandably it is more difficult to determine the obvious starting point of

science in football than for example the founding of the Football Association on October 26th, 1863 (Drust, 2019; FIFA, 2007). However, there are milestones that can be understood as starting points. For example, Reep and Benjamin (1968) published one of the very first articles and provided probabilities of shots, passes, and goals. Along with Reilly and Thomas (1976), who investigated the work-rate associated with different positions in football. Another milestone was the first World Congress of Science and Football in 1987 (Hughes & Franks, 2004). The first academic program in science and football was offered in 1991 at the University of Liverpool (Reilly & Williams, 2003). From there, research grew steadily and was primarily driven by the research in the United Kingdom (Drust, 2019). Today, the growing body of research can be categorized into biology and exercise physiology, biomechanics and technology, sports medicine, behavioral science and coaching, youth development and performance profiling as well as match analysis (Drust et al., 2015; Reilly & Williams, 2003).

This thesis contributes to the category match analysis. Match analysis subsumed all research with regards to "…recording and examination of behavioral events occurring during competition" (Carling et al., 2005, p. 2). A similar term often used is "performance analysis". Performance analysis can be understood as the investigation of performance gathered during actual competition or training, in contrast to data from laboratory settings or self-reports (O'Donoghue, 2009). In this thesis, both terms will used exchangeable and refer to recording and examination of behavioral events occurring during actual sports competition or training.

This being the case, one of the first published articles, the above-mentioned study by Reep and Benjamin (1968), was also one of the first match analyses. However, subsequent research remained limited for the following years, partly due to the absence of suitable academic journals (Hughes & Franks, 2004). Since the 1990s, more specific journals, research societies and conferences have increased the quantity and quality of research in match analysis (Sarmento et al., 2014). The growth in match analyses was also supported by technological progress, resulting in new systems specifically for football (Mackenzie & Cushion, 2013).

Sarmento et al. (2014) published a systematic review about match analysis in football. They found 2732 articles in their initial search but included only 53 articles in the review. The 24 articles published in 2010 and 2011 represented the last two years of their review, but half of the articles. Sarmento et al. (2014) concluded that match analysis was mainly done using descriptive and comparative approaches. The advances of predictive designs have only been used in the recent years. Mackenzie and Cushion (2013) also raised methodologic concerns in their critical review of performance analysis. They criticized small sample sizes, a lack of operational

definitions, and conflicting classifications of activity. In addition, they criticized the deficiency of conceptual clarity as well as the need for a relationship between research and practice, and researchers and practitioners. Consequently, they proposed a checklist for performance analysis research in football (Mackenzie & Cushion, 2013):

- The nature of the competition that is to be investigated
- Providing statistical justification for the sample size
- Context to the sample used (i.e. location, period of season, opposition faced etc.).
- Comprehensive and published operational definitions for the variable(s) under investigation and ensure specific contextual information is included.
- When researching the physical aspects of football performance, considering previous research in order to better inform the thresholds adopted to ensure research that is comparable.

The focus of performance analysis has been mainly on frequency distributions of certain game events like shots or running distance. A new approach, triggered by advances in sensor technology, now allows for positional data of individual players and the ball to be analyzed (Memmert & Rein, 2018). Recently, performance analysts also investigated tactical behaviors in football based on collective activities. The variables used in many of those studies can be put into the broad categories of measures of position, distances, playing spaces and numerical relations (Low et al., 2019). Both approaches allow for a more comprehensive analysis of performance in the future.

All the above underscores the importance of continuing this line of research, since not only rules, and tactics change over time, but also the body of research is growing at a much faster rate than it has in the twentieth century.

1.4. General methodology

1.4.1. Analyzing a football match

The definition of a performance indicator or performance factor needs to be clear prior to the beginning of an analysis. Hughes and Bartlett (2002) defined a performance indicator as "… a selection, or combination of action variables that aims to define some or all aspects of a performance. Clearly, to be useful, performance indicators should relate to successful performance or outcome" (p. 739). In a second step, the identification of performance factors also depends on the classification of the game that should be analyzed. Read and Edwards (1992) structured formal games into three categories, net/wall games, invasion games, and striking/fielding

games. Football belongs to the category invasion games, within that it fits to the subcategory goal-striking games (Hughes & Bartlett, 2002). The performance factors can now be structured in four types: match classifications (e.g., crosses), biomechanical (e.g., kicking), technical (e.g., tackles), and tactical (e.g., shot types) which makes clear that performance in football is a multifaceted concept that can only be explained by a combined approach (Hughes & Bartlett, 2002).

A football match can be also analyzed in many ways depending on the research scope. For example, if the aim of the study is to determine the effects of the position of the shots on goal scoring probability, the position data of shots fired are an essential part of the data collection. In contrast, if the research aim is to examine the effects of running distance on the outcome of a match, the position data of shots fired are not essential. Thus, it needs to be determined how to gather the required data and information before a football match can be analyzed. In general, the decision needs to be made whether primary data, also called raw data, are needed and accessible or whether secondary data are available and sufficient for the research purpose (Hox & Boeije, 2005).

The method of collecting primary data related to performance in football is better known as notational analysis. With this method, movements are analyzed, tactics and techniques are evaluated and statistically compiled (Hughes & Franks, 2004). The first publication in notational analyses in any sports was conducted by Fullerton in 1912 (Hughes & Franks, 2004). Two of the earliest articles in football using hand notation systems were conducted by Reep and Benjamin (1968) and Reilly and Thomas (1976). Reep and Benjamin (1968) collected data from 3,213 match of the English League between 1953 and 1968 and found that 80 percent of goals were scored after three or more passes and 50 percent of goals originated from possession gained in the last quarter of the field. Reilly and Thomas (1976) studied the intensity and extent of activities during a match, described the distance covered for different positions and discovered that a player is only in possession of the ball for less than two percent of the game.

Despite being considered accurate and inexpensive hand notational systems have some disadvantage such as a considerable learning time and many man-hours of work. Computerized notation systems helped to overcome some of those disadvantages (Hughes, 1988). Also, methods have progressed with the advances in technology to include more objective and quantitative measures of performance (Hughes et al., 2007). Nowadays, hardware and software enable companies to collect live data of football matches efficiently and also to store those data for years (Liu et al., 2013). However, these systems still involve human operators who can make mistakes, limiting their reliability. Therefore, reliability evaluations needs to be done to ensure the

understanding of the measurement errors (O'Donoghue, 2007). For example, the accuracy and reliability of Prozone Sports Ltd®, Gecasport, Amisco Pro®, and Opta Sportdata has been shown in the past. Most recently, Liu et al. (2013) showed kappa values of 0.92 (home team) and 0.94 (away team) for a match in the Spanish La Liga, respectively. This indicates that the involved observers counted the same action or events into the same performance indicator. Correspondingly, a high inter-operator reliability is essential for further use of those data in scientific research. The use of those data is an example of secondary data.

The analysis of a football match cannot only be viewed in terms of the data source. It can also be differentiated by the type of analysis into descriptive, comparative and predictive studies (Marcelino et al., 2011; Sarmento et al., 2014). Descriptive studies simply describe actions and events of a football match (e.g., distance covered, passes played). Comparative analyses not only describe performance indicators they also compare those to a reference (e.g., shots on goal of top three compared to bottom three using a t-test). Predictive analyses as well compare performance indicators also provide information to predict future events (e.g., discriminant analysis of winning and losing teams). To carry out a comparative analysis or a predictive analysis, the dependent variable needs to be defined. This can be the final table (Oberstone, 2009), the points earned (Coates et al., 2016), scoring a goal (Wright et al., 2011), remaining in the competition (Delgado-Bordonau et al., 2013) or winning/losing a match (Lago et al., 2016). In terms of winning or losing, the analysis can be further differentiated between a result-based or goal-based approach (Goddard, 2005). In the result-based approach, only the result of the match is used in terms of win, draw, or loss. In contrast, the goal-based approach also accounts for the difference in goals scored, which is assumed to carry more information than the result-based approach. Nonetheless, the goal-based approach is not resulting in a better model performance (Goddard, 2005).

An alternative approach to assess the outcome is to view matches as close and unbalanced. Here, the sample is split into two groups of matches, one with a narrow goal difference (close matches) and one with a wide goal difference (unbalanced matches) (Vaz et al., 2010). This method appears to have a better model performance then the goal-based approach and can overcome the moderator effect of one team which does not play at its best level (Gómez et al., 2014; Higham et al., 2014; Vaz et al., 2010; Sampaio et al., 2010; Vaz et al., 2010). The result-based approach, focusing on close matches only, can be used to achieve a sufficient model performance despite using only a subset of the available information. The result-based approach also allows for an ordered-logit regression because of the scale of measure is ordinal (McCullagh,

1980). Assuming wining is the favored outcome; the result variable can be rearranged to 0 being a loss, 1 being a draw and 2 being a victory. In addition, a logistic regression, unlike a linear regression, does neither require a linear relationship between the dependent and independent variables nor homoscedasticity (Greene, 2011). Finally, the error distribution does not need to be normally distributed, which could be violated in football analysis because results in football mostly follow a Poisson distribution (Dixon & Coles, 1997; Maher, 1982; Myers, 1990; Rue & Salvesen, 2000).

Nevertheless, the ordered-logit regression also makes some assumptions. The order of the dependent variables has already been mentioned above. Secondly, there needs to be no multicollinearity because this would lead to unreliable data (Kleinbaum & Klein, 2010). Multicollinearity describes the situation in which the covariate of one independent variable correlates with the covariate of another independent variable (Zuur et al., 2010). The variance inflation factor (VIF) can be used to control for the level of multicollinearity. The cut-off value is usually between 5 and 10 (Craney & Surles, 2002). Independent variables with a higher value than the cut-off would need to be excluded in the analysis to allow for reliable results. However, the process should be iterative, starting with the variable with the highest VIF value. Afterwards, the VIF values should be calculated again. If there is another variable with a VIF value above the cut-off value the process is repeated (Craney & Surles, 2002). Finally, the remaining independent variables should have a VIF value below the cut-off value before analyzing the data further.

The last assumption of the ordered logit regression is called proportional odds that is why the model is also called proportional odds model. This means that in the model the relationship between each pair of outcome groups is the same (Kleinbaum & Klein, 2010). The violation of proportional odds can lead to biased results (Fullerton, 2009). However, proportional odds can be tested using the Brant test. This test evaluates whether the observed deviations from the ordered logit regression model are larger than what might be credited to chance alone (Brant, 1990; Williams, 2016). A significant Brant test means that the assumption of proportional odds is violated (Williams, 2016). However, the use of a multinomial logistic regression, which would fit the data in case of a violation of the proportional odds, is not desirable here, since the information from the ordering would be not fully accounted for (Kleinbaum & Klein, 2010). Therefore, the generalized ordered logit approach can be a better alternative (Williams, 2016). The generalized ordered logit can be defined as (Williams, 2006):

$$P(Y_i > j) = \frac{exp(\alpha_j + X_i \beta_j)}{1 + [exp(\alpha_j + X_i \beta_j)]}, j = 1, 2, ..., M - 1$$

Unquestionably, the ordered logit model is a distinct case of the generalized ordered logit model, where the betas are the same for each j (Williams, 2016). In case of no violation of the proportional odds, the generalized ordered logit model would produce the same results as the ordered logit model. However, the generalized ordered logit model can also reduce errors in statistical significance, which could lead to conclude inaccurately that an independent variable has no effect on the result. Since the software is available to calculate the model effortlessly, the generalized ordered logit model should be considered if it can better serve the needs of the research goal (Williams, 2016).

Regardless of the logistic regression model used, the results of the analysis need to be interpreted to draw meaningful conclusions. The results of the logistic regression indicate whether an independent variable has a significant effect and whether this effect is positive or negative, but it can be challenging to determine the value of the effect on the dependent variable. A popular method of making the results more intuitively meaningful are marginal effects (Williams, 2012). Cameron and Trivedi (2010) noted that the marginal effects measure the effect on the conditional mean of y of a change in one of the regressors, for example xj., which equals the relevant slope coefficient in a linear regression.

Three common choices for the evaluation of marginal effects are the average marginal effects (AME), marginal effects at mean (MEM), and marginal effects at a representative value (MER). In the current practice it is favorable to use the AME over the MEM whenever possible (Greene, 2011). Williams (2012) described the main argument for AME as a demand for realism because the sample means used in MEM might refer to either absent or inherently senseless observations. He noted that the reason MEM is most often used is that it is a good approximation of AME. MER can be preferable over the two alternatives if more than a single estimate of the marginal effects is required. For example, in a hypothetical experiment about diabetes and gender, AME and MEM could lead to the conclusion that being female leads to an increased chance of diabetes by 0.6%. However, age has a great effect on diabetes, which could be incorporated using MER. This could lead to a more sophisticated conclusion, like at the age of 20, the effect is 0.09% but at age 70 it is 1.5% (Williams, 2012). In general, marginal effects also make it possible to draw intuitive figures to demonstrate the effect (for example see Figure 2, Chapter 3.4). However, even the most powerful statistical approach cannot compensate for a lack of transparency and operational definitions (Mackenzie & Cushion, 2013).

1.4.2. Definition of variables

As stated above, the definition of success in my thesis is winning the match. Therefore, the result of a match in terms of win, draw or loss is the dependent variable. Since the independent variables were collected from public website, their operational definition is also used here and as follows (Liu et al., 2013; Liu et al., 2015; Opta, 2018):

- *Total Shots:* Is the sum of shots on target (see below), shots off target (a clear attempt to score that goes over or wide of the goal without making contact with another player or would have gone over or wide of the goal but for being stopped by a goalkeeper's save or by an outfield player or directly hits the frame of the goal and a goal is not scored) and blocked shots (a blocked shot is defined as any clear attempt to score which is going on target and it is blocked by an outfield player, where there are other defenders or a goalkeeper behind the blocker and includes shots blocked unintentionally by the shooter's own teammate).

- *Shots on target:* Any goal attempt that goes into the net or a clear attempt to score that would have gone into the net but saved by the goalkeeper or stopped by a player who is the last-man with the goalkeeper having no chance of preventing the goal (last line block). Shots directly hitting the frame of the goal are not counted as shots on target, unless the ball goes in and is awarded as a goal. In addition, shots blocked by another player, who is not the last man, are not counted as shots on target.

- *Shots from counter attack:* Any goal attempt produced from a counter attack. A counter attack is an attempt created after the defensive quickly turn defense into attack winning the ball in their own half. A counter-attack situation is recorded after (a) the ball is turned over in the defensive half; (b) the ball is quickly played (6 s, 3 passes) into the attacking third (the ball must be under control); (c) the defense had four or less defenders in a position to defend the attack and attacking players must match or outnumber the defensive teams players and (d) the ball is fully under control in the oppositions defensive third.

- *Shots from inside 6-yard box:* Any goal attempt occurred in the 6-yard box. A shot on the 6-yard line will count as being inside the box

- *Shots from inside penalty area:* Any goal attempt occurred in the 18-yard box. A shot on the 18-yard line will count as being inside the box.

- *Goal efficiency:* Calculated through goals multiplied by 100 and divided by total shots.

- *Ball possession (%):* Possessions are defined as one or more sequences in a row belonging to the same team. A possession is ended by the opposition gaining control of the ball. The value is calculated as the duration of ball possession as a proportion of total duration when the ball was in play.

- *Passes:* Any intentional played ball from one player to another. Passes include open play passes, goal kicks, corners and free kicks played as pass – but exclude crosses, keeper throws and throw-ins.

- *Pass accuracy (%):* Successful passes as a proportion of total passes. A successful pass is a pass that goes to a teammate directly without a touch from an opposition player.

- *Long passes:* Any attempted pass of 25 yards or more.

- *Short passes:* Any attempted pass of less than 25 yards.

- *Average pass streak:* The average number of passes attempted in each series of consecutive passes.

- *Crosses:* Any intentional played ball from a wide position intending to reach a teammate in a specific area in front of the goal.

- *Successful dribbles:* A dribble is an attempt by a player to beat an opponent when they have possession of the ball. A successful dribble means the player beats the defender while retaining possession.

- *Offsides:* Given to the player regarded to be in an offside position where a free kick is awarded. If two or more players are in an offside position when the pass is played, the player considered being most active and trying to play the ball is given offside. The total of all given offsides to players of one team is the amount of offsides for the respective team.

- *Corners:* When the ball goes out of play resulting in a corner kick.

- *Aerials won:* This is where two players challenge in the air against each other. The player that wins the ball is deemed to have won the duel.

- *Distance:* The total distance in kilometer covered by a team during the match at any speed. The distance covered by each player of the team is totalized to get the distance of the team.

- *Successful tackles:* A tackle is defined as where a player connects with the ball in a ground challenge where he successfully takes the ball away from the player in possession. The tackled player must clearly be in possession of the ball before the tackle is made. It is not a tackle, when a player cuts out a pass by any means.

- *Tackles success (%):* Successful tackles as a proportion of the total of successful tackles and missed tackles. A missed tackle is where a player attempts to challenge for the ball and does not make it.

- *Fouls:* A foul is defined as any infringement that is penalized as foul play by a referee. Offsides are not given as a foul conceded.

- *Yellow cards:* Every yellow card given to a player

- *Red cards:* Every red card given to a player, including straight red card and a red card from the second yellow card

- *Defensive errors:* A mistake made by a player losing the ball that leads to a shot or a goal.

- *Duel success (%):* A duel is a 50-50 contest between two players of opposing sides in the match. For every duel won there is a corresponding duel lost depending on the outcome of the contest. This is the proportion of duels won divided by duels lost.

- *Clearances:* A defensive action where a player kicks the ball away from his own goal with no intended recipient.

- *Interceptions:* This is where a player anticipates an opponent's pass and intercepts the ball by moving into the line of the intended pass.

Liu et al. (2015) were able to show a high inter-operator reliability for the system used by OPTA Sports so that their definitions seem to be sufficient for identifying the correct actions on the field.

The data for the market value and the average age of teams (i.e., average age of the starting formation) was drawn from the website Transfermarkt.de. The average age of the starting formation is the average of the age of the first eleven players who start the match for the respective team. The age is an integer which is not rounded, for example if a players' birthday is March 22nd, 1990 and game day is March 21st, 2010, the age of the player used is 19. The market value is estimated based on performance (e.g., successful passes, goals) including stability of the performance (recent performance has a higher value than past performance), experience (number of games played nationally and internationally including national team), perspectives for the future (anticipated value for younger players results in additional value), and prestige (public perception of the player and public perception of the club) (Transfermarkt.de, 2017).

1.5. Aim and scope of this thesis

This thesis aims to identify the success factors in professional football. Therefore, the above-mentioned variables were analyzed in two different settings considering the methodological caveats discussed before. However, to narrow down the most influential variables, a systematic literature review was conducted first (Chapter 2). This also allowed to incorporate existing findings into the design of the subsequent studies (Chapter 3 and 4). Hence, the purpose of this thesis can be split into two main goals:

i. A comprehensive review of the available literature on success factors in football focusing on physical and contextual factors related to win a match.

ii. A comprehensive investigation of success factors in two different settings using a novel methodological approach as well as a broad selection of variables.

Due to the absence of an existing review specifically dealing with success factors in football as well as conflictive previous research findings, a review of the existing literature seemed indicated. For example, Lago et al. (2010) showed that possession is a significant success factor analyzing a full season of the Spanish La Liga. In contrast, Collet (2013) studied the Top 5 leagues in Europe as well as the UEFA Champions League and UEFA Europa League and showed that in both the Spanish La Liga and the Top 5 European leagues overall possession time is negatively linked to success. Furthermore, the available data of a football match are extensive at the present time. For example, the website www.whoscored.com provides almost 200 individual types of data for one match. This wealth of information cannot be put into one model of success because of the multicollinearity problem (Graham, 2003). Rather, an educated selection of variables providing the most value for the research topic has to be made. Moreover, a review can also reveal overarching gaps in current research and highlight methodological concerns (Eagly & Wood, 1994). Consequently, the review (Chapter 2) deals with peer-reviewed research regarding success factors in professional football to reveal the most promising variables and to identify questions for future research.

Subsequently, the insights of the review were used for the design of the two subsequent empirical studies. At first, the German Bundesliga was selected as the subject of further research due to the small number of existing studies about it despite being one of the top football leagues in Europe. Secondly, a large set of variables was selected based on the literature; notably adding market value for the first time. Additionally, the variables belong three types of the four types of performance factors described earlier (see 1.4.1). Consequently, the aim of the study in Chapter 3 was to reveal the success factors of the German Bundesliga for three consecutive seasons.

Chapter 4 consists of a similar set-up, but focusses on national teams; specifically the success factor of the World Cup 2014 and the World Cup 2018. This approach allows for a comparison of the identified success factors between club teams and national teams and an identification of future research questions.

In summary, this thesis will be guided by the research question about the identity of quantitative performance factors in professional football, their predictive power for the outcome of a football match, and the possible importance of differentiating between home and away teams as well as club and national competition, respectively.

2. Review of the state of research

This is an adaption of an article published by Bentham Science Publishers in The Open Sports Sciences Journal on 29/06/2018, available online: https://doi.org/10.2174/1875399X01811010003

The original research article was published as:

Lepschy, H., Wäsche, H., Woll, A. (2018). How to be Successful in Football: A Systematic Review. The Open Sports Sciences Journal, 11 (1). doi: 10.2174/1875399X01811010003

Review of the state of research

2.1. Abstract

Background

Despite the popularity of football, the analysis of success factors in football remains a challenge. While reviews on performance indicators in football are available, none focuses solely on the identification of success factors and addresses the large and growing body of recent research up until 2016.

Objective

To find out what determines success in football and to organize the body of literature, a systematic literature review analyzing existing studies with regard to success factors in football was undertaken.

Method

The studies included in this review had to deal with performance indicators related to success in football. The studies were published in 2016 or before. The initial search revealed 19,161 articles. Finally, sixty-eight articles were included in this review. The studies were clustered with regard to comparative analyses, predictive analyses and analyses of home advantage.

Results

In total, 76 different variables were investigated in the reviewed papers. It appeared that the most significant variables are efficiency (number of goals divided by the number of shots), shots on goal, ball possession, pass accuracy/successful passes as well as quality of opponent and match location. Moreover, new statistical methods were used to reveal interactions among these variables such as discriminant analysis, factor analysis and regression analysis. The studies showed methodological deficits such as clear operational definitions of investigated variables and small sample sizes.

Conclusion

The review allows a comprehensive identification of critical success factors in football and sheds light on utilized methodological approaches. Future research should consider precise operational definitions of the investigated variables, adequate sample sizes and the involvement of situational variables as well as their interaction.

Keywords: *match analysis, soccer, success, performance, indicator, football*

2.2. Introduction

Football or soccer (in this paper the term 'football' is used) is the most popular sports in the world. According to the "Big Count" study of FIFA (FIFA Communications Divisions, 2007) there are 270 million people involved in the match (players and referees). Moreover, football attracts millions of spectators around the world. For example, the global TV audience that followed the 2015 UEFA Champion's League final between FC Barcelona and Juventus Turin was estimated to be 180 million people from more than 200 territories (UEFA, 2015). Due to its high popularity, football stands out among sports and games. In contrast to games such as basketball or handball, football is a low scoring game, and scoring a goal is usually a rare event. For this reason, the final match score does not provide a clear picture of the teams' technical and physical performances. To understand success factors in football, various other performance indicators next to goals scored have to be considered. Football is also a sport which has elements of chance but nevertheless this does not mean successful teams are just luckier than others (Dufour W., 1993; Reilly & Williams, 2003)

To identify the factors which lead to success in football it is necessary to find performance indicators which significantly discriminate winners and losers. However, the identification of critical factors for successful performance poses a major challenge (Hughes & Franks, 2004). In 1912, Fullerton did the first work in this area of performance analysis for baseball (Eaves, 2017). In football, Reilly and Thomas (1976) performed one of the first systematic notational analyses. They used hand notation and audio tapes to analyze in detail the movements of English First Division football players (Hughes, 2003), and found out, inter alia, that a player is usually in touch with the ball for only two percent of the time. In another early performance analysis, Reep and Benjamin (1968) developed a new approach to study 3,213 matches in England between 1953 and 1968 using frequency distributions. Their analysis revealed that about 80 percent of all goals are scored after three or fewer passes and about 10 shots are needed for one goal.

A milestone for science and football was the first World Congress of Science and Football which was held in Liverpool in 1987 (Hughes & Franks, 2004). Various themes were discussed such as team management, computer-aided performance analysis and decision-making by referees (Reilly et al., 2011). In the following years, the numbers of research papers concerning football and performance analysis increased steadily (Carmichael et al., 2000; Clarke & Norman, 1995; Lago & Martín, 2007; Oberstone, 2009; Pollard & Reep, 1997). Hughes and Bartlett

(2002) reviewed and analyzed research on performance indicators in sports and defined a performance indicator as "… a selection, or combination of action variables that aims to define some or all aspects of a performance. Clearly, to be useful, performance indicators should relate to successful performance or outcome" (p. 739). Researchers also monitored match structures, summarized some performance indicators and utilized them (e.g., numbers of shots, passes, dribbles or ball possession) in various subsequent papers which provided more insight into possible success factors in football (Eaves, 2017; Hughes & Franks, 2005).

In the context of this paper, two review studies regarding performance analysis in football are noteworthy. Mackenzie and Cushion (2013) critically reviewed 60 articles (articles published up to 2010) with a focus on methodological approaches, and concluded that there is an over-emphasis of research on predictive and performance controlling variables (e.g., location, shots). They suggested an alternative approach that focuses on research that investigates athlete and coach learning to enhance our understanding of football performance. However, these factors cannot readily be operationalized as success factors. Sarmento et al. (2014) systematically reviewed 53 articles (articles published up to 2011) with a focus on major research topics and methodologies. They concluded that most studies used a comparative analysis to analyze differences between players or teams. Unlike Mackenzie and Cushion (2013), they identified a lack of predictive studies. While it was not the focus of their research, they also identified some success factors for a team such as the number of shots and shots on goal. They concluded that match location, quality of the opposition, match status and match half seem to have a greater importance for success due to the large number of studies that focused on these aspects.

Both aforementioned reviews comprised a wide variety of possible outcomes in the included articles, such as physical conditions or contextual variables. In this study, we focus solely on predictive or comparative studies that considered success as outcome (win/loss, league ranking, etc.). This allows a clear identification of the critical factors for success. Moreover, this review also considers studies published after 2011, addressing a large and growing body of recent research that has not been covered in previous reviews, and enables an assessment of the current state of the art.[4] Not only has the amount of the articles related to performance analysis in football grown substantially since 2011, also various new methodological approaches have

---

[4] The body of research on this topic has grown significantly in the last years. For example, in the three years between this review and the review of Sarmento et al. (2014) the number of predictive studies, which are the most promising studies to deliver new insights to the of success in football, has grown by more than 40 percent (see also tables 6 to 8).

been utilized. For example, Grund (2012) introduced network analysis into the research about success factors and Collet (2013) revealed new insights into the effect of ball possession using an ordered logit regression. Liu et al. (2015) used a k-means cluster analysis and a cumulative logistic regression to reveal the factors that differentiate the between winning and losing teams. Overall, the aim of this study is to provide a systematic review of the available literature on performance analysis in elite male football concerning methodologies and results to find out critical factors for success in football and to provide guidance for future research[5].

2.3.   Material and methods

The systematic review of performance indicators in elite men's football was done in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) statement (Moher et al., 2009). The last search was conducted on June 24th, 2017.

To search for relevant publications and ensure the quality of the articles, the following databases were utilized: Web of Science (the modules "Core" and "Medline"), Scopus and PubMed. Articles that were published in 2016 or before and in English were considered. The search strategy comprised search terms that combined one of two primary keywords (soccer OR football) with a second keyword (e.g., success, win, loss) using the Boolean operator AND. All utilized search terms are presented in Table 1.

Table 1. Search terms systematic review

Keyword 1	OR Keyword 1	AND Keyword 2
soccer	football	possession
soccer	football	goal
soccer	football	pass
soccer	football	success
soccer	football	shot
soccer	football	sprint
soccer	football	duel
soccer	football	corner
soccer	football	win

---

5 Actual results of the selected articles are found in the discussion section

soccer	football	lose
soccer	football	loss
soccer	football	performance indicator
soccer	football	match performance
soccer	football	indicator
soccer	football	distance
soccer	football	home advantage

For inclusion, the articles had to meet the following criteria:

- The data had to deal with performance analysis in football.
- The variables of interest were linked to success (win/loss, goals, continuance in league/tournament, league ranking and points won).
- Adult elite football was investigated.
- The study was written in English.
- The study was published in an academic journal.
- The study design was comparative or predictive or focused on home advantage in football.

It should be noted that we included studies on home advantage in this review as a separate category besides comparative and predictive studies utilizing inferential statistics. Although most of the studies on home advantage used a descriptive approach to reveal the influence of home advantage, we considered these non-inferential studies because home advantage is one of the most investigated variable regarding success factors (see Mackenzie & Cushion, 2013).

The initial search revealed 19,161 articles (Web of Science [Core and Medline]: 9,706; Scopus: 6,038; PubMed: 3,417). After excluding the duplicates 10,833 articles remained. The articles were screened based on an assessment of both the title and the abstract. All articles without a focus on the investigation and analysis of data on the conditions of competition results in elite adult football were excluded. In total, 185 articles were relevant for this review. These articles were read in detail and assessed for relevance and quality. Articles which did not meet the criteria were excluded. After this step, 53 articles remained. Subsequently, the literature references of these 53 articles were screened for more articles meeting the criteria. Fifteen additional articles were identified. Finally, 68 articles were included in the review (Figure 1).

Then, the articles that met the inclusion criteria were indexed, and each article was summarized. The summaries comprised the study purpose and design, methods of data collection and analysis, and key findings. This enables an overview and comparison of the articles and allows an assessment of the current state of research on performance indicators in football.
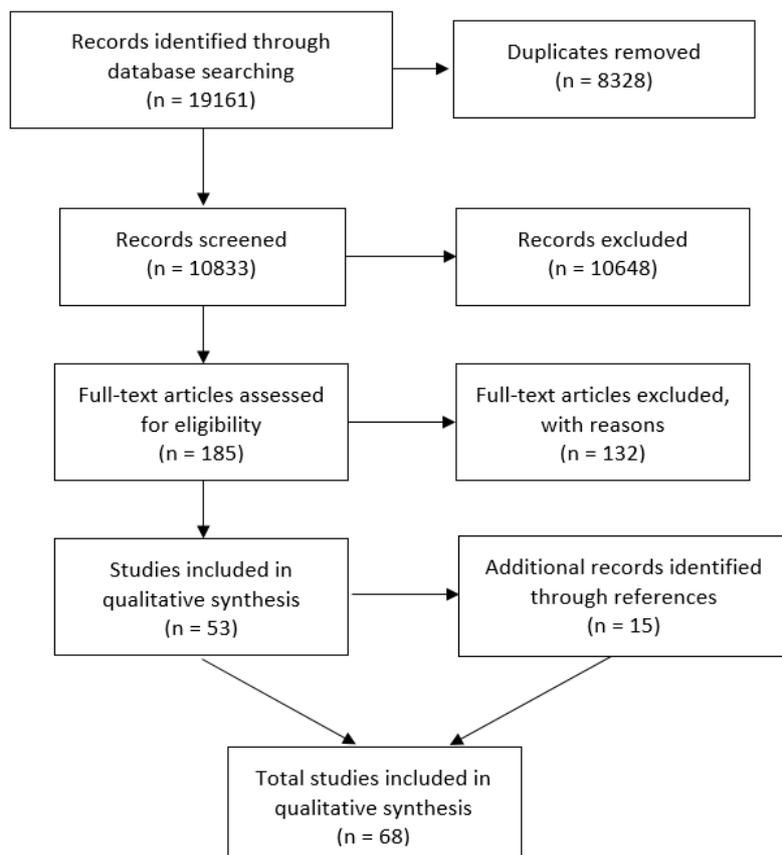
Figure 1. Flow diagram of this systematic review (based on Moher et al., 2009)

2.4. Results

The identified articles were published between 1986 and 2016, covering a time span of 31 years. More than half of the articles (exact 61.8 %; 42 articles) were published within the last seven years (2010-2016) of the searched time period, indicating that this field of research has recently gained momentum.

To organize the identified analyses, the articles were categorized following a system used by Sarmento et al. (2014) and Marcelino et al. (2011). In the first step the articles were assigned to predictive (e.g., Carmichael & Thomas, 2005; Mechtel et al., 2011), comparative (e.g., Armatas, Yiannakos, Papadopoulou et al., 2009) or home advantage (HA) analyses (e.g., Lago et al., 2016). In the second step articles were assigned to one of the three types of analysis from

above according to different operationalization of success (i.e., win/loss, goals, continuance in league/tournament, league ranking, and points won) (see Table 2).

Table 2. Number of articles in each category.

		Variables of interest					
Design	win / loss	goal difference	goals	league / tournament ranking	points	continuance in league / tournament	Row total
Comparative	7	2	1	9	1	2	22
Predictive	14	5	7	3	3		32
Total*[6]	21	7	8	12	4	2	54
Home advantage	20						20

* Multiple responses possible

Of the articles, 30 were predictive analyses, 22 were comparative analyses, and 20 focused on the analysis of home advantage. One of the articles (Oberstone, 2009) covers both types of analyses (predictive and comparative). In total, 21 articles over all three types of analysis utilized "win/loss" as the success variable. "Goal difference" was used by seven articles, "goals" by eight, "league/tournament ranking" by 12, "points" by four and "continuance in league/tournament" by two.

2.5. Discussion

In the following section, methods and major results of the identified articles will be presented within the three different categories of type of analysis. Finally, all findings will be summarized and the most frequent and significant variables regarding success factors in football will be discussed.

---

[6] Oberstone (2009) used comparative and predictive methods; Mechtel et al. (2011) used win/loss and goal difference; Collet (2013) used win/loss and points; Carmichael and Thomas (2005) used predictive methods and home advantage; Armatas, Yiannakos, Zaggelidis et al. (2009) used comparative methods and home advantage; Lago et al. (2016) used predictive methods and home advantage.

2.6. Comparative analyses

In seven of the 21 comparative analyses researchers compared wins and losses. In three of the seven papers draws were also included, and in one instance the percentage of wins was considered alongside wins and losses (see Table 3). In the three papers that compared only wins and losses (Broich et al., 2014; Kapidžić et al., 2010; Szwarc, 2007) the authors tried to find variables that explain differences between winners and losers. Broich et al. (2014) identified goal efficiency (number of goals divided by the number of shots), shots, passes and ball contacts as the most important team parameter for winning. Efficiency was also analyzed by Szwarc (2007). He showed that players of winning teams are more efficient than their opponents. As a result of the small sample (seven matches) only shots on goal ($p < 0.05$) and shots defended by goalkeeper ($p < 0.01$) differed significantly between winners and losers. Kapidžić et al. (2010) did not analyze efficiency but they also found that the numbers of shots within 16 meters ($p < 0.05$) and accurate passes ($p < 0.01$) are significant indicators for winning teams at the European Championship in 2008. Winners also scored more goals than losing teams in the Championship. Three more papers investigated the differences between wins, losses and draws (Armatas, Yiannakos, Papadopoulou et al., 2009; Janković, Leontijević, Pašić et al., 2011; Ruiz-Ruiz et al., 2013). These studies reported various significant differences between winning, drawing and losing teams. Winners have more entries into the penalty area ($p < 0.01$) (Ruiz-Ruiz et al., 2013), more successful attacks ($p = 0.003$) and passes ($p = 0.015$) as well as a higher ball possession rate ($p = 0.001$) (Armatas, Yiannakos, Papadopoulou et al., 2009; Janković, Leontijević, Pašić et al., 2011). Armatas, Yiannakos, Papadopoulou et al. (2009) revealed that 71.4 percent of teams that scored the first goal subsequently won the match ($p < 0.05$). In contrast to the other studies, one study focused on the total winning percentage (Carron et al., 2002). Another difference is the use of group cohesion as the independent variable. The authors showed a statistically significant relationship between individual attraction to the group-task and performance with a very high effect size of 1.94 ($p < 0.05$). The higher the positive feelings of each group member to the group-task, that is, to play football successfully, the higher were the likelihood of winning.

Review of the state of research

Table 3. Comparative articles with regard to wins and losses.

Author(s)	Year	Sample	Data collection	Key findings
Carron, Bray and Eys	2002	Nine football teams in Canada	GEQ questionnaire and secondary data	Individual attraction to group-task with significant performance link (p<0.05); group-integration-task not significant; both with high effect size (1.94 und 1.16)
Szwarc	2007	Seven finals European Champions League 1997-2003	Video analysis	Efficiency of shots (p<0.05) and goalkeeper efficiency (p<0.01) significant higher in the winners; Losers significant more efficient in general defense (p<0.05) such as interrupt of action, intercepting pass with ball
Armatas, Yiannakos, Papadopoulou and Skoufas	2009	240 matches in first division of Greece 2006-2007	Video analysis	71.4% of the teams that score the first goal win the match
Kapidžić, Mejremić, Bilalić and Bečirović	2010	13 matches European Championship 2008 and 12 matches first division Bosnia and Herzegovina 2008-2009	Secondary data	European Championship: winners score more goals, and more shots on goal within penalty area. First division: winners perform more successful passes, shots on goal, goals, throw-in and offensive actions
Janković, Leontijević, Pašić and Jelušić	2011	60 matches World Cup 2010	Secondary data	Winning teams perform more successful attacks (ending with a shot) and passes than losing team and in draws; winners have more ball possession and

				pass accuracy compared to losers
Ruiz-Ruiz, Fradua, Fernandez-Garcia and Zubillaga	2013	64 matches Word Cup 2006	Video analysis	Winners perform more entries into penalty area as teams in draws and losing teams
Broich, Mester, Seifriz and Yue	2014	118 matches first division Germany 2013-2014	Secondary data	Goal efficiency, shots, passes and ball contacts (in this order) are the most important team parameters for wins

In nine of the articles the authors compared teams with different positions in the league/tournament ranking (see Table 4). Luhtanen et al. (2001) investigated the influence of offensive and defensive variables on the final ranking of the European Championships in 1996 and 2000. In 1996, interceptions and the success rate of all defensive actions showed a significant correlation ($p<0.05$) with the final ranking. In 2000, significant correlations with the ranking were found for success rate in passes ($p<0.05$) and attempts ($p<0.05$) on goal. In the other papers, different football leagues were investigated, and it was shown that better ranked teams (top-teams) need less shots for a goal than worse ranked teams (Armatas, Yiannakos, Zaggelidis et al., 2009; Lago-Ballesteros & Lago, 2010; Oberstone, 2009). This parameter corresponds to Broich et al. (2014) 'goal efficiency'. It was also found that top teams have more successful attacks, complete their offensive attacks more frequently between zero and 11 meters in front of the goal (Janković, Leontijević, Pašić et al., 2011), have more successful passes (; Janković, Leontijević, Jelušić et al., 2011; Oberstone, 2009; Rampinini et al., 2009), score more goals (Armatas, Yiannakos, Zaggelidis et al., 2009; Bekris et al., 2013; Lago-Ballesteros & Lago, 2010;), perform more crosses (Bekris et al., 2013; Oberstone, 2009), have more ball possession (Lago-Ballesteros & Lago, 2010; Rampinini et al., 2009), shoot more often on the goal (Lago-Ballesteros & Lago, 2010; Rampinini et al., 2009), have more assists (Armatas, Yiannakos, Zaggelidis et al., 2009; Lago-Ballesteros & Lago, 2010; Rampinini et al., 2009) and take more shots (Bekris et al., 2013; Lago-Ballesteros & Lago, 2010; Oberstone, 2009; Rampinini et al., 2009). The best teams in the league also perform fewer fouls (Oberstone, 2009) and allow fewer shots and crosses (Bekris et al., 2013). The worst ranked teams have fewer counter attacks, have less possession with zero to four passes and have less possession longer than 12 seconds (Tenga &

Sigmundstad, 2017). Worse teams also have more very high-intensity running, high-intensity running and total distance covered (Rampinini et al., 2009). Better teams cover more total distance with the ball and very high-intensity running with the ball (Rampinini et al., 2009). Furthermore, the top teams show a faster recovering (recapture is 1.3 to 1.7 seconds faster than mean times) of ball possession (Vogelbein et al., 2014). Obviously, top teams score more goals per match (Armatas, Yiannakos, Zaggelidis et al., 2009; Bekris et al., 2013; Lago-Ballesteros & Lago, 2010; Oberstone, 2009). The cited studies showed that a lot of factors influence success (operationalized as league ranking) in football. Overall, it appears that goal efficiency, passes and shots are the most important factors in this research area.

Table 4. Comparative articles with regard to league / tournament ranking.

Author(s)	Date	Sample	Data collection	Key findings
Luhtanen, Belinskij, Häyrinen and Vänttinen	2001	31 matches European Championship 1996 – 2000	Video analysis	Interceptions and success rate interceptions and defensive actions have highest correlation with final ranking (1996). % Successful passes and % successful goals attempts (2000)
Armatas, Yiannakos, Zaggelidis, Skoufas, Papadopoulou and Fragkos	2009	10 seasons second division in Greece	Secondary data	Top ranked less shot per goal, more goals, more shots in penalty area and more assists
Oberstone	2009	380 matches in first division England 2007-2008	Secondary data	Goals per match, number of shots, short passes, total passes, pass completion are higher for better teams; goals conceded per match and fouls are lower for better teams
Rampinini, Impellizzeri, Castagna, Coutts and Wisloff	2009	416 matches in first division Italy 2004-2005	Video analysis	Worse teams more total distance, high intensity running (>14km/h) and very high intensity running (>19km/h); Top teams more total distance with ball and high intensity running with ball, more short passes, tackles, dribbles, shots and shots on goal

Lago-Ballesteros and Lago-Peñas	2010	380 matches in first division Spain 2008-2009	Secondary data	Top teams more goals, shots and shots on goal; worse teams need more shots per goal
Janković, Leonti-jević, Jelušić, Pašić and Mićović	2011	228 matches in first division Serbia 2009-2010	Video analysis	Successful attacks (end up with a shot) and pass rate higher for top teams; top teams kick the ball more often form 0-11m to the goal
Tenga and Sig-mundstad	2011	997 goals from 1922 matches in first division in Norway 2008-2010	Video analysis	Worst teams less goals through counterattack, less possession with 0-4 passes, less possession for 12 seconds or more and less posses-sion started in the midfield
Bekris, Mylonis, Sarakinos, Gissis, Gioldasis and Sotiropoulos	2013	240 matches in first division Greece	Secondary data	Goals per match, shots, shots in penalty area, crosses and assists are higher for top teams; they conceded less shots, shots in penalty area and crosses
Vogelbein, Nopp and Hoekelmann	2014	306 matches in first division Germany	Video analysis	Top teams have a faster recovering of ball possession after losing it (defensive reaction time)

Six more studies used a comparative approach to investigate success factors operationalized differently to the articles discussed previously (see Table 5). Two papers focused on goal dif-ference (Bekris et al., 2014; Yue et al., 2014). Bekris et al. (2014) compared matches with one-goal differences (short range results) as well as matches with three-goal differences or more (wide range results). Their analysis showed that winners in wide range results have more ball possession, perform more passes, win more duels (overall and aerial), and have more shots, shots on target and a higher shot accuracy. In the short range results these differences were not found. A winner-winner comparison showed that wide range winners perform more passes, have a higher pass accuracy, more short distance shots and shots on-target. Yue et al. (2014) used a similar approach. They analyzed matches with a difference of two or more goals and matches with a difference of three or more goals. Goal efficiency, shots, passes and ball contacts were found to be the most important factors for scoring a goal (in this order). Clemente (2012) and Delgado-Bordonau et al. (2013) operationalized success as continuance in a tournament.

They compared teams with a different number of matches respectively teams that got to the semi-final. Both analyzed matches of the World Cup 2010. Clemente (2012) revealed that teams with more matches in a tournament (the successful ones) score more goals through open play, have more shots inside the penalty area and perform more passes. Delgado-Bordonau et al. (2013) showed that successful teams perform more shots on-target, have a higher efficiency and concede fewer shots. They also revealed that the first goal in the match leads to a victory for 66.7 percent in the group stage and for 81.3 percent in the knockout stage. Hughes and Franks (2005) used a new and different approach to analyze football. They normalized the data into "goals/shots per 1000 possessions" to analyze the relative importance of ball possession. The authors used this parameter to compare successful teams (getting to the quarterfinals) and unsuccessful teams (first round losers) in the 1990 World Cup. Accordingly, successful teams show a strong trend to be better in converting possession into shots on goal (no significant difference). For ball possessions with more than eight passes there is a significantly higher chance for successful teams to create a shooting opportunity ($p<0.05$). In contrast, the necessary shots for a goal increase with more passes per possession (Hughes & Franks, 2005). Hoppe et al. (2015) used the final points accumulated by each team during one season in the German Bundesliga. They analyzed the running performance with and without ball possession of the teams. Only total distance with ball possession was a significant predictor for final points ($p<0.01$). They concluded that not only running performance is important for success, but rather the relation to technical/tactical skill regarding ball possession (Hoppe et al., 2015).

Table 5. Comparative articles with regard to other operationalization of success.

Author(s)	Date	Sample	Data collection	Key findings
Hughes and Franks	2005	52 matches World Cup 1990	Secondary data	Variable of interest is goal scored; successful teams are better in converting possession into shots on goal; for possession with more than 8 passes there is a significant ($p<0.05$) better chance for successful teams to create a shooting opportunity; shots necessary for a goal increased with more passes per possession

Clemente	2012	208 matches World Cup 2010	Secondary data	Variable of interest is continuance in tournament; teams with more matches score more goals per match, through open play, from within the penalty area, and play more passes
Delgado-Bordo-nau, Domenech-Monforte, Guzmán and Mendez-Villanueva	2013	56 matches World Cup 2010	Secondary data	Variable of interest is continuance in tournament; successful teams score more goals, perform more shots, have better efficiency, conceded less goals per match, conceded less shots; during group stage, teams scoring the first goal had a 66.7% chance to win (81.3% for knockout stage)
Bekris, Gioldasis, Gissis, Komsis and Alipasali	2014	64 matches European Leagues 2013-2014	Video analysis	Variable of interest is goal difference; wide range results: winners have better performance in duels (aerial and overall), ball possession, passes, shots, shot accuracy, shots on goal; comparison of wide range with short range winners: wide range winners perform more passes, shots, and have a higher passing accuracy and more shots on goal
Yue, Broich and Mester	2014	74 matches in first division Germany 2011	Secondary data	Variable of interest is goal difference; in matches with a goal difference of 2 and more or with 3 and more the most important factors are efficiency, shots, passes and ball contacts (in this order); correlation of this four factors

				with number of goals shows the same result
Hoppe, Slomka, Baumgart, Weber and Freiwald	2015	306 matches in first division Germany 2012/13	Secondary data	Variable of interest is points accumulated; total distance with ball possession only significant predictor for final points accumulated (p<0.01)

2.7. Predictive analyses

Fourteen of the predictive analyses focused on differences between wins, draws and losses (two of these papers considered two groups: winners and non-winners) (see Table 6). Four of these papers used a discriminant analysis to reveal the most discriminating factors (Castellano et al., 2012; Lago et al., 2010; Lago et al., 2011; Moura et al., 2014). Shots on goal was a discriminant factor in all four studies. Crosses, match location and ball possession (Lago et al., 2010; Lago et al., 2011) as well as the quality of the opponent (similar to strength or team ability) (Lago et al., 2011) were other identified factors. Collet (2013) and Harrop and Nevill (2017) used a regression analysis/model and showed that higher pass accuracy is a good predictor for success. More shots, fewer passes, fewer dribbling, and match location are further predictors (Harrop & Nevill, 2017). Collet (2013) investigated the influence of possession on success and showed that possession is not as relevant as assumed. If the strength of a team is controlled, the influence of possession on success will range from -5.7% (in German Bundesliga; significant (p<0.05)) to +1.8% (all national teams; not significant). The fact that possession has a potential negative link to success may be worth further examination. Efficiency measures seem to be better predictors for success (Collet, 2013; Broich et al., 2014; Delgado-Bordonau et al., 2013; Szwarc, 2007; Yue et al., 2014). Liu et al. (2015), Liu et al. (2016) and Mao et al. (2016) used cumulative logistic regression in a generalized linear model. They also divided the sample into close matches and unbalanced matches (a cluster analysis based on the goal difference was used) with a cluster analysis and cut-off values. In past research it appeared to be more likely in close matches that both teams play at their best (Liu et al., 2015; Vaz et al., 2010). They showed that shots on goal, shot accuracy, tackles and aerial advantage have positive effects on winning (Liu et al., 2015; Mao et al., 2016). Liu et al. (2016) also investigated the within-team effects (changes in team values between matches) and between-team effects (differences between average team values over all matches). Shots on target and total shots have positive within-team effects on winning. Game location showed a small positive within-team effect. Ball possession

showed a small negative within-team effect but also a small positive between-team effect. Within-team effects varied depending on strength of team and opponent (Liu et al., 2016).

Gómez et al. (2012) used a factor analysis with several factors and the zone of the pitch. For the zone of the pitch they divided the field into five zones from goal to goal and into three to five subzones in each of these zones. They identified four factors. All factors are highest for winners. The best discrimination is given for ball recovery in zone two (2.1, 2.2 and 2.3) (penalty zone to center circle) and offensive actions with long passing sequences in zone 5.1 (six-yard box) and 5.2 (within penalty zone). Bar-Eli et al. (2006) and Mechtel et al. (2011) investigated the impact of a player's dismissal. Both found out that a sending-off decreases (sanctioned team) respectively increases (opponent) the chance of winning. Mechtel et al. (2011) also identified strength (points earned in the last three seasons) and home advantage as success factors. Torgler (2004) applied an economic win function to determine the influences on winning or not winning during the FIFA World Cup 2002. He showed that a higher number of shots on goal leads to a higher probability to win. He also revealed the negative effect of a player's dismissal. Hosting the tournament was a strong advantage as well. It increases the chance of winning by 45 percentage points (Torgler, 2004). Hanau et al. (2014) investigated the difference between the expected outcome of a football match and the actual outcome. They found out that the actual outcome is determined by the standing in the last season and home advantage.

Table 6. Predictive analyses with regard to wins and losses.

Author(s)	Date	Sample	Data collection	Key findings
Torgler	2004	63 matches World Cup 2002	Secondary data	Higher number of shots on goal higher probability to win than not to win; dismissal has strong negative effect; hosting the tournament is a strong advantage
Bar-Eli, Tenenbaum and Geister	2006	743 matches in first division Germany 1963-2004	Secondary data	Chance of winning decreases after a red card dependent on match status and match location
Lago-Penas, Lago-Ballesteros, Dellal and Gomez	2010	380 matches in first division	Secondary data	Shots, shots on goal, effectiveness, assists, crosses, conceded crosses, possession and match

		Spain 2008-2009		location discriminate best between win, draw and loss
Lago-Penas, Lago-Ballesteros and Rey	2011	288 matches European Champions' League group-stage 2007-2010	Secondary data	Winners perform more shots, better effectiveness, more passes, higher possession and receive less cards; shots on goal, crosses, possession, match location and quality of opponent discriminate best
Mechtel, Baker, Brandle and Vetter	2011	2962 matches in first division Germany 1999-2009	Secondary data	Players dismissal increase chance of winning for opponent; team strength (overall and at home) increase chance of winning
Castellano, Casamichana and Lago	2012	177 matches World Cup 2002-2010	Secondary data	Shots, shots on goal, shots received and shots on goal received discriminate best
Gómez, Gómez-Lopez, Lago and Sampaio	2012	1900 matches in first division Spain 2003-2008	Secondary data	Field subdivided in 19 zones; 7 variables recorded; factor analysis revealed four factors (First: Turnovers in Zone 5.2 and Crosses in zone 4; Second: Goals in zone 5.1, Shots in zone 5.1, Turnovers in zone 4 and Ball recover in zone 1; Third: Goals in zone 5.2, Shots in zone 5.2 and Ball recover in zone 1; Fourth: Turnovers in zone 5.1), factors highest for winners; draw data closer to lose
Collet	2013	6172 matches from several leagues and tournaments	Secondary data	More time with ball leads to more points and goals; passes and pass accuracy correlate with points and goals; more points on smaller pass to shots on goal relation; if team

				strength is controlled negative effect for possession; pass and shot accuracy are better predictors
Harrop and Nevill	2014	46 matches in second division England 2012-2013	Secondary data	Less passes (p=0.006), more successful passes (p=0.042), more shots (p=0.027), less dribbles (p=0.018) and the match location (p=0.044) are significant in prediction of success; passes (p=0.000), successful passes (p=0.001), and passes in opposition half (p=0.005) are different between wins, draws and losses
Moura, Martins and Cunha	2014	96 matches in group stage World Cup 2006	Secondary data	Cluster analysis to generate two groups of data; 70.3% of the winning team were classified into the same group; shots, shots on goal and possession discriminate best the winning teams
Hanau, Wicker and Soebbing	2015	306 matches in first division Germany 2010-2011	Secondary data	Actual winning is influenced by difference in ranking last year and home match
Liu, Gomez, Lago-Penas and Sampaio	2015	48 matches World Cup 2014	Secondary data	Shots, Shots on goal, Shots from Counter Attack, Shot from Inside Area, Ball Possession, Short Pass, Average Pass Streak, Aerial Advantage and Tackle clear positive effects on winning, Shots Blocked, Cross, Dribble and Red Card negative relationship to winning

Liu, Hopkins and Gomez	2016	320 matches in first division Spain 2012/13	Secondary data	Shots on target and total shots have positive within team effect to winning; ball possession: small negative within-team effect but a small positive between-team effect; Game location showed a small positive within-team effect; Within-team effects varied depending on the strength of team and opposition.
Mao, Peng, Liu and Gomez	2016	480 matches in first division China 2014-2015	Secondary data	Shots on goal (positive), shot accuracy (positive), cross accuracy (trivial), tackle (trivial) and yellow cards (trivial) have effects on winning

The second most frequent kind of predictive analyses are studies that used goal scoring as the indicator of success (see Table 7). Pollard and Reep (1997) developed a quantitative variable, called the 'yield', defined as the probability of a goal being scored minus the probability of one being conceded. The yield for the penalty area as starting zone of ball possession and open play is 78.3 (per 1000 possessions you can expect 78.3 more goals scored than goals conceded). They also found that open play always has a higher yield than set play (Pollard & Reep, 1997). Carmichael and Thomas (2005) established a match-based production function. They found that shots on goal, shots that hit woodwork, tackles, own goals and free kicks are significant predictive factors ($p<0.05$) for the home teams. Kapidžić et al. (2009) also identified shots on goal as a significant predictor for goal scoring ($p=0.027$). Wright et al. (2011) postulated position of attempt, goal keepers' position and type of shot as the three predictors for goal scoring. Tenga, Holme et al. (2010) and Tenga, Ronglan et al. (2010) used the same data set with different methods for their analysis. Both papers showed that counter attacks are more effective than elaborated attacks in producing goals. Grund (2012) used a network analysis to identify success factors. He revealed that networks with high intensity and low centralization have a better performance. An increased passing rate lead to a better performance in this study (Grund, 2012).

Table 7. Predictive analyses with regard to goal scoring.

Author(s)	Date	Sample	Data collection	Key findings
Pollard and Reep	1997	22 matches World Cup 1986	Video analysis	Calculation of "yield" (probability of a goal being scored, minus the probability of one being concede); starting zone of ball possession, open/set play and playing strategy as factors for the yield calculation; open play higher yield as set play; the closer it gets to the opponent goal the higher the yield
Carmichael and Thomas	2005	380 matches in fist division England 1997-1998	Secondary data	Attacking play seems more important for home team and defensive play for away teams; shots on goal, tackles, free kicks and cards given are important factors
Kapidžić, Bećirović and Imamović	2009	31 matches European Championship 2008	Secondary data	Shots within penalty area are the only significant single predictor (p=0.003), shots on goal, shots off goal, shots blocked, pass completion, long, middle and short passes and completion explained 36% of the variance
Tenga, Holme, Ronglan and Bahr	2010	163 matches in first division Norway 2004	Video analysis	More goals during counter attacks; counter attacks better than elaborate attacks; attacks starting in the last third better as first third; long possession is better than short possession
Tenga, Ronglan and Bahr	2010	163 matches in first division Norway 2004	Video analysis	Counter attacks better than elaborate attacks; scoring

				opportunities and score box possessions (shooting opportunities) can be used as a proxy for goals scored under certain circumstances
Wright, Atkins, Polman, Jones and Sargeson	2011	167 goals in first division England 2010-2011	Video analysis	Three factors are significant predictors of goal success (p<0.05), position of attempt, goal keepers' position and type of shoot
Grund	2012	76 matches in first division England 2006-2008; 283,259 passes to create network	Secondary data	A clear network intensity effect is found. Increases in the passing rate lead to increased team performance. a clear network centralization effect is present; Increases in the centralization of team play lead to decreased performance

In the last group of predictive analyses three variables of interest were collected (see Table 8). The most frequent variable is goal difference as utilized in five papers (Carmichael et al., 2000; García-Rubio et al., 2017; Mechtel et al., 2011; Papahristodoulou, 2007). In all articles match location is positively linked to goal difference. Quality of the opponent was also identified as a significant predictor (p<0.05) (García-Rubio et al., 2017; Mechtel et al., 2011; Papahristodoulou, 2007). Moreover, Carmichael et al. (2000) showed that passes, tackles, interceptions, clearances, blocks, interceptions, free kicks and ball caught by goalkeeper are significant predictors for a positive goal difference(p<0.05). A red card was associated with a negative goal difference (Carmichael et al., 2000; Mechtel et al., 2011; Papahristodoulou, 2007). García-Rubio et al. (2017) showed that scoring first is the strongest predictor for a positive goal difference. Lago et al. (2016) used a tree analysis to determine the effects of scoring first on the outcome of a match. They showed that the first scoring team scored 1.88 goals more than their opponent on average. This is influenced by the quality of the teams and the match period in which the first goal was scored (Lago et al., 2016)., Hall et al. (2016), Kringstad and Olsen (2016) and Oberstone (2009) investigated relevant factors for the league ranking in a predictive design. Hall et al. (2016) focused on the relationship between payroll and performance. They found

that there is a higher winning probability of 0.614 for 50% more spending in payroll. The top level is more sensitive to spending. Oberstone (2009) developed a regression model to predict the league ranking. He revealed six variables which are sufficient for predicting the league ranking (in terms of points earned). These six variables are the percentage of goals to shot (goals divided by shots), the percentage of goals outside penalty area (goals from outside penalty area divided by goals within penalty area), ratio of short to long passes, total crosses, average goals conceded per match and yellow cards. Kringstad and Olsen (2016) studied budgeted revenue and success. They showed that budgeted revenues are a significant factor (p<0.05) but only for the bottom-half of the teams and not for the top-half of the teams. The remaining three papers focused on points as the variable of interest. Lago (2007) defined performance as shots performed minus shots conceded and found that this is a predictor for more points. Furthermore, he showed that the higher the FIFA ranking is, the higher the chance to win. Collet (2013) focused on ball possession. His result was that more time with the ball leads to more points and goals, but if it is controlled by team strength a negative effect for possession can be observed. Passes and shot accuracy turned out to be better predictors for points. Coates et al. (2016) investigated the relationship between salary structure and success. They revealed that salary inequality has a negative effect on success, but the wage bill of a team has a positive relationship with success by a similar amount. This result support the cohesion theory (Coates et al., 2016).

Table 8. Predictive analyses with regard to other operationalization of success.

Author(s)	Date	Sample	Data collection	Key findings
Carmichael, Thomas and Ward	2000	380 matches in first division England 1997-1998	Secondary data	Variable of interest is goal difference; fixed effects for relative performance of teams; match location, differences in successful passes, passes in penalty area, tackles, clearances, blocks, interceptions, free kicks, red card and ball caught by goalkeeper are significant predictors (p<0.05)
Hall, Szymanski and Zimbalist	2002	39 teams in the first four divisions England 1974-1999	Secondary data	Variable of interest is league ranking; 50% more spending in payroll leads to 0,614 higher winning probability; Granger causality from higher payrolls to

				better performance cannot be rejected
Lago-Penas	2007	64 matches World Cup 2006 Germany	Secondary data	Variable of interest is points earned; performance (shots minus shots conceded) is a predictor for more points; the higher the FIFA-Ranking, the higher the chance to win
Papahristodoulou	2008	806 matches European Champions League 2001-2007	Secondary data	Variable of interest is goal difference; goals are an effect of shooting; red cars are negative for winning probability; match location important for winning probability
Oberstone	2009	380 matches in first division England 2007-2008	Secondary data	Variable of interest is league ranking; % goals to shot, % goals outside penalty area, proportion (ratio) short/long passes, total crosses, average goals conceded per match and yellow cards are sufficient to predict league ranking/point earned
Mechtel, Baker, Brandle, and Vetter	2011	2962 matches in first division Germany 1999-2009	Secondary data	Variable of interest is goal difference; players' dismissal increase chance of winning for opponent; team strength (overall and at home) increase chance of winning
Collet	2013	6172 matches from several leagues and tournaments	Secondary data	Variable of interest is points earned; higher ball possession leads to more points and goals; passes and pass accuracy correlate with points and goals; more points with lower pass-to-shots-on-goal-ratio (how many passes before a shot); if team strength is

				controlled there is a negative effect for possession; pass and shot accuracy are better predictors
Garcia-Rubio, Gomez, Lago-Penas and Ibanez	2015	475 matches European Champions League 2009-2013	Secondary data	Variable of interest is points earned; Positive influence of match location, scoring first and quality of opposition in match outcome, scoring first strongest predictor then match location, then quality of opposition, Structural coefficient significant underlines that teams that score first achieve more shots on goal in both stages of competition ($p<0.01$)
Coates, Frick and Jewell	2016	138 team year observations in first division USA 2005-2013	Secondary data	Variable of interest is points earned; Negative relationship between salary inequality and team success; the best-fit model suggests that increasing salary inequality and the team wage bill work in opposite directions by similar magnitudes
Kringstad and Olsen	2016	720 matches in first division Norway 2011-2013	Secondary data	Variable of interest is league ranking; Budgeted revenues are a significant factor of success for the bottom-half teams but not for the top-half teams ($p<0.05$); money could be a significant driver of success, but only to a certain extent
Lago-Penas, Gomez-Ruano, Megias-Navarro and Pollard	2016	1826 matches in France, Italy, Spain, England and	Secondary data	Three independent variables were significant factors on the final outcome: the quality of the opposition ($p<0.001$), the minute

		Germany 2014/15		in which the first goal is scored (p<0.01) and the team scoring first (p<0.001); teams that scored first scored 1.88 goals more than the opponent

2.8. Analyses of home advantage

The review of predictive analyses already showed that match location (home advantage) is an important factor in explaining success in football (Bar-Eli et al., 2006; Carmichael & Thomas, 2005; Lago et al., 2010; Lago et al., 2011; Liu et al., 2016; Mechtel et al., 2011; Papahristodoulou, 2007; Torgler, 2004). Seventeen papers that focused mainly on match locations specifically home advantage were identified in this the review (see Table 9). In one of these papers (Carmichael & Thomas, 2005) further factors related to success, besides home advantage, were also investigated. The first analysis of home advantage in football was done by Pollard (1986). He investigated different team sports including the first four football divisions in England from 1888 to 1984. There was very little variation between 85 seasons (between 1939 and 1945 there were no official seasons due to World War II). The points won by the home team differed between 62.5 percent and 67.9 percent. Clarke and Norman (1995) provided an approach to quantify team ability and home advantage at a team level due to the influence of the quality of opponent (team ability or strength). This approach was also used by other authors to define home advantage for a team (Lago et al., 2011; Mechtel et al., 2011; Papahristodoulou, 2007). Clarke and Norman (1995) stated that it is necessary to consider difference in ability to calculate home advantage. In their research the home advantage relating to goals differed from year to year and between teams. The average home advantage between 1981 and 1990 in England resulted in 0.528 goals per match. Another result is that team ability is more important than home advantage (Clarke & Norman, 1995). Overall, home advantage explains around 60 percent with some variations (Armatas & Pollard, 2014; Goumas, 2014a; Goumas, 2014b; Goumas, 2014c; Goumas, 2015; Lago & Lago-Ballesteros, 2011; Pollard, 2006; Pollard, 2008; Pollard & Gómez, 2009; Pollard & Pollard, 2005; Pollard et al., 2008; Poulter, 2009; Saavedra García et al., 2015; Sánchez et al., 2009; Seçkin & Pollard, 2008; Thomas et al., 2004) (see also Table 9). Before the 1980s, the explaining percentage of home advantage was moderately higher (Thomas et al., 2004). Saavedra García et al. (2015) investigated home advantage in the first division in Spain between 1928 and 2011. Home teams won 70.8 percent of the points for the period when 2 points were awarded for a victory and 56.7 percent when three points were

awarded for a victory. Lago et al. (2016) showed a consistent home advantage for all five major leagues in Europe (France, Italy, Spain, England and Germany) for the season 2014/15. Home teams won between 56.47 percent (Italy) and 61.84 (Germany) of the awarded points for a victory.

Lago and Lago-Ballesteros (2011) investigated the variables that discriminate best (discriminant value ≥|.30|) between home and away teams. Home teams score more goals, perform more crosses, more passes, have more ball possession and commit more fouls. Away teams show more losses of possession and gather more yellow cards. Armatas and Pollard (2014) found shots, clearances, headed shots, corners and saves to have the highest effect size for match variables between home and away teams. Goumas (2015) analyzed home advantage on a team level adjusted for team ability (operationalized by UEFA ranking points). Home advantage did not vary between teams despite a home advantage of 73% for Arsenal London and a home advantage of 58% for Inter Milan. Away disadvantage varied between teams ranging from 45% (F.C. Barcelona) to 68% (Olympiacos F.C.). There was also a tendency that teams with a higher home advantage had lower away disadvantage. Home advantage and away disadvantage differed significant between countries from 70% (English teams) to 52% (Turkish teams) (p=0.01) (Goumas, 2015). The major causes for home advantage discussed are crowd support, travel fatigue, familiarity, territoriality, referee bias, special tactics, rule factors and psychological factors as well as the interaction of these (Pollard, 1986; Pollard, 2006; Pollard, 2008).

Table 9. Analyses of home advantage.

Author(s)	Date	Sample	Key findings
Pollard	1986	58,123 matches in England 1888-1984	Little variation between the centuries and divisions; no difference between two- and three-point system; home advantage in percent of obtained point is around 64%; local derbies show significant lower home advantage (p<0.01)
Clarke and Norman	1995	20,306 matches in England 1981-1991	Home advantage in terms of goals per match; team ability included; home advantage 0.528 goals per match in average
Thomas, Reeves, and Davies	2004	7834 matches in England 1985-2003	Slightly lower home advantage in recent years (2%-5% lower); home advantage still stable phenomenon

Carmichael and Thomas	2005	380 matches in England 1997-1998	57% of the points obtained at home; home teams won 48% of the matches
Pollard and Pollard	2005	Over 70,000 matches in England 1888-2003	Home advantage was highest in the early years of each league; home advantage seems stable around 60% of the point obtained at home
Pollard	2006	89813 matches around the world 1997-2003	Home advantage is found in all big leagues in the world; in the Balkan countries and in the Andean region home advantage is much higher; home advantage varies from 48.87 (Andorra) to 78.95 (Bosnia) around the world
Pollard, Silva, and Medeiros	2008	2326 matches in Brazil 2003-2007	Average home advantage 65%, calculated by the points obtained at home; north and south teams have a higher advantage
Seckin and Pollard	2008	3672 matches in Turkey 1994-2006	61.5% average home advantage; calculated by the points obtained at home; local derbies (matches in Istanbul) show lower home advantage
Armatas, Yiannakos, Papadopoulou, and Skoufas	2009	240 matches in Greece 2006-2007	47.3% of the matches are won by home team, 26.3% draws and 26.4% won by away team
Pollard and Gomez	2009	81,185 matches in France, Italy, Spain and Portugal 1928 (or beginning) - 2007	About 66% average home advantage of the points obtained at home; recent general decline in home advantage since the 1980s; home advantage in Spain highest with an average of 69%; increased home advantage for teams from islands; lower home advantage in capital cities
Poulter	2009	808 matches in European Champions League 2001-2007	Home teams won 67.7% of the matches; home team is 1.98 times more likely to score in match than the away team; home teams perform more shots, shots on goal and corners; away teams have more fouls committed, offside and cards
Sanchez, Garcia-Calvo, Leo, Pollard, and Gomez	2009	20,992 matches in Spain 1980-2007	About 66% average home advantage calculated by the points obtained at home; slightly significant decrease of home advantage after introduction of the 3-point system (p=002)

Lago-Penas and Lago-Balles-teros	2011	380 matches in Spain 2008-2009	61.95% victories for home and 38.05% victories for guests (draws excluded); 4 groups according to league ranking; inferior teams benefit less from home advantage than superior teams
Armatas and Pollard	2014	2160 matches in Greece 1994-2011	About 65% average home advantage calculated by the points obtained at home; shots, clearances, headed shots, corners and saves have highest effect size for match variables between home and away teams
Goumas	2014a	1384 matches in European Champions League and Europa League	58.8% (CL) and 58.0 (EL) home advantage in terms of goals scored; in terms of competition points gained in the group stage home advantage was 57.8% in the CL and 59.2% in the EL; crowd density is important in influencing referee bias; more yellow cards against away teams
Goumas	2014b	765 matches in Australia 2005-2012	57.7% average home advantage of the points obtained at home and 56.5% home advantage in terms of goals scored; home advantage increases with increasing time zones crossed by away teams
Goumas	2014c	3277 matches in Europe, Asia, South America and Africa 2007-2013	59% (Europe), 60% (Asia), 63% (South America) and 70% (Africa) home advantage in terms of goals scored; absolute distance travelled, and time zones crossed associated with poorer match performance
Saavedra García; Gutiérrez Aguilar, Fernández Romero and Sa Marques	2015	22015 matches in Spain 1928-2011	70.8% average home advantage for the period when 2 points were awarded for a victory; 56.7% average home advantage when three points were awarded for a victory
Goumas	2015	1058 matches European Champions League 2003-2013	Home advantage measured on a team level; home advantage did not vary between teams despite 58% for Inter Milan and 73% for Arsenal London; away disadvantage vary between teams significantly (p<0.05); tendency of higher home advantage and lower away disadvantage; home advantage differs significant between countries 70% English teams to 52% Turkish teams (p=0.01)

Lago-Penas, Gomez-Ruano, Megias-Navarro and Pollard	2016	1826 matches in France, Italy, Spain, England and Germany 2014/15	Results showed that home teams scored first in 57.8 % of matches and went on the obtain 84.85% of points; Away team scored first, they obtained only 76.25% of subsequent points

2.9.   Integrative discussion

The aim of this study was to review performance analyses in adult male football in order to identify success factors and utilized methods. The review revealed that there is an extensive and growing body of performance analyses literature in football. In contrast to early studies that were often based on descriptive designs (Reep & Benjamin, 1968), analyses with predictive designs, explaining more and more success factors (Collet, 2013; Lago et al., 2011; Liu et al., 2015;), have gained momentum in recent years. The most frequently studied variables were shots (27 times)/shots on goal (23 times) followed by passes (20 times). Overall, 76 different variables were investigated in the reviewed papers. Based on the results in the papers, the most influential variables are efficiency (Broich et al., 2014; Delgado-Bordonau et al., 2013; Liu et al., 2015), shots on goal (Lago et al., 2011; Mao et al., 2016), possession (Rampinini et al., 2009), pass accuracy/successful passes (Janković, Leontijević, Pašić et al., 2011; Luhtanen et al., 2001), quality of opponent (Lago et al., 2016; Mechtel et al., 2011; Papahristodoulou, 2007), and match location (García-Rubio et al., 2017; Lago et al., 2011; Pollard, 2006)[7].

It became apparent that performance in football depends on a high number of variables. For example, Oberstone (2009) investigated 24 different variables. Using a 6-variable regression (percentage of goals to shots, percentage of goals scored outside of box, ratio of short/long passes, total crosses, average goals conceded per match and yellow cards) he predicted the points earned by English football teams in the 2007/2008 season. The fit delivered an $R^2=0.990$ (p<0.0000) indicating strong evidence for his model. Similarly, Kapidžić et al. (2010) investigated 21 variables in the first division in Bosnia and Herzegovina 2008/2009 (12 matches) and in the 2008 European Championship (13 matches). While in the first division 13 variables (e.g., shots, passes, and offensive structure) significantly discriminate between winners and losers (p<0.05), in the European Championship only three variables were significant (shots on goal,

---

[7] The most influential variables were assessed based on specific evidences the authors provided. For example, Broich et al. (2014) defined the parameter q (relative size of the difference) and calculated a highly significant value of 103.4 for efficiency, which is more than four times higher than the value of the second most important variable (number of shots). To quantify the importance and influence of success factors, a meta-analytical approach would be needed. However, this goes beyond the scope of this paper.

number of goals scored within penalty area and number of goals scored outside penalty area) (p<0.05). Although both studies considered many variables, it were the obvious variables such as shots and goals that became significant, explaining only little of the underlying mechanisms of success in football. Liu et al. (2015) and Mao et al. (2016) studied very similar variables in two different samples. Shot on target and tackle were the only two discriminating variables in both studies. Other variables had no clear effect or the effect depended on the context (Liu et al., 2015, Mao et al., 2016). Based on these results, it seems that not many success factors in football are stable over different contexts and samples. It should be noted, however, that an exclusive focus on statistical data (e.g., shots, possession) will probably be not sufficient to explain these mechanisms. A more sophisticated approach is needed to reveal these mechanisms. This includes more variables and the use of more complex statistical approaches such as ordered logit regressions to determine the influence of these variables. Also, the inclusion of qualitative variables e.g., self-perception and social perception or the evaluation of motivation can help to reveal the nature of performance. A third area of investigation should be more player centric such as questionnaires e.g., about group cohesiveness or personality traits.

Moreover, the review revealed that to date many different types of matches and settings have come into the focus of researchers, providing a more holistic view on success factors in football. Regarding comparative and predictive analyses, 34 articles focused on league matches, 13 on cup matches for national teams and six on cup matches for clubs. Especially studies that integrate different types of matches and settings provide useful insights allowing for generalizable statements. For example, Collet (2013) analyzed more than 6,000 matches including league matches from England, Italy, France and Germany, matches from the European Champions League and the Europe League as well as national matches from Europe, America, Africa and Asia. In this way, he found that in the leagues pass accuracy and shot accuracy are more important for success than ball possession, in contrast to the assumptions of many scholars and professionals (for Germany one percent more possession even leads to a winning probability that is reduced by 5.7 percent). Also, Lago et al. (2016) studied over 1,800 matches in the five top leagues across Europe. They could show that scoring first is a crucial part of winning a match. In total, 27 studies chose a design that comprised an international comparison, while among the studies that focused on one nation, England showed to be the most studied country in football (11 articles), followed by Germany (7 articles) and Spain (7 articles) (see

Table 10).

Review of the state of research

Table 10. Design and country of the reviewed studies.

Country of sample	Study design			Total
	Comparative	Predictive	Home Advantage	
Australia			1	1
Brazil			1	1
Canada	1			1
England*	1	7	5	13
Germany	4	3		7
Greece*	3		2	5
International*	9	12	7	28
Italy	1			1
Norway	1	3		4
Serbia	1			1
Spain	1	3	3	7
Turkey			1	1
USA		1		1
China		1		1
Total[8]	22	30	20	72

* Multiple responses

Methodologically, the review showed that in recent years new ways of statistical analyses were introduced. Lago et al. (2010) were the first authors who used a discriminant analysis to identify differences between winners and losers. Moura et al. (2014) combined this approach with a factor analysis. They investigated 14 variables and performed a factor analysis. Subsequently, a cluster analysis was used to classify the teams into two groups. Finally, they showed that 70.3 percent of the winning teams were classified into the same group (67.8 percent for drawing and losing teams). Shots, shots on goal, playing time with ball possession and percentage of ball

---

[8] Oberstone (2009) used comparative and predictive methods; Carmichael and Thomas (2005) used predictive methods and home advantage; Armatas, Yiannakos, Papadopoulou et al. (2009) used comparative methods and home advantage; Lago et al. (2016) used predictive methods and home advantage

possession were the most important variables to discriminate between winning teams and drawing or losing teams in this study. Liu et al. (2015) used a cluster analysis to identify only close matches. This approach has the advantage that both teams give probably their best and do not lean back because the match is already decided (Liu et al., 2015; Vaz et al., 2010). The concept of close and unbalanced matches also improved the analysis of success factors in football (Broich et al., 2014; Liu et al., 2015). Close matches are defined by a small goal difference. In unbalanced matches one team dominates the other team in terms of goal difference very obviously (Gómez et al., 2014; Gómez et al., 2017; Lupo et al., 2014; Lupo & Tessitore, 2016; Vaz et al., 2010). This concept was first introduced in a discrimination study about rugby in 2010 (Vaz et al., 2010) and is widely used since then (Broich et al., 2014; Gómez et al., 2014; Gómez et al., 2017; Liu et al., 2015;; Lupo et al., 2014; Lupo & Tessitore, 2016; Vaz et al., 2010)

However, most researchers (comparative and predictive design) used a form of regression analysis (22 studies). Discriminate analysis (six studies) and ANOVA (five studies) are the second and third most frequently used statistical methods. For example, Mechtel et al. (2011) and Collet (2013) used an ordered logit regression to identify the influence of a dismissal respective ball possession. An advantage of this method is that it controls for other variables and to investigate a goal-based and result-based approach. Liu et al. (2015) and Mao et al. (2016) used a generalized linear model. First, they ran a cluster analysis to define cut-off values (see above). Then they applied a cumulative logistic regression to predict winning probabilities. Afterwards they employed non-clinical magnitude-based inferences to evaluate the true effect of the variable (Liu et al., 2015; Mao et al., 2016). This approach allows a more realistic and intuitive interpretation of effects (Hopkins et al., 2009). Since much of current research is still descriptive or comparative, these two approaches are promising with regard to providing new, valuable insights to performance in football.

Finally, a crucial point that was found is sample size. Many studies, such as Kapidžić et al. (2010) who analyzed 25 matches, rely on small sample sizes. Of the reviewed papers, the sample sizes varied from seven matches (Szwarc, 2007) to 89,813 matches (Pollard, 2006). In total, only 28 papers analyzed all matches of a whole or several seasons. It appears that many studies lack sample sizes that are adequate to produce generalizable results.

2.10. Practical implications

A critical question is how the results can support football coaches and their staff. Based on the findings of this review, coaches could be advised to instruct their teams to shoot extensively while at the same time considering shot accuracy. However, advice of this kind would not do

justice to the complex nature of football and the demands of coaches. Bishop (2008) emphasized that only results providing performance-enhancing knowledge will be applied in practice. Hence, research has to deliver results that make it more likely to win. This also includes findings with regard to training, match preparation and coaching. Nash and Collins (2006) stated that coaching is a very complex and dynamic process. The actions of coaches are based on knowledge that has been acquired over years of experience and reflection, that is, tacit knowledge (Nash & Collins, 2006; Sternberg, 2003). For coaches, the importance of shots for scoring goals is more than obvious. It is also hardly surprising that pass accuracy, the opponent's quality and home advantage have a positive impact. A benefit for football coaches would be to reveal the partial influence of these variables including their interactions (e.g., by analyzing regression models).

However, there are less obvious findings that provide empirical evidence for beneficial tactical behaviors. First, possession is not as important as might be assumed (Collet, 2013; Liu et al., 2015; Mao et al., 2016). Second, a focus on counter attacks can be very effective and can be utilized as a successful tactical strategy, especially for underdogs (Tenga & Sigmundstad, 2017). Ball recovery in the zone between a team's own penalty area and center circle (Gómez et al., 2012) and a quick ball recovery (Vogelbein et al., 2014) can result in significantly more successful attacks respectively goals ($p<0.001$). Coaches can build on this evidence to improve tactical concepts. For example, coaches could put more emphasis on the practice of counter attacks, as a tactical element, to overwhelm the opponent's defense and produce more good scoring opportunities. Also pressing, the attempt to recover the ball as close as possible to the opponent's penalty area seems to be a promising tactic. It shortens not only the space between the attackers and the goal, it can also cause confusion within the opposing defense. This could lead to more goals since counterattacks are more effective against an imbalanced defense (Tenga, Holme et al., 2010).

2.11. Conclusions

The aim of this work was to review research in performance analysis relating to success factors in elite men's football. In total, 68 articles were identified and clustered based on their study design with regard to comparative, predictive or home advantage analyses. It was found that the most influential variables are efficiency, shots on goal, ball possession, pass accuracy/successful passes, as well as quality of opponent and match location. New statistical approaches, such as discriminant analysis, factor analysis, regression analysis and magnitude-based inferences reveal interactions between these variables.

Concerning study design, an increase of predictive studies was found. For future studies, we suggest considering more often one of the 'Big 3' leagues (Spain, England and Germany) or all of them to get more representative samples. Furthermore, the consideration of other influences on success such as psychological factors and/or weather conditions would be of interest. Additionally, new methodological ways of analyzing success factors in football could be beneficial. For example, Borrie et al. (2002) presented a method to investigate time-based events in sports. Moreover, more advanced statistical methods should be applied to ensure a broader insight into the mechanisms of performance such as regressions and magnitude-based inferences (Collet, 2013; Liu et al., 2015; Mechtel et al., 2011).

Most of the studies did not consider the influence of contextual (e.g., home advantage, quality of opponent) and interactional variables (e.g., first goal scored by time of goal scoring). In some studies, the influence of variables is also computed without a clear definition of the investigated variables. This lack of operational definitions poses a problem and, inter alia, does not allow valid comparisons between the studies. In future research, variables should be clearly defined to enable comparable and reproducible results (see also Mackenzie and Cushion (2013); Sarmento et al. (2014)). The consideration of interacting variables such as quality of opponent and match location should also be considered in future investigations to provide more insights. Future study designs should also make sure to take the differences between different competitions (e.g. leagues, cup competitions) into account, especially the differences between a league match and a knockout match.

Moreover, we found very different approaches regarding the sample size required for generalization. Sample sizes of considered matches varied between very low numbers and thousands of matches. A small sample size is clearly a limitation in some of the reviewed papers, resulting in no generalizability. Studies investigating league matches should consider at least a sample size of one season. Hence, our review supports the finding of Mackenzie and Cushion (2013) with regard to small sample sizes that remains a major deficit of performance analyses in football. Additionally, future studies should use effect sizes to interpret the results properly (see also Broich et al. (2014)). A last important aspect to consider when designing a study is the context of the analyzed sample. For example, the tactic that is used (e.g., counterattacks vs. elaborate attacks) could vary regarding the opponent.

Based on the idea that performance is a consequence of prior learning, inherent skills, situational factors and influence of the opposition (James, 2012), the assumption holds that future performance is to a large extent a consequence of previous performance. Again, this underlines

the aforementioned importance of considering the context of a sample as well as the operational definition of the investigated variables. Prior learning and inherent skills are two variables that were not considered in research about success factors in football as defined in this review. Both are exciting new possibilities for future research.

Finally, we would like to point to two methodological approaches that might lead to new insights in analyzing football performance. First, social network analysis provides new methods to analyze different aspects utilizing relational data, (e.g., the passing network of football teams), that have the potential to contribute substantially to a better understanding of success (Duch et al., 2010; Grund, 2012; Wäsche et al., 2017). Second, psychological factors could be taken into account for future research (e.g., reversal theory, see Apter (1984)). The investigation of psychological factors is in fact more difficult than the analysis of statistical data. The operationalization of cohesion found in this review (Carron et al., 2002) is a good example for the use of psychological concepts.[9]

As this review, has shown, generalizable knowledge about success factors in football can be a helpful resource for coaches to gain a better understanding of the match. While significant progress in the field of performance in football was made in the last years, the review identified various deficits that future research has to address to provide more valuable information about what determines success.

Acknowledgements

---

[9] Bar-Eli et al. (2006) focused also on a psychological factor. However, they focused on the factor that leads to a dismissal and not to a psychological factor that contributes directly to performance.

3. Success factors in the German Bundesliga

51

The original research article was published as:

Lepschy, H., Wäsche, H., & Woll, A. (2020). Success factors in football: An Analysis of the German Bundesliga. International Journal of Performance Analysis in Sport. https://doi.org/10.1080/24748668.2020.1726157

3.1. Abstract

Knowledge of success factors in professional football is critical. However, the identification of the success factors primarily focused on English and Spanish leagues. In this paper, three seasons of the German Bundesliga (2014/2015 until 2016/2017) with a total of 918 matches were analyzed. To facilitate a more precise analysis of success factors only close matches were included and the home and away team perspective was analyzed separately. Therefore, 29 variables were included in a generalized ordered logit approach. The results showed that, defensive errors, market value, goal efficiency, shots from counter attacks, shots on target, and total shots have the greatest impact. Furthermore, crosses showed a negative relationship with success. Besides, the opponent and home advantage were important contextual effects. Overall, eleven and twelve variables are significant, respectively. Duel success was only significant for away teams and a higher market value seems to have a more positive impact for them. This study provides novel data and contributes to prior results from other European leagues. Future research should further investigate the impact of ball possession and distance covered. Coaches should focus on accuracy rather than on quantity as well as train fitness (physically and mentally) to lower the risk of errors.

Keywords: performance analysis, soccer, sport analytic, match analysis, performance indicators

3.2. Introduction

Football has become the most popular sport in the world with billions of people around the globe watching the games (FIFA, 2015). To further improve the performance of football teams and players, various data is being produced in professional football leagues that provide multiple opportunities to analyze games and identify critical factors for success. In the past decades, a lot of research on performance analysis in football (association football or soccer were used as synonyms) has been conducted (for a detailed overview see Lepschy et al., 2018, Mackenzie & Cushion, 2013, Sarmento et al., 2014).

The knowledge of performance indicators that can determine success in football is critical. This is especially true since football is a sport where the outcome is not always free of chance (Dufour W., 1993; Reilly & Williams, 2003). Predictive studies enable the identification of new and useful insights on indicators of performance that can inform future efforts for performance improvement (Sarmento et al., 2014). However, in a recent review Lepschy et al. (2018) found that less than half of the studies dealing with success factors in football utilized predictive analyses. They concluded that there is a need for more predictive analyses to better understand determinants of success in football. Moreover, success in football cannot be explained with just a few variables. By analyzing the FIFA World Cup 2014 Liu et al. (2015) showed that most of the 24 variables investigated have an influence on the match outcome. In contrast, existing work regarding success factors especially in the German Bundesliga only focused on a few variables (Broich et al., 2014; Schauberger et al., 2017). To our knowledge there is no study on the German Bundesliga which investigated more than 10 variables at once. There are also only few studies in other European Leagues which included more than 20 variables (e.g., Liu et al., 2015; Liu et al., 2016; Oberstone, 2009).

While the performance of older players will most likely decrease after the age of 30 (Baker & Tang, 2010), the effect of age has yet not been considered in studies about success factors in football. Therefore, the average age of the starting formation was included in this paper. Another variable that has not been considered in previous research is the market value of the starting formation. Three studies investigated financial figures (e.g., revenue and salary) and their relationship to success so far. All three showed a positive relationship between success and the financial figures (Coates et al., 2016; Kringstad & Olsen, 2016; Torgler & Schmidt, 2007). To address the influence of financial power regarding success, this paper is the first – to our knowledge – that includes market value.

Moreover, it is not only the investigated variables that are relevant. It is also the perspective taken on the match outcome that should be well-considered. Outcome in football can be described as goal-based (goals scored and conceded by each team) or result-based (win, draw, lose) (Goddard, 2005). However, the goal-based approach does not result in a better model performance (Goddard, 2005). The closeness of the game seems to be the better approach to account for the goal difference and to overcome the moderator effect of one team that does not play at its best level (Gómez et al., 2014; Higham et al., 2014; Lupo et al., 2014; Sampaio et al., 2010; Vaz et al., 2010). For this reason, the sample used in this paper is divided into a group of matches with a narrow goal difference (close matches) and a group of matches with a wide goal difference (unbalanced matches) (Vaz et al., 2010).

In sum, the German Bundesliga has not been investigated as thoroughly as other European leagues (Lepschy et al., 2018). Notwithstanding, the quality of teams in the German Bundesliga is high, as it is reflected UEFA ranking for club competitions where Germany is ranked third (UEFA, 2019). In their review, Lepschy et al. (2018) found only seven studies which analyzed performance and success in football based on data from the Bundesliga (Lepschy et al., 2018). Interestingly, these analyses came to different conclusions. For example, Broich et al. (2014) and Yue et al. (2014) identified efficiency as the most influential variable in the Bundesliga. However, Schauberger et al. (2017) showed that (running) distance is the most important variable. This issue about the German Bundesliga will be addressed later. Besides the need to analyze the German Bundesliga in more detail, there is another reason to focus on one national competition. Mitrotasios et al. (2019) showed clear tactical differences in the top four European leagues in terms of goal scoring opportunities. These finding underlines that – instead of pooling data from different leagues – European football leagues should be analyzed separately.

The goal of this paper is to identify success factors in the first football division in Germany (Bundesliga) using an explorative approach with a broad variety of variables. Some of the variables have not been included yet in comparable studies. Moreover, a novel methodological approach is applied to considering only close matches and analyzing both, the home and away team perspectives. We aim to contribute to a better understanding of performance in German professional football. Furthermore, the objective is to add results to prior research on success factors in other European leagues that help to identify overarching patterns of success in football.

3.3. Methods

Sample

The match related data (except for duel success, distance, average age and market value) was obtained from the website www.whoscored.com. Data on duel success and distance were collected from www.kicker.de. Data for both websites is provided by OPTA Sports. Liu et al. (2013) showed a high inter-operator reliability for the system used by OPTA Sports. All match results were validated through the public website www.kicker.de. Data was collected for all matches from season 2014/2015 through season 2016/2017. This equals 102 match days with a total of 918 matches.

Variables

In alignment with previous research, 25 performance indicators and four contextual variables were included (see Table 11) (Broich et al., 2014; Castellano et al., 2012;; Lago et al., 2010; Lago et al., 2011; Liu et al., 2015; Liu et al., 2016; Oberstone, 2009; Yue et al., 2014). Operational definitions of the variables are given on the OPTA website (https://www.optasports.com/news/opta-s-event-definitions) (Opta, 2018). Home advantage/away disadvantage (negative value of home advantage) and quality of opponent (team rating) were calculated based on Clarke and Norman (1995) and included into the model as control variables.

Table 11. Performance indicators and contextual variables Bundesliga

Group	Variables
Variables related to goal scoring	Total shots, Shots on target, Shots from counter attack, Shots from inside 6-yard box, Shots from inside penalty area, Goal efficiency (Goals*100∕Total shots)
Variables related to passing and organizing	Ball possession (%), Passes, Pass accuracy (%), Long passes, Short passes, Average pass streak, Crosses, Successful dribbles, Offsides, Corners, Aerials won, Distance
Variables related to defense	Successful tackles, Fouls, Yellow cards, Red cards, Defensive errors, Duel success (%), Clearances
Contextual variables	Home advantage/Away disadvantage, Quality of opponent, Average age starting formation, Total market value starting formation

The data for the market value and the average age of the starting formation was drawn from the website Transfermarkt.de. Market value is estimated based on performance (e.g., successful passes, goals) including stability of the performance (recent performance has a higher value than past performance), experience (number of games played nationally and internationally including national team), perspectives for the future (anticipated value for younger players results in additional value), and prestige (public perception of the player and public perception of the club).

Data from Transfermarkt.de is used in various scientific analyses and the database is considered to be a reliable source (Göke et al., 2014). Although values are estimated, there is a high correlation with actual values (Frick, 2011). However, market value is not a standardized factor for the quality of the team. Nevertheless, football clubs pay enormous amounts for players. These amounts are reflected in the market value and can be used to anticipate the quality of a player and the team respectively.

Procedures

A Kolmogorov-Smirnov-Test of normal distribution showed that only the variables ball possession in percentage and distance are normally distributed. Results in football are mostly a Poisson distribution (Dixon & Coles, 1997; Maher, 1982; Rue & Salvesen, 2000).

To decide which matches are close and which are unbalanced, a two-step cluster analysis was performed (Gómez et al., 2014; Lupo et al., 2014; Liu et al., 2015; Sampaio et al., 2010). The analysis revealed one cluster containing 774 matches (.96 ± .759, 0 to 2 goal difference) and another cluster holding 144 matches (3.60 ± .934, 3 to 8 goal difference). The 774 matches with a close match result were used for statistical analyses.

Each match was analyzed twice (home team perspective and away team perspective). Team's tactical preparation, and team selection can vary depending on the location of the match (home vs. away) (Carmichael & Thomas, 2005). Home teams perform more attacking actions than away teams (Lago & Lago-Ballesteros, 2011; Poulter, 2009). Goumas (2014a) showed a negative referee bias towards away teams. Therefore, match statistics were modelled separately for home teams and away teams to account for possible differences in success factors.

Statistical analysis

A one-way analysis of variance (ANOVA) was performed to determine significant differences between home teams and away teams (Mechtel et al., 2011; Weiss, 1997). Match results were translated to be able to be modelled with an ordered approach, taking into account the order of

desirability (home team approach: 0=lost, 1=tied, 2=won; away team approach: 0=lost, 1=tied, 2=won).

The assumption of proportional odds in the ordered logit regression is often violated (Brant, 1990; Kleinbaum & Klein, 2010). This is also the case in this study. The test of the parallel regression was significant for the full model (home team approach: $Chi^2 = 54.65$; df = 29; p = 0.003; away team approach: $Chi^2 = 56.06$; df = 29; p = 0.002). Therefore, the generalized ordered logit regression (user-written *gologit2* in STATA) was used to calculate the effects (Williams, 2006; Williams, 2016).

The Variance Inflation Factor (VIF) was used to determine potential multicollinearity (Ender, 2010). A value of VIF $\geq$ 10 was set as the cut-off value based on the specified model (Craney & Surles, 2002). The variables passes (home: VIF= 463.50; away: VIF= 506.88) and short passes (home: VIF= 530.13; away: VIF= 582.40) were removed from both models. The remaining variables showed a VIF value < 10.

Pseudo R^2 was 0.2751 (home approach) and 0.2540 (away approach). A model fit between 0.2 and 0.4 is considered an excellent model fit (McFadden, 1977).

Marginal effects (command *margins* in STATA) were used to interpret the result (Cameron & Trivedi, 2010; Mechtel et al., 2011; Williams, 2012). The margins value indicates that on average a one unit increase in the independent variable changes the probability of the desired outcome by that number. This enables an interpretation of the importance of a factor. The significance level was set to p < 0.05 for all statistical analyses.

Data was analyzed with IBM SPSS Statistics and STATA.

3.4. Results

The descriptive statistics of close matches are presented in

Table 12 with results of the ANOVA to provide a comprehensive picture of the studied data. Home teams executed significantly more total shots (+2.08; p < 0.001), more shots on target (+0.66; p < 0.001), more shots from inside the 6-yard box (+0.27; p < 0.001), as well as more shots from inside the penalty area (+1.21; p < 0.001). Home teams also performed more crosses (+2.33; p < 0.001) and have a higher duel success rate (+0.76%; p < 0.01). The away teams conceded more fouls (+0.89; p < 0.001), more defensive errors (+0.07; p < 0.05) as well as more clearances (+3.18; p < 0.001). Away teams also got more yellow cards (+0.24; p < 0.001). Home advantage was significant and resulted on average in +0.36 goals (±0.58; p < 0.001).

Table 12. Descriptive Statistics and ANOVA Bundesliga

N=774	Away=0 Home=1	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	Upper Bound	Minimum	Maximum
Total shots***	0	11.68	4.61	.17	11.36	12.01	1.00	28.00
	1	13.76	5.23	.19	13.39	14.13	1.00	37.00
Shots on target***	0	4.15	2.28	.08	3.98	4.31	.00	13.00
	1	4.81	2.45	.09	4.63	4.98	.00	14.00
Shots from counter attack	0	.67	.92	.03	.60	.73	.00	6.00
	1	.74	.99	.04	.67	.81	.00	7.00
Shots from inside 6-yard box***	0	.63	.87	.03	.56	.69	.00	5.00
	1	.90	1.00	.04	.83	.98	.00	7.00
Shots from inside penalty area***	0	6.27	3.01	.11	6.06	6.48	.00	19.00
	1	7.48	3.55	.13	7.23	7.73	.00	19.00
Goal efficiency	0	10.40	9.92	.36	9.70	11.10	.00	50.00
	1	10.34	8.67	.31	9.72	10.95	.00	57.14
Ball possession (%)	0	49.63	12.28	.44	48.76	50.50	16.60	84.50
	1	50.37	12.28	.44	49.50	51.24	15.50	83.40
Passes	0	440.11	131.12	4.71	430.85	449.36	171.00	940.00
	1	444.86	130.30	4.68	435.66	454.05	118.00	972.00
Pass accuracy (%)	0	73.41	8.80	.32	72.79	74.03	44.00	92.00
	1	73.91	8.67	.31	73.29	74.52	42.00	92.00
Long passes	0	71.29	13.50	.49	70.34	72.25	26.00	113.00
	1	70.18	13.32	.48	69.24	71.12	26.00	113.00
Short passes	0	410.31	138.10	4.96	400.56	420.05	131.00	915.00
	1	419.08	137.42	4.94	409.38	428.78	100.00	990.00
Average pass streak	0	4.39	1.16	.04	4.31	4.47	2.00	10.00
	1	4.41	1.18	.04	4.33	4.49	2.00	10.00
Crosses***	0	12.85	5.73	.21	12.44	13.25	.00	35.00
	1	15.18	6.45	.23	14.73	15.64	1.00	42.00
Successful dribbles	0	9.53	4.52	.16	9.22	9.85	.00	30.00
	1	9.81	4.79	.17	9.47	10.15	.00	28.00
Offsides	0	2.35	1.76	.06	2.22	2.47	.00	10.00
	1	2.52	1.77	.06	2.40	2.65	.00	11.00
Corners	0	4.33	2.49	.09	4.15	4.50	.00	14.00

	1	5.12	2.86	.10	4.92	5.32	.00	18.00
Aerials won	0	24.75	9.32	.34	24.09	25.41	3.00	62.00
	1	24.91	9.30	.33	24.25	25.56	4.00	56.00
Distance	0	114.84	4.42	.16	114.53	115.15	99.54	128.61
	1	115.14	4.27	.15	114.84	115.44	102.65	128.95
Successful tackles	0	19.02	5.77	.21	18.61	19.43	5.00	38.00
	1	18.85	5.58	.20	18.46	19.25	3.00	41.00
Fouls***	0	15.42	4.39	.16	15.11	15.73	3.00	30.00
	1	14.53	4.19	.15	14.24	14.83	3.00	29.00
Red cards	0	.07	.26	.01	.06	.09	.00	1.00
	1	.08	.29	.01	.06	.10	.00	2.00
Yellow cards***	0	2.02	1.24	.04	1.93	2.11	.00	6.00
	1	1.78	1.24	.04	1.69	1.86	.00	6.00
Defensive errors*	0	.43	.67	.02	.38	.48	.00	3.00
	1	.36	.66	.02	.32	.41	.00	5.00
Duel success (%)**	0	49.62	4.85	.17	49.28	49.96	34.00	64.00
	1	50.38	4.85	.17	50.04	50.72	36.00	66.00
Clearances***	0	23.20	9.98	.36	22.49	23.90	3.00	64.00
	1	20.02	8.83	.32	19.40	20.64	2.00	56.00
Home advantage***	0	-.36	.58	.02	-.40	-.31	-1.66	1.15
	1	.36	.58	.02	.31	.40	-1.15	1.66
Team rating	0	-.04	.57	.02	-.08	.00	-1.15	1.67
	1	.01	.59	.02	-.03	.06	-1.15	1.67
Average age starting formation	0	26.38	1.23	.04	26.29	26.47	23.20	30.10
	1	26.40	1.24	.04	26.31	26.49	22.90	30.20
Total market value starting formation	0	77.12	78.02	2.80	71.62	82.63	9.35	422.00
	1	77.47	78.02	2.80	71.97	82.98	10.00	421.00

Significant differences between home teams and away teams ***$p < 0.001$ **$p < 0.01$ *$p < 0.05$.

The ANOVA showed that home teams performed more offensive actions, such as shots, which is consistent with previous research (e.g., Lago & Lago-Ballesteros, 2011; Poulter, 2009). A significant difference in yellow cards was found, which is not in line with an earlier study about European cup competitions (Goumas, 2014a). The average home advantage measure in goal difference was 0.36 goals, which is lower than the 0.5 goals found by Clarke and Norman (1995). This supports the hypothesis that home advantage is lowering in general (e.g., Pollard & Gómez, 2009). Sánchez et al. (2009) proposed discouraged defensive play and a weakened relationship between players and their fans as possible reasons.

Success factors in the German Bundesliga

The results of the marginal effects calculations for the desired outcome of a win are presented in Table 13 (home approach) and Table 14 (away approach). Home advantage/away disadvantage showed a significant positive/negative impact on the probability of winning the match. On average a one unit increase of the variable increased the probability of winning by 0.0929 ($p < 0.001$) and 0.0806 ($p < 0.001$) respectively. Moreover, quality of opponent (team rating) posed a significant negative influence for home teams and away teams (-0.0415; $p < 0.05$ and -0.0564; $p < 0.01$, respectively). From the home team perspective, shots from counter attacks, goal efficiency, clearances, shots on target, shots from inside the penalty area, total market value starting formation and total shots all had a significant positive influence on the probability of a home team win. In contrast, defensive errors and crosses had a significant negative influence. From the away team perspective, total shots, goal efficiency, clearances, total market value starting formation, shots from counter attack, duel success (%) and shots on target had a significant positive influence on the probability of an away team win. In contradiction, successful tackles, defensive errors and crosses all had a significant negative influence.

Table 13. Marginal effects from a home team perspective for the outcome win of home team Bundesliga

	dy/dx	Std. Err.	z	P>z	95% Conf. Interval	
Total shots home*	0.0085	0.0043	2.0000	0.0460	0.0002	0.0168
Shots on target home**	0.0165	0.0063	2.6000	0.0090	0.0041	0.0289
Shots from counter attack home***	0.0648	0.0133	4.8700	0.0000	0.0387	0.0908
Shots from inside 6-yard box home	-0.0035	0.0118	-0.3000	0.7640	-0.0267	0.0196
Shots from inside penalty area home*	0.0129	0.0057	2.2700	0.0230	0.0018	0.0240
Goal efficiency home***	0.0236	0.0018	12.8200	0.0000	0.0200	0.0272
Ball possession (%) home	0.0014	0.0018	0.8000	0.4210	-0.0021	0.0049
Pass accuracy (%) home	0.0034	0.0026	1.2800	0.2020	-0.0018	0.0085
Long passes home	0.0015	0.0010	1.5500	0.1220	-0.0004	0.0035
Average pass streak home	-0.0115	0.0112	-1.0300	0.3050	-0.0335	0.0105
Crosses home***	-0.0120	0.0028	-4.3000	0.0000	-0.0175	-0.0066
Successful dribbles home	-0.0032	0.0025	-1.3200	0.1860	-0.0080	0.0016
Offsides home	-0.0025	0.0058	-0.4300	0.6680	-0.0138	0.0088
Corners home	0.0075	0.0058	1.2900	0.1990	-0.0039	0.0189
Aerials won home	-0.0014	0.0017	-0.8600	0.3880	-0.0047	0.0018
Distance home	0.0043	0.0035	1.2000	0.2290	-0.0027	0.0112
Successful tackles home	-0.0028	0.0024	-1.1700	0.2420	-0.0075	0.0019
Fouls home	0.0021	0.0031	0.6700	0.5040	-0.0040	0.0081

Yellow cards home	-0.0165	0.0090	-1.8400	0.0660	-0.0342	0.0011
Red cards home	0.0496	0.0374	1.3200	0.1850	-0.0238	0.1229
Defensive errors home***	-0.0975	0.0162	-6.0100	0.0000	-0.1293	-0.0657
Duel success (%) home	0.0055	0.0029	1.9200	0.0540	-0.0001	0.0111
Clearances home***	0.0074	0.0014	5.0800	0.0000	0.0045	0.0102
Home advantage***	0.0929	0.0197	4.7300	0.0000	0.0544	0.1315
Team rating away*	-0.0415	0.0191	-2.1800	0.0290	-0.0789	-0.0041
Average age starting formation home	-0.0030	0.0086	-0.3400	0.7320	-0.0199	0.0140
Total market value starting formation home*	0.0003	0.0001	2.1500	0.0320	0.0000	0.0006

***$p < 0.001$ **$p < 0.01$ *$p < 0.05$.

Table 14. Marginal effects from an away team perspective for the outcome win of away team Bundesliga

	dy/dx	Std. Err.	z	P>z	95% Conf. Interval	
Total shots away***	0.0186	0.0046	4.0300	0.0000	0.0095	0.0276
Shots on target away*	0.0125	0.0061	2.0600	0.0400	0.0006	0.0245
Shots from counter attack away**	0.0327	0.0102	3.2200	0.0010	0.0128	0.0526
Shots from inside 6-yard box away	0.0098	0.0119	0.8200	0.4110	-0.0136	0.0332
Shots from inside penalty area away	0.0034	0.0057	0.5900	0.5530	-0.0077	0.0145
Goal efficiency away***	0.0153	0.0015	10.0900	0.0000	0.0123	0.0182
Ball possession (%) away	-0.0007	0.0016	-0.4200	0.6730	-0.0039	0.0025
Pass accuracy (%) away	0.0037	0.0024	1.5500	0.1210	-0.0010	0.0083
Long passes away	0.0006	0.0008	0.6900	0.4900	-0.0010	0.0022
Average pass streak away	-0.0013	0.0090	-0.1400	0.8870	-0.0189	0.0163
Crosses away**	-0.0071	0.0024	-2.9100	0.0040	-0.0118	-0.0023
Successful dribbles away	0.0000	0.0023	-0.0100	0.9930	-0.0046	0.0046
Offsides away	0.0034	0.0058	0.5900	0.5570	-0.0080	0.0149
Corners away	0.0066	0.0055	1.1800	0.2360	-0.0043	0.0174
Aerials won away	0.0003	0.0016	0.2000	0.8410	-0.0028	0.0035
Distance away	0.0041	0.0022	1.8300	0.0670	-0.0003	0.0085
Successful tackles away***	-0.0072	0.0020	-3.5400	0.0000	-0.0112	-0.0032
Fouls away	0.0011	0.0028	0.4200	0.6770	-0.0042	0.0065
Yellow cards away	0.0114	0.0078	1.4500	0.1460	-0.0040	0.0268
Red cards away	-0.0238	0.0359	-0.6600	0.5080	-0.0941	0.0465
Defensive errors away***	-0.0751	0.0149	-5.0600	0.0000	-0.1042	-0.0460
Duel success (%) away**	0.0078	0.0028	2.8200	0.0050	0.0024	0.0133

Clearances away***	0.0068	0.0014	5.0100	0.0000	0.0042	0.0095
Away disadvantage***	0.0806	0.0180	4.4700	0.0000	0.0453	0.1159
Team rating home**	-0.0564	0.0181	-3.1100	0.0020	-0.0920	-0.0209
Average age starting formation away	-0.0006	0.0078	-0.0800	0.9380	-0.0159	0.0147
Total market value starting formation away**	0.0005	0.0001	3.5600	0.0000	0.0002	0.0007

***$p < 0.001$ **$p < 0.01$ *$p < 0.05$.

The four variables with the highest margins value are displayed for the home team perspective in Figure 2 and for the away team perspective in Figure 3.
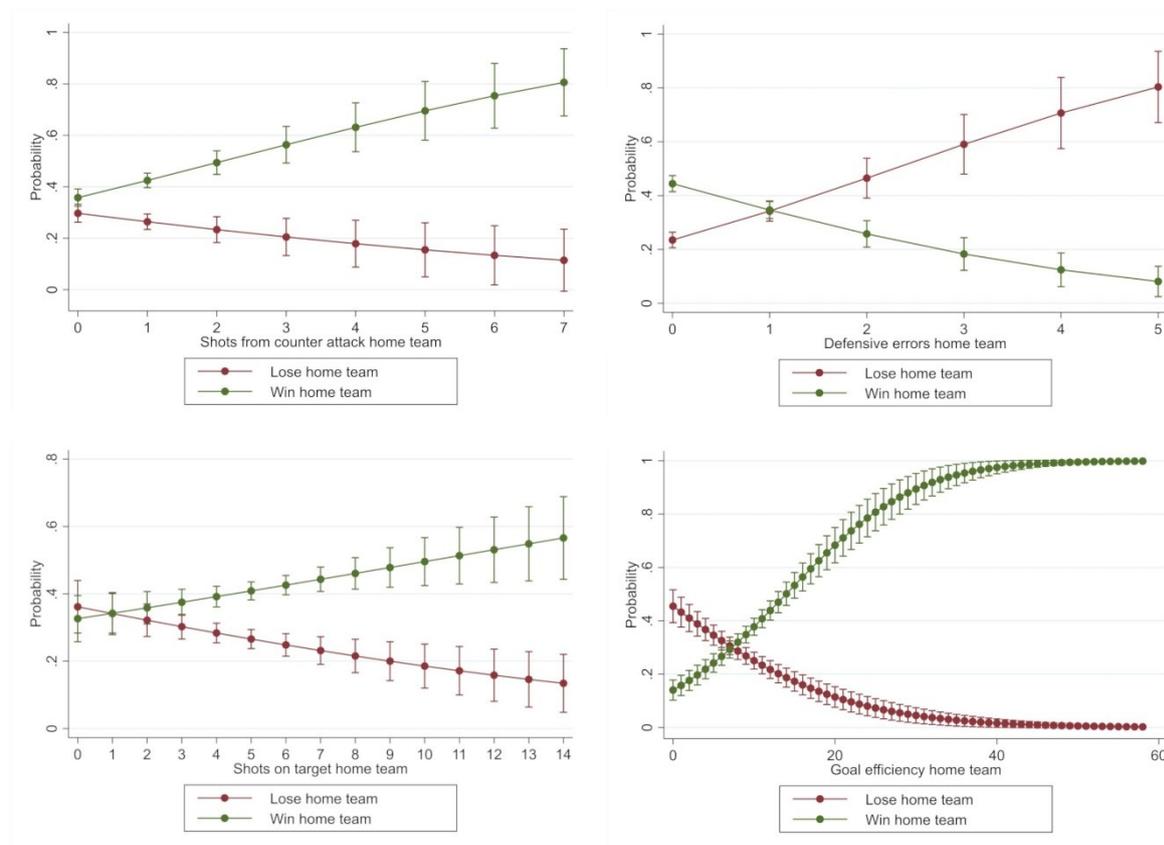

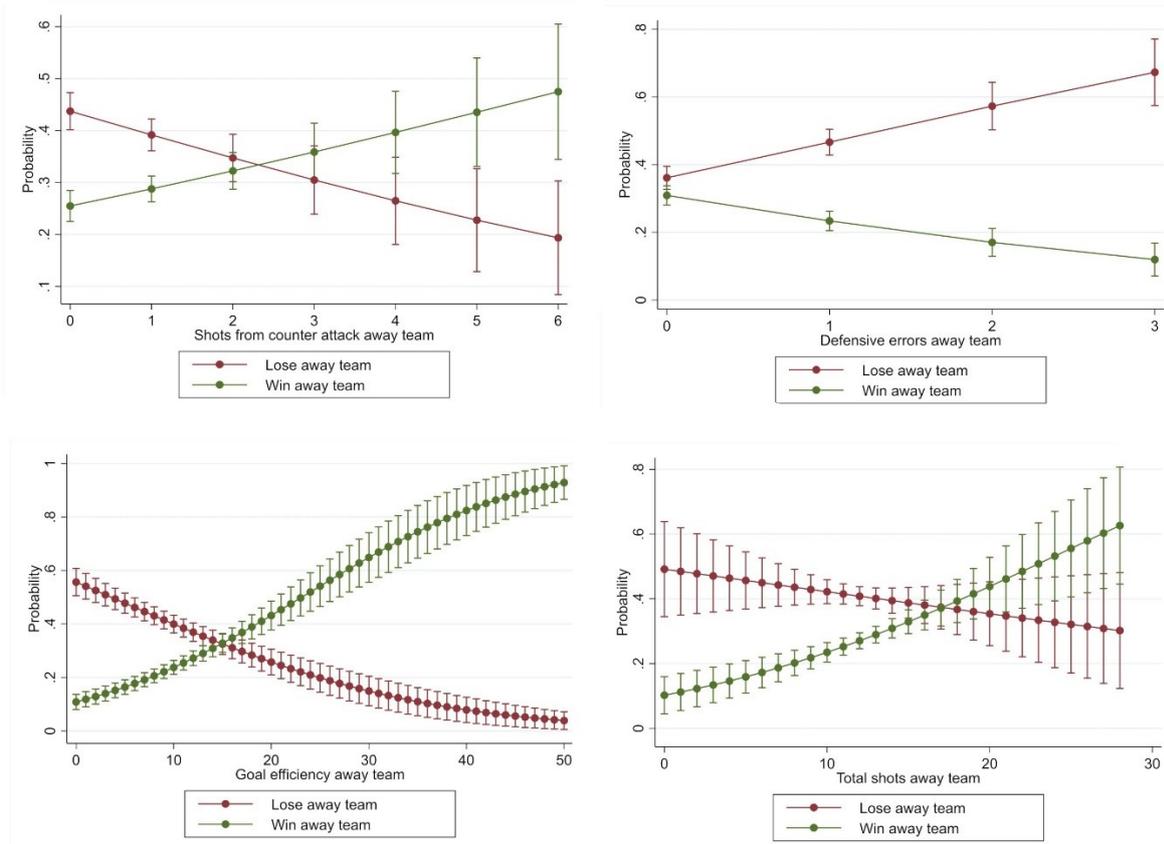Figure 2. Highest margin values for the home team perspective Bundesliga

Figure 3. Highest margin values for the away team perspective Bundesliga

3.5. Discussion

This study aims to determine the success factors in football in the German Bundesliga using a broad variety of variables including market value. The results are discussed in the following order. First, an overview of significant variables, second the most influential factors, third less influential variables and fourth, and lastly non-significant factors in contrast to previous research. Finally, practical implications as well as limitations and directions for future research are discussed.

Overview of significant variables

The analysis revealed that, if controlled for home advantage and quality of opponent, defensive errors, shots from counter attacks, goal efficiency, clearances, shots on target, shots from inside the penalty area, crosses, total market value starting formation and total shots are significant predictors for success from a home team perspective. For the away team, defensive errors, total

shots, goal efficiency, clearances, total market value starting formation, shots from counter attack, duel success (%), shots on target, successful tackles, and crosses had a significant influence on winning.

Most influential factors

Defensive errors showed the highest influence of all performance statistics. Even though less errors seem to be an intuitive success factor, it was rarely analyzed in previous research. The high value of -9.75% (home) and -7.51% (away) influence of defensive errors can also be explained by its operational definition "A mistake made by a player losing the ball that leads to a shot or a goal." (Opta, 2018). In such situations, the defense is usually imbalanced since the team possessed the ball and focused on the next attack. Playing against an imbalanced defense also increases the chance of goal scoring (Tenga, Holme et al., 2010). However, getting the chance to score does not mean that you will score. Subsequently, it was revealed that goal efficiency (Goals*100∕Total shots) is one of the most important success factors. The results underline that not only frequency of shots, but also quality of shots is critical. This is in line with the findings of Broich et al. (2014) and Yue et al. (2014).

The results about shots and shots on target support the conclusions from previous research showing that these two variables have a significant impact on success (Dufour et al., 2017; Lago et al., 2010; Liu et al., 2015; Liu et al., 2016; Lago, 2007; Oberstone, 2009; Pappalardo & Cintia, 2018; Yang et al., 2018)

Less influential variables

More successful tackles were linked to a negative outcome, but only for the away team. It is likely that many defensive actions are leading to a loss since the team is forced to defend more than to attack even if the amount of successful tackles is high, but this does not say anything about the amount of unsuccessful tackles. This is also supported by the positive impact of duel success for away teams (see Table 14). There the relationship between successful and unsuccessful duels is considered. This is in line with Schauberger et al. (2017) who showed that the tackling rate (rate of won tackles) has a significant positive effect on winning (see also Liu et al., (2016) with regard to the first division in Spain). Additionally clearances showed a significant positive effect on winning which is in line with previous research (Carmichael et al., 2000). Even the small effect size of clearances might hide the fact that this effect could become substantial since the difference in clearances can be sizeable (see Table 12).

More crosses were associated with a decrease in the probability of winning. The operational definition of crosses is: "Any ball sent by a player into the opposition team's area from a wide position" (Opta, 2018). Accuracy is not considered here, pointing towards the fact that the sheer sending of the ball into the opposition team's area does not say anything about the quality of the cross. Lago et al. (2010) and Liu et al. (2015) also showed that more crosses are negatively linked to success. Reis et al. (2017) similarly showed long distance passes are mostly not effective and result in losing ball possession.

The market value of the starting formation showed significant but low values of 0.0003 for home and 0.0005 for away teams. This is in alignment with previous research regarding success and financial figures (Coates et al., 2016; Kringstad & Olsen, 2016). However, the low value may be misleading. It means that a one million Euro increase in market value increases the probability of winning by 0.03% and 0.05% respectively. The range between minimum and maximum value was more than 400 million Euros. In this case the probability of winning would increase by 12.3% (411x0.03%) and 20.6% (413x0.05%) respectively. Hence, market value seems to be a substantial success factor for the German Bundesliga. To our knowledge, market value was investigated for the first time in the context of football-specific success factors and therefore needs to be addressed in future research.

Non-significant factors in contrast to previous research

Ball possession has been widely discussed with different results ranging from a positive effect to a negative effect (Lepschy et al., 2018). In this study, the effect of ball possession was not significant and showed no clear tendency. This result supports other studies which controlled for possible moderating effects (Collet, 2013; Liu et al., 2015). The possession of the ball seems to be less important. Similarly, distance was not significant in the study which contradicts Schauberger et al. (2017). They showed that distance is strongly connected to match outcome but analyzed only eight variables. In contrast, Yang et al. (2018) showed that total distance without ball possession has no significant influence on winning. Hoppe et al. (2015) only focused on match running performance, and showed that only distance with ball possession is a significant predictor for accumulated points. This shows that the true influence of the distance and ball possession remains unclear and needs to be investigated in future studies.

In contrast, the average age of the starting formation appears to exert no influence on the result. While this might sound counterintuitive, it can be explained by a well-distributed age structure (see Table 12).

Practical implications

The results point towards various aspects that could foster the performance of football teams in the German Bundesliga. In general, there seems to be a tendency that accuracy (e.g., goal efficiency, shots on target) is critical. Regarding training planning and match preparation, a stronger emphasis on accuracy rather than on quantity (number of game actions such as shots, passes and tackles) could be more beneficial. Moreover, coordination, accuracy of shots, and tactical and physical ability to get into a favorable position (e.g., to shoot on target from close range) are trainable skills that can all increase goal efficiency. Due to the high influence of defensive errors on match outcome, another emphasis could be put on increasing the quality of the defense and defenders to lower the rate of defensive errors. Moreover, the data indicate that fitness (physically and mentally) should be well trained to lower the risk of errors (see also Njororai (2012)). Since shots from counterattacks are an important success factor, match preparation could take into account possible benefits of counterattacks. This could be relevant, for example, when playing against a team which favors ball possession. Finally, it is important to know for coaches and managers that external factors like the market value and the venue of the games must be considered to explain success.

Limitations and future research

Due to concerns about multicollinearity, passes and short passes were removed. Hence, no conclusion about the influence on success of these variables could be made. Additionally, data on market value is not a standardized factor, which can be easily assessed or counted. It is also noteworthy, that the average age of the starting formation can be the same for two teams, while the age structure is different. Hence, an influence of age cannot be ruled out completely. Finally, during the analyzed seasons Bayern Munich was the dominating team of the Bundesliga and had the highest market value. However, despite some limitations this research provided the first comprehensive and broad overview for the German Bundesliga that included for 29 success factors.

This study showed that the true nature of ball possession and distance is a field for future research. While we showed that both variables had no significant influence in the past three seasons of the German Bundesliga, previous studies revealed significant effects for theses variables or parts of them (e.g. Dufour et al., 2017; Lago et al., 2011; Schauberger et al., 2017). Furthermore, future research needs to control for home advantage and quality of opponent to reveal the true influence of other performance factors. Both variables showed a high influence in this

study. The determination of absolute effects in the German Bundesliga should only be the first step to discover the differences between the leagues in different countries. For further insights, non-physical factors such as motivational and social factors should also be considered as well as the above-mentioned age structure of a team. Moreover, new methodological approaches such as social network analysis enable the analysis of further performance indicators (Wäsche et al., 2017). Finally, a complex systems' view on football can provide additional insights (Pappalardo & Cintia, 2018).

3.6. Conclusions

This study showed that avoidance defensive errors are the most important success factor for home teams and away teams in the German Bundesliga where the negative effect tends to be greater for home teams. The following three most influential factors are goal efficiency, shots from counterattacks and shots on target (home teams) and total shots (away team) respectively. Some factors differ in the amount of influence between home teams and away teams, but, successful tackles (negative effect) and duel success (positive effect) are only significant for away teams. For the first time it was shown that the total market value of the starting formation has a significant positive influence on the winning probability which is slightly higher for the away team. Interestingly, more crosses are associated with a lower probability of winning. Overall, it seems that efficiency and accuracy are more important than the absolute number of game actions. This is not only the case for shots but also for passes and tackles. The results can support coaches in training improvements and match preparation. The quality of opponent and home advantage are important contextual variables which should be accounted for when analyzing success factors in football.

4. Success factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil

The original research article is under review as:

Lepschy, H., Woll, A. & Wäsche, H. (under review). Success factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil.

4.1. Abstract

Research on success factors in football focusing on national teams is sparse. The current study examines the success factors during the World Cup 2018 in Russia and the World Cup 2014 in Brazil. A total of 128 matches were analyzed using a generalized order logit approach. 29 variables were identified from previous research. This is a novel method for analyzing football matches. The results showed that defensive errors, goal efficiency, duel success, tackles success, shots from counter attacks, clearances, and crosses have a significant influence on winning a match during those tournaments. Ball possession, distance and market value of the teams had no influence on success. Overall, most of the critical success factors and those with the highest impact on winning close games were defensive actions. Moreover, the results suggest that direct play and pressing were more effective than ball possession play. The study contributes to a better understanding of success factors and can help to improve effectiveness of training, match preparation and coaching.

Keywords: match analysis, performance analysis, performance indicators, soccer, sport analytics

4.2. Introduction

To understand the mechanisms underlying success in football remains a challenge. Various attempts have been undertaken to identify and quantify indicators of performance, but results vary and are partly inconsistent. Most studies focused on domestic leagues consisting of club teams, while studies on the performance of national teams at tournaments are sparse. Only eleven studies involving data of success factors from a World cup were published in recent years (Lepschy et al., 2018). Of these studies only six used a predictive study design which can provide more sophisticated conclusions (Lepschy et al., 2018).

The most studied variables with regard to success factors in football are shots and shots on goals followed by variables like goal efficiency (number of goals divided by shots), passing, and possession (Lepschy et al., 2018). Goal efficiency and shots on goal were shown to be important factors for winning a football match (Broich et al., 2014; Lago et al., 2010; Lepschy et al., 2020). Ball possession and passing showed mixed results but seem to be no significant success factor if studies are controlled for other variables (Collet, 2013; Lepschy et al., 2020; Liu et al., 2015; Oberstone, 2009). Lepschy et al. (2020) studied the success factors of the German Bundesliga and showed that defensive errors are an influential success factor. They also revealed a significant effect for the total market value of the starting formation. Home advantage and the

quality of opponent are two further important contextual variables explaining success (Clarke & Norman, 1995; Goumas, 2015; Pollard, 2006).

Success in football games is usually evaluated based on results (win, draw, loss) or based on goals (goals scored and conceded by each team). Despite providing more information, the goal-based approach does not perform better than the result-based approach (Goddard, 2005). An alternative method is the approach of the closeness/balance of the game which allows to over-come the moderator effect that one team does not play at its best level when the game is seem-ingly decided (Gómez et al., 2014; Higham et al., 2014; Lepschy et al., 2020; Lupo et al., 2014; Sampaio et al., 2010; Vaz et al., 2010). The approach of unbalanced matches and close matches divides the sample into a group of matches with a narrow goal difference (close matches) and a group of matches with a wide goal difference (unbalanced matches) (Vaz et al., 2010). This approach will be also used in this study to avoid the inclusion of biased data.

The goal of this study is to identify the success factors for the FIFA World Cup 2018 in Russia and the FIFA World Cup 2014 in Brazil using an elaborated statistical approach. 29 variables will be investigated using a result-based approach. This will be the first study to include market value as success factor of a FIFA World Cup.

4.3. Methods

The data used for this study were freely available. Most data (except duel success, distance, average age and market value) for all 128 matches were collected from www.whoscored.com. The data for duel success were gathered from www.kicker.de. The data on both websites are provided by OPTA. The data for distance covered were collected from www.fifa.com. The data about market value of the starting formation and average age were retrieved from www.trans-fermarkt.de. The market value is an estimated figure, which is built on different aspects. The following factors are part of the estimation: performance and stability of the performance, ex-perience, perspectives for the future, and prestige. These data have been used in various studies and are considered to be reliable (Göke et al., 2014) and show a high correlation with actual values (Frick, 2011). The average age of the starting formation is the age of each player at the day of the match day summarized and divided by eleven. The operational definition of the 25 performance variables can be found on opta.com (Opta, 2018) and whoscored.com (Whoscored.com, 2018).

To take into account the effect of home advantage (twelve matches were played by the host nations in 2014 and 2018) a binary dummy variable for home advantage was included in the

analysis. To control for the strength of the opponent the last FIFA coefficient prior to the tournament was used (FIFA, 2014; FIFA, 2018). The FIFA coefficient was used since it is the only official rating of national teams playing in a World Cup (Gásquez & Royuela, 2016). Eventually, the 29 variables related to goal scoring, to passing and organizing, to defense, and context were included in the analysis (Table 15).

Table 15. Performance variables and contextual variables World Cups

Group	Variables
Variables related to goal scoring	Total shots, Shots on target, Shots from counter attack, Shots from inside 6-yard box, Shots from inside penalty area, Goal efficiency (Goals*100∕Total shots)
Variables related to passing and organizing	Ball possession (%), Passes, Pass accuracy (%), Long passes, Short passes[10], Average pass streak, Crosses, Successful dribbles, Corners, Aerials won, Distance in kilometers
Variables related to defense	Tackles success (%), Fouls, Yellow cards, Red cards, Defensive errors, Duel success (%), Clearances, Interceptions
Contextual variables	Quality of opponent (FIFA coefficient), Average age starting formation, Total market value starting formation, Home advantage (0;1)

The tournament rules allow matches to be only decided after 30 minutes of extra time and/or a penalty shootout. Eight matches were decided through a penalty shootout, these were counted as tied. Five matches were decided after extra time, these were counted as a win for the respective team. The dependent variable was in all cases the result-based outcome of the match, described as win, draw or loss.

A K-means cluster was used to determine the balance of the game. 108 matches classified as close (goal difference 0 to 2 goals) and 20 matches as unbalanced (goal difference 3 or more

---

[10] Removed after test of multicollinearity

goals). The 108 matches were analyzed twice since the home team (first mentioned team) on the schedule is not playing at home except for the twelve matches mentioned before. Hence, this analysis is based on 216 observations of which all available information could be used.

The test of parallel regression was significant (Brant: Chi^2=260.7; p=0.000) therefore the assumption of proportional odds is violated (Brant, 1990). Consequently, the generalized ordered logit regression was used for the analysis (Williams, 2016). To test for the multicollinearity the command *collin* was used (Ender, 2010). A Variance Inflation Factor (VIF) above 10 was set as the cut off value (Craney & Surles, 2002). The variables passes (VIF= 654.69) and short passes (VIF= 662.10) showed higher values. The variable short passes was removed from the model. Pseudo R2 of the analyzed model was 0.3622. To interpret the results marginal effects (command *margins*) were calculated (Mechtel et al., 2011; Williams, 2012). The significance level was set to p<0.05 for all statistical analyses.

The data were analyzed in IBM SPSS Statistics 24 and STATA 15. The study received ethical approval by the Institutional Review Board of the Institute of Sports and Sports Science, Karlsruhe, Germany.

4.4. Results

The descriptive statistics are presented in Table 16. The average goals per match were 2.66 (2.64 in 2018 and 2.67 in 2014).

Table 16. Descriptive statistics World Cups

	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
				Lower Bound	Upper Bound		
Total shots	13.00	5.43	0.37	12.27	13.72	3.00	39.00
Shots on target	4.14	2.41	0.16	3.82	4.47	0.00	17.00
Shots from counter attack	0.37	0.80	0.05	0.26	0.48	0.00	5.00
Shots from inside 6-yard box	0.73	0.92	0.06	0.60	0.85	0.00	4.00
Shots from inside Penalty Area	6.40	3.34	0.23	5.96	6.85	1.00	23.00
Goal efficiency	10.12	9.68	0.66	8.83	11.42	0.00	57.14
Ball possession (%)	50.00	12.46	0.85	48.33	51.67	21.00	79.00

Passes	447.32	137.83	9.38	428.83	465.80	156.00	1115.00
Pass accuracy (%)	79.97	7.10	0.48	79.02	80.92	57.00	93.00
Long passes	58.92	13.96	0.95	57.05	60.79	29.00	107.00
Short passes	434.05	143.17	9.74	414.85	453.25	147.00	1104.00
Average pass streak	4.61	1.28	0.09	4.44	4.78	2.00	10.00
Crosses	18.78	8.69	0.59	17.61	19.94	3.00	53.00
Successful dribbles	10.14	4.62	0.32	9.52	10.76	1.00	23.00
Corners	5.07	2.75	0.19	4.71	5.44	0.00	19.00
Aerials won	17.57	7.59	0.52	16.55	18.58	2.00	49.00
Distance	109.63	11.73	0.80	108.05	111.20	93.00	155.00
Tackles success (%)	64.63	10.76	0.73	63.19	66.08	33.33	94.44
Fouls	14.26	5.03	0.34	13.58	14.93	4.00	31.00
Yellow cards	1.57	1.15	0.08	1.42	1.73	0.00	6.00
Red cards	0.05	0.21	0.01	0.02	0.08	0.00	1.00
Defensive errors	0.40	0.65	0.04	0.31	0.49	0.00	3.00
Duel success (%)	50.00	5.46	0.37	49.27	50.73	36.00	64.00
Clearances	25.25	10.68	0.73	23.82	26.68	4.00	67.00
Interceptions	11.66	5.00	0.34	10.99	12.33	2.00	29.00
FIFA coefficient	964.53	249.78	17.00	931.03	998.03	457.00	1558.00
Average age starting formation	27.84	1.38	0.09	27.65	28.02	24.40	30.90
Total market value starting formation	191.52	180.41	12.28	167.32	215.71	4.83	710.00

The marginal effects for the outcome 'win' of all analyzed variables are displayed in Table 17. Shots from counterattack, goal efficiency, crosses, tackle success (%), defensive errors, duel success (%) and clearances had a significant influence on winning a match. Defensive errors showed the highest influence (dy/dx = -0.1025, $p < 0.05$) with one defensive error decreasing the probability of winning by 10.25%. One additional shot from a counter attack increased the chance of winning by 6.51% (dy/dx = 0.0651, $p < 0.05$). However, duel success (%) and goal efficiency showed to be important as well and highly significant (dy/dx = 0.0214, $p < 0.01$ respectively dy/dx = 0.0193, $p < 0.01$). None of the contextual variables showed a significant

impact. However, the contextual variable home advantage had the highest positive value (dy/dx = 0.0822, p = 0.4780) of all variables[11].

Table 17. Marginal effects for the outcome 'win' World Cups

	dy/dx	Std. Err.	z	P>z	95% Conf. Interval	
Total shots	0.0149	0.0099	1.5100	0.1310	-0.0044	0.0343
Shots on target	0.0182	0.0153	1.1900	0.2340	-0.0118	0.0481
Shots from counter attack*	0.0651	0.0326	2.0000	0.0460	0.0012	0.1291
Shots from inside 6-yard box	0.0090	0.0278	0.3200	0.7460	-0.0454	0.0634
Shots from inside penalty area	0.0003	0.0124	0.0200	0.9810	-0.0239	0.0245
Goal efficiency**	0.0193	0.0034	5.7300	0.0000	0.0127	0.0259
Ball possession (%)	0.0091	0.0052	1.7500	0.0810	-0.0011	0.0192
Passes	-0.0004	0.0005	-0.7600	0.4450	-0.0014	0.0006
Pass accuracy (%)	-0.0082	0.0065	-1.2700	0.2050	-0.0209	0.0045
Long passes	-0.0013	0.0022	-0.5700	0.5710	-0.0057	0.0031
Average pass streak	0.0121	0.0393	0.3100	0.7590	-0.0650	0.0891
Crosses*	-0.0111	0.0046	-2.4100	0.0160	-0.0201	-0.0021
Successful dribbles	-0.0066	0.0065	-1.0300	0.3050	-0.0193	0.0060
Corners	0.0044	0.0127	0.3400	0.7320	-0.0205	0.0293
Aerials won	-0.0021	0.0036	-0.5900	0.5540	-0.0092	0.0049
Distance	-0.0021	0.0028	-0.7600	0.4470	-0.0075	0.0033
Tackles success (%)*	0.0057	0.0022	2.5600	0.0100	0.0013	0.0100
Fouls	0.0065	0.0065	1.0000	0.3150	-0.0062	0.0192
Yellow cards	-0.0148	0.0192	-0.7700	0.4400	-0.0525	0.0228
Red cards	-0.0165	0.0768	-0.2100	0.8300	-0.1669	0.1339
Defensive errors*	-0.1025	0.0448	-2.2900	0.0220	-0.1903	-0.0148
Duel success (%)**	0.0214	0.0062	3.4900	0.0000	0.0094	0.0335
Clearances*	0.0084	0.0034	2.4900	0.0130	0.0018	0.0150
Interceptions	-0.0038	0.0046	-0.8300	0.4080	-0.0130	0.0053
FIFA coefficient	0.0002	0.0001	1.3600	0.1730	-0.0001	0.0004
Average age starting formation	-0.0190	0.0171	-1.1100	0.2660	-0.0525	0.0145
Total Market value starting formation	0.0001	0.0002	0.5300	0.5980	-0.0003	0.0005
Home advantage	0.0822	0.1158	0.7100	0.4780	-0.1448	0.3093

**p < 0.001 *p < 0.05.

The seven significant variables including the 95% confidence intervals are also shown in Figure 4. All graphs show a clear development of the predictors regarding the probability of winning or losing. The higher or lower the value of the predictor the higher is the probability of winning or losing.

[11] Non-significance is due to the small sample size for home advantage (n = 12).

Success factors in the FIFA 2018 World Cup in Russia and FIFA 2014 World Cup in Brazil
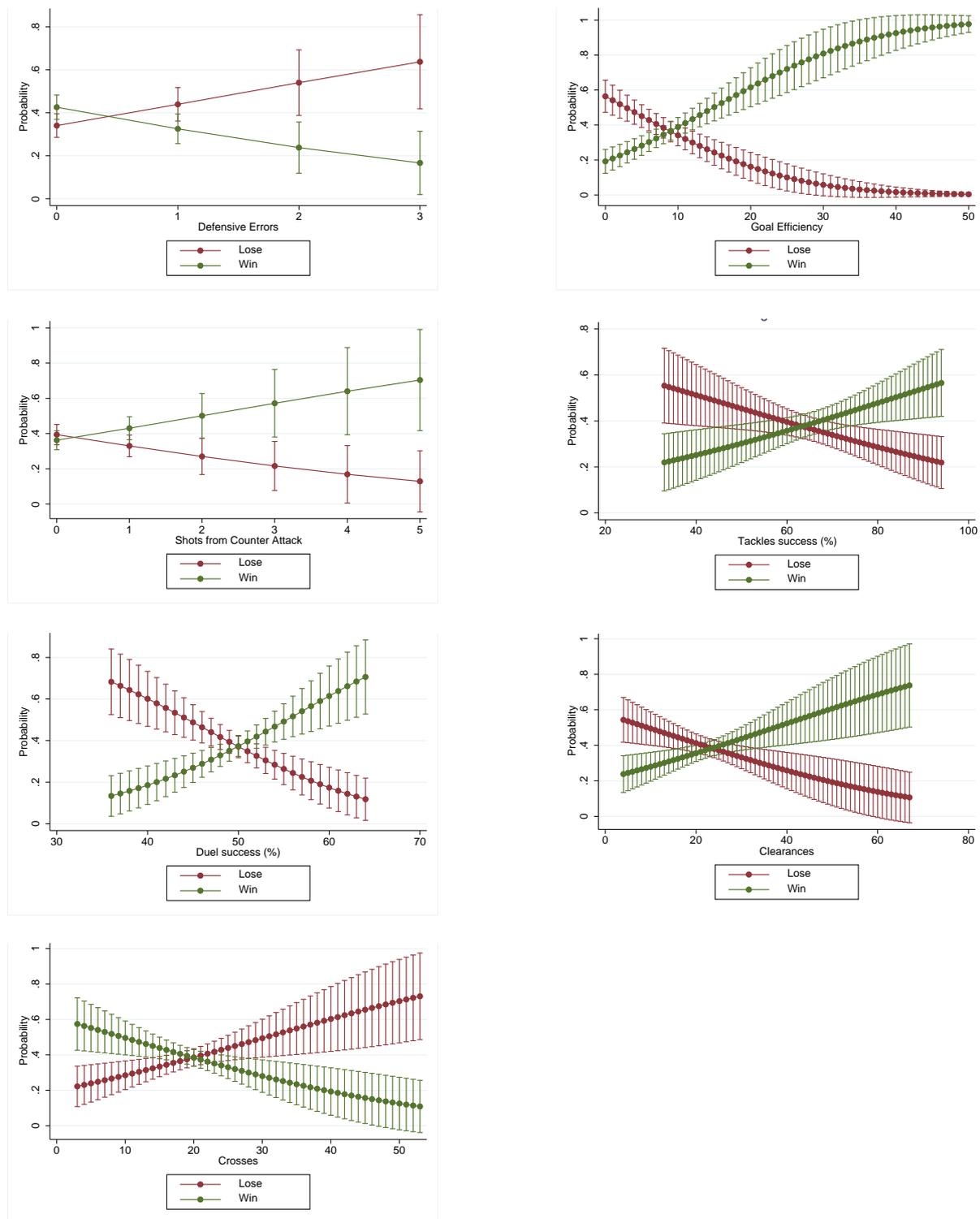

Figure 4. Margins with 95% CIs of the significant variables World Cups

4.5. Discussion

The purpose of this research was to identify success factors in the games played at the Football World Cups in 2018 and 2014. The significant positive success factors during the World Cup 2018 and 2014 were shots from counter attack, duel success (%), goal efficiency (%), clearances

and tackles success (%). On the other hand, defensive errors and crosses had a significant negative impact on the probability of winning. Despite none of the contextual factors in this study were significant it is still worth noting that the effects of those variables were substantial.

Overview of significant variables

Of the significant variables, most variables related to defense (defensive errors, tackles success (%), duel success (%), clearances) were significant. Two variables related to goal scoring (goal efficiency (%), shots from counter attack), and one variable related to passing and organizing (crosses) showed significant influence. No contextual variables were significant.

Significant defensive factors

The most influential success factor was defensive errors. Each defensive error decreases the probability of winning by 10.25% ($p < 0.001$). Despite being an intuitive result, defensive errors were rarely analyzed in recent studies, and this study permits a quantification of the impact. Lepschy et al. (2020) showed similar results for the German Bundesliga. The impact of errors in this study is slightly higher than the impact in the German Bundesliga. The operational definition of a defensive errors could also contribute to the big impact, "A mistake made by a player losing the ball that leads to a shot or a goal."(Opta, 2018). Losing the ball by a mistake usually also leaves the defense in an imbalanced status. Tenga, Holme et al. (2010) showed that playing against an imbalanced defense increase the chance of a goal for the attacking team. Several studies showed that the chance of a defensive errors is also increasing toward the end of a match because of physical deterioration and diminished cognitive function (Simiyu, 2014). The next significant factor related to defense is duel success in percentage, showing the third highest value of all significant success factors. Each additional percentage increases the chance of winning by 2.14% ($p < 0.001$). However, duel success has the lowest standard deviation of the following significant defensive factors and has the lowest range. Therefore, it could be argued that, despite the higher value, the positive effect of duel success is limited. Furthermore, the percentage of successful tackles was a significant positive success factor as well (0.57%; $p < 0.05$). However, previous research has yielded inconclusive results about whether the percentage of successful tackles is significant or not (Liu et al., 2015; Liu et al., 2016; Oberstone, 2009). Future research should investigate this further and focus on identifying possible interacting factors such as the location of the tackles or the direction of the tackles. Finally, clearances showed a significant positive effect (0.84%, $p < 0.05$) on success. This confirms previous research by Carmichael et al. (2000) and Lepschy et al. (2020). However, clearances were only

rarely included in past research. In the light of those results, future research should consider including clearances for an analysis of success factors in football.

Significant offensive factors

Besides the multitude of significant defensive factors, the analysis revealed that there also actions of offensive performance that can make the difference. Notably, each shot from a counter attack increased the chance of winning by 6.51% ($p < 0.05$). Moreover, the conversion of shots into goals showed to be a very important success factor. In agreement with previous research it was shown that goal efficiency has a significant positive effect on winning (Broich et al., 2014; Lepschy et al., 2020). A positive change of one percentage in goal efficiency increase the chance of winning by 1.93% ($p < 0.001$).

Significant factors related to passing and organizing

Crosses are the only significant variable related to passing and organizing. The number of crosses had a significant negative effect (-1.11%, $p < 0.05$). Again, this confirms previous research (Lago et al., 2010; Lepschy et al., 2020; Liu et al., 2015). The reason might be that only quantity and not quality of crosses was considered. This assumption is supported by a study which found that long passes are linked to losing ball possession (Reis et al., 2017). Unsuccessful crosses are likely to initiate a counterattack. Moreover, crosses from the midfield could be an indicator of limited technical and tactical skills or a compact defense of the opponent. Nevertheless, there is also indication of a positive effect for crosses (Oberstone, 2009). Hence, future research should consider the quality of crosses.

Non-significant factors in contrast to previous research

The effect of ball possession has been discussed controversially. It was not a significant predictor in past FIFA tournaments if other variables were included in the model (Collet, 2013). However, studies related to success factors in football leagues show ambiguous results (Collet, 2013; Liu et al., 2015; Liu et al., 2016). In this study ball possession showed no effect, supporting the assumption that ball possession is losing significant impact if the results are controlled for other influencing variables (Collet, 2013).

Interestingly, total shots, shots from inside 6-yard box and from inside penalty area did not affect the outcome of the games. Total shots and subgroups of shots (shots from inside 6-yard box and shots from inside penalty area) were widely studied in the past and the results showed mostly a significant positive effect on success (Lago et al., 2010; Lago et al., 2011; Lepschy et

al., 2020; Liu et al., 2015; Liu et al., 2016; Oberstone, 2009; Pappalardo & Cintia, 2018). However, our non-significant results might be due to goal efficiency and points towards the importance of precision over quantity of shots. In our analysis distance showed no effect, although a recent study identified it as the most influential variable in the German Bundesliga (Schauberger et al., 2017). However, in the latest study on the Bundesliga, including a wide range of variables, distance had also no effect on success (Lepschy et al., 2020). In contradiction to prior results of the German Bundesliga, market value was not a significant predictor of success (Lepschy et al., 2020). Seemingly, the market value of national teams at world cups is less important than in club football. A reason might be the different character of tournament games including single knock-out games to games played during a regular season. However, further research is needed to determine if this hypothesis can be supported. Other explanations could be that not enough matches with a distinct difference in market value were included or a mediator variable, which is not yet identified, is present.

In general, it showed that actions related to defense had a high impact of success in the last two world cups. Moreover, it appears that variables related to efficiency such as duel success (%), goal efficiency (%), and tackles success (%) are more important than the quantity of single factors, a finding that is supported by Collet (2013). Finally, ball possession seems to be of less importance also on a national team level. A more pressing/direct style, as reflected in defensive errors of the opponent and shots form counter attacks as well as duel and tackles success, seems to be more successful. This finding is in line with other studies (Pollard, 2019).

Practical implications

The results of this study have various implications for coaches of national teams but could also be helpful for coaches of club teams. Our findings point towards aspects that can make the difference at high-level football matches. Shot accuracy during matches is critical and should be properly addressed in training sessions. The development and utilization of apt training methods could be beneficial for the goal efficiency and eventually lead to more success. Accuracy instead of quantity should be the maxim. Furthermore, more effective ways to lower the probability of defensive errors should be found and implemented in specific training sessions. Next to technical and tactical skills, the improvement of endurance, speed and mental strength could be critical in this context. To increase the duel and tackles success rate, specific training methods could be utilized, and players should be focused on the importance of these factors in match preparations. Substitutions to accommodate for physical and mental fatigue of the starting formation can also contribute to a lower error rate and can help to win a match. Instead of

substituting forwards in during the second half, coaches could consider strengthening the defense through specific substitutions. On the tactical side, coaches should be aware of the significance of counterattacks especially when playing against stronger opponents. The play against an imbalanced defense can lead to more scoring opportunities especially if played at a faster pace (Almeida, 2019).

Limitations and future research

By interpreting the results of this study, three restrictions have to be taken into account. First, the sample size contained only matches of national teams during a tournament including only 128 matches. Therefore, the possible generalization of the results is limited. In addition, the sample consisted of matches from the group stages and knock out stages. The tactics used in the different stages could have interfered with the results. Second, the variable short passes was dropped in favor of reduced collinearity. Any effects of this variable were not accounted for. Third, the variable market value of the starting formation was gathered from a public website and is not a standardized factor.

With regard to future research, the study points towards several aspects that need further investigation. The influence of ball possession needs to be analyzed in more detail. This study showed no significant influence which is in agreement with previous research (e.g. Collet, 2013; Lepschy et al., 2020). However, other recent studies found a significant effect of ball possession but in opposite directions (e.g. Lago et al., 2011; Liu et al., 2015; Schauberger et al., 2017). Future research also needs to analyze the effects of the distance covered, since results are inconsistent. In addition, the negative impact of crosses should be analyzed. Lepschy et al. (2020) found similar results for the Bundesliga. It needs to be determined when crosses are a negative predictor and in which cases they are not. Moreover, the non-significant influence of shots, except shots from counter attack, should be investigated further to confirm previous results which showed a clear positive effect (e.g. Lago et al., 2010; Lepschy et al., 2020; Liu et al., 2015; Oberstone, 2009; Pappalardo & Cintia, 2018). Also, the effect of home advantage at World Cups also needs to be studied further considering crowd support, climate and possible influences of a "once in a lifetime experience" for players.
Methodologically, predictive analyses are the methods of choice. However, alternative methodological approaches such as social network analysis (Wäsche et al., 2017) should be considered. Social network analysis already revealed some new insights (Grund, 2012; Pina et al., 2017).

4.6. Conclusions

The study showed that defensive errors had the strongest influence on the probability of winning or losing a football match during the World Cups 2018 and 2014. In addition, goal efficiency, duel success in percentage and tackles success in percentage were shown be of high significance. It appears that efficiency factors are more important than single factors alone. Shots from counter attacks and clearances also revealed a positive impact. In contrast, the number of crosses showed a negative impact on winning. In total, four different variables related to defense, two variables related to goal scoring, one variable related to passing and organizing and no contextual variables were significant. Interestingly, shots from counterattacks, tackles and duel success are significant predictors of success whereas ball possession and passes are not significant. This supports the assumption that tactics dominated by pressing could be a better strategy than tactics solely based on ball possession. However, national teams and club teams cannot readily be compared due to different contexts such as the competition format. For example, market value of the starting formation was shown to be significant factors for the Bundesliga but not in the last two World Cups (Lepschy et al., 2020). Future research needs to determine possible differences and if those differences are significant like shots and shots on target or the market value of the starting formation. In addition, the ambiguous results for ball possession and number of crosses from different studies needs to be addressed in future research. Further research on success factors, building on existing knowledge and utilizing apt methods will further contributes to the knowledge of coaches, managers and other practitioners to improve team performance in football.

5. General Discussion and Conclusions

The goal of this thesis was to investigate the success factors in football. To this end, the state of current research was analyzed (Chapter 2). Based on the results of the review, key factors were subsequently analyzed empirically, using data from the German Bundesliga and the past two World Cups (Chapter 3 and Chapter 4).

In the first study (Chapter 2), it was shown that a broad variety of variables can influence success in a football match. Shots, including the subcategories shots on target, shots from counter attack and shots form inside the penalty area, have been widely proven to have a significant positive impact. Moreover, successful passes, pass accuracy and goal efficiency were also identified clearly by past research to have a positive effect on winning. Less clear were the effects of crosses and ball possession. Both have been shown with positive and negative effects as well as no significant effect. A clear negative effect was confirmed for a red card. Lastly, the contextual variables with a clear impact on winning in the past were home advantage and the quality of the opponent.

Besides the identified variables, a lack of predictive studies concerning success factor in football, especially in the German Bundesliga, as well as questionable sample sizes were revealed. Consequently, the second study dealt with three consecutive seasons of the German Bundesliga. It was found that goal efficiency, total shots, shots on target, shots counter attack and clearances had a positive effect on winning, whereas defensive errors and crosses had a negative effect. Interestingly, shots inside the penalty area only had a positive impact for home teams, while duel success (positive) and successful tackles (negative) were only significant for away teams. Also, the contextual variables home advantage (positive), quality of opponent (negative) and market value (positive) were significantly linked to success.

In the third study, matches of the two latest World Cups in 2014 and 2018 were analyzed. It was revealed that goal efficiency, shots from counter attack, clearances, duel success and tackles success rate were positively linked to winning a match, whereas defensive errors and crosses had a significant negative impact on winning. Interestingly, market value and quality of opponent, measured by the FIFA coefficient, did not have any significant influence.

5.1. Positive influence on winning

In the analysis of the German Bundesliga matches, is was shown that, if controlled for home advantage and quality of opponent, goal efficiency, shots from counter attacks, clearances, total shots, shots on target, total market value starting formation and shots from inside the penalty area are significant positive success factors for home teams. For away teams, goal efficiency, shots from counter attack, clearances, total shots, shots on target, total market value starting formation and duel success (%) are the significant positive success factors. In the World Cups 2014 and 2018 goal efficiency, shots from counter attack, clearances, duel success (%) and tackles success (%) were the significant positive success factors.

Significant variables in Bundesliga and World Cup

Based on these results, it seems that goal efficiency is one of the most influential success factors by increasing the chance of winning between 1.53% and 2.36% for each percentage point more in goal efficiency. This also supports the findings of Broich et al. (2014) and Yue et al. (2014), who reported similar results previously.

Furthermore, shots from counter attacks had a positive impact in both competitions and regardless of the match venue. Each additional shot from counter attack increased the probability of winning by 6.51% in the World Cups, 6.48% for home teams in the Bundesliga and 3.27% for away teams. This importance of the development of the shot was also reported by Liu et al. (2015). The results also confirm earlier studies form the Norwegian leagues (Tenga, Holme et al., 2010; Tenga, Ronglan et al., 2010) as well as the Spanish league (Lago-Ballesteros et al., 2012), which showed that counter attacks are more effective than regular offensive plays. In general, a more pressing/direct style looks more promising to be successful, which was also found in another study (Pollard, 2019).

Finally, clearances was the last variable with a significant positive impact on winning. Each additional clearance increases the chance of winning by 0.0068 to 0.0084. Taking into consideration that the number of clearances can vary by more than 60 in a game (see Table 12 and Table 16), this seemingly small effect can become substantial over the course of a match and shows the importance of good defensive work. This result also confirms previous research in the English league (Carmichael et al., 2000). However, the study in Chapter 4 was the first time that linked clearances to winning a match in World Cup tournaments. Considering these results, future research should also include clearances in the identification of success factor in football.

Significant variables only in Bundesliga or World Cups

There are also some variables, which only had significant influence in the Bundesliga as discussed in Chapter 3. Irrespective of the match venue, total shots and shots on target had a significant positive influence on winning for home teams and away teams with an effect size between 0.0085 and 0.0186 (see Table 13 and Table 14), confirming results of previous studies (Dufour et al., 2017; Lago et al., 2010; Liu et al., 2015; Liu et al., 2016; Lago, 2007; Oberstone, 2009; Pappalardo & Cintia, 2018; Yang et al., 2018). Interestingly both variables did not show a significant effect in the World Cups, which needs to be addressed in future research.

The newly introduced variable total market value of the starting formation also only had a positive influence in the Bundesliga. Future research needs to determine whether market value is not a factor in World Cups. It is possible, however, that not enough matches with a distinct difference in market value were included or that an unknown mediator variable obscured this effect. In previous research, financial figures have been shown to be of substantial influence (Coates et al., 2016; Kringstad & Olsen, 2016). The effect of market value in the Bundesliga was 0.0003 (home teams) and 0.0005 (away teams), showing a slightly higher influence for away teams. Even with an apparently small effect size, the influence can be important since the effect multiplies with every million Euros in difference between the two teams. The range between minimum and maximum value for the Bundesliga was more than 400 million Euros. In this case the probability of winning would increase by 12.3% (411x0.03%) and 20.6% (413x0.05%), respectively as discussed in Chapter 3.

Additionally, shots from inside the penalty area (home teams) and duel success (%) (away teams) only shows a positive effect for one of the opponents in the Bundesliga. The results for shots form inside the penalty area can be explained by the fact that home teams perform more offensive actions, which subsequently lead to a higher presence within the 18 yard box (Lago & Lago-Ballesteros, 2011; Poulter, 2009). As for away teams in the Bundesliga, duel success (%) also showed a significant positive influence on winning in the World Cups. The effect there was about three times higher than in away teams. However, duel success (%) also showed a low range between minimum and maximum, which limits the positive effect of duel success notably (see Table 12 and Table 16). Noteworthy, duel success (%) for home teams slightly missed the statistical significance level (p = 0.054). Additionally, home teams have a significant higher duel success (%) than away teams (see Table 12). In summary, duel success is a significant success factor in football matches, but its impact seems limited.

Lastly, tackles success (%) also influenced winning positively only in World Cups. This variable was included after the study of the Bundesliga because it was suggested that efficiency has a more positive effect than quantity alone. This was confirmed by the results of the World Cups. Did the amount of successful tackles had a negative effect for away teams in the Bundesliga, for the World Cups tackles success (%) has shown an increase in the probability of winning by 0.57%, for each percentage point increase in the tackle success rate. Nonetheless, previous research has produced mixed results about whether the percentage of successful tackles is significant or not (Liu et al., 2015; Oberstone, 2009; Schauberger et al., 2017). Future research should explore this effect further and put more emphasis on identifying interacting factors, such as the location of the tackles.

5.2. Negative influence on winning

In comparison with the above-mentioned variables with a positive effect on wining, there were only three variables that significantly influenced the chance of winning in a negative way. In the German Bundesliga as well as in the World Cups, defensive errors and crosses had a negative influence on winning. In addition, away teams in the Bundesliga also had a negative effect for the number of successful tackles.

Significant variables in Bundesliga and World Cup

Defensive errors were identified as the biggest single factor by effect size for the Bundesliga as well as the World Cups. The decrease in the winning probability ranged from 10.25% (World Cups) to 7.51% (Bundesliga away teams). Despite being an intuitive result, defensive errors were rarely analyzed in recent studies (see Chapter 2). However, the result can be also explained by the operational definition "A mistake made by a player losing the ball that leads to a shot or a goal." (Opta, 2018). In these conditions, the defense can be imbalanced since the team possessed the ball and already focused on the attack. Playing against an imbalanced defense also increases the chance of goal scoring (Tenga, Holme et al., 2010). Several studies also showed that the chance of defensive errors is increasing towards to the end of a match because of physical fatigue and diminished cognitive function (Abt et al., 2001; Simiyu, 2014).

Furthermore, crosses showed a consistent negative impact on winning across the Bundesliga and the World Cups. Each additional cross resulted in a decrease in the probability of winning between -0.0071 (Bundesliga away teams) and -0.012 (Bundesliga home teams). As before, the operational definition can provide an indication for the underlying reasons. Crosses are defined as: "Any ball sent by a player into the opposition team's area from a wide position" (Opta,

2018). Accuracy is not considered in this, leading to the point that the sheer sending of the ball into the opposition team's area does not say anything about the quality of the cross. This finding is also in line with previous studies. Lago et al. (2010) and Liu et al. (2015) revealed that more crosses are negatively linked to success. Additionally, Reis et al. (2017) similarly showed long distance passes are mostly not effective and result in losing ball possession. The result of unsuccessful crosses could be to concede a goal because the opponent starts a counterattack, which is an effective attack style as shown before. Nevertheless, there is also indication of a positive effect for crosses (Oberstone, 2009). Hence, future research should investigate further to clarify the effect of crosses and consider the quality of crosses as well as the following attack pattern of the opponent.

Significant variables only in Bundesliga for away teams

The sole variable that showed significant negative effects for only a subgroup of the studies was the number of successful tackles (see Table 14). It is possible that the action of a tackle, regardless of the success, indicates a high number of defensive actions, which can lead to a defeat because of unsuccessful tackles. The positive effect of the percentage of successful underlines this assumption. However, as stated earlier, the variable successful tackles was replaced in the study of the World Cup with the percentage of successful tackles (tackles success %), which led to a positive impact on winning but also limits the comparability of these variables.

5.3. Noteworthy non-significant influences

Despite the aim of this thesis is to find success factors in football, it is noteworthy that some variables did not emerge as significant success factor. First, ball possession has been ambiguously discussed in previous research as shown in Chapter 2. In the study of the Bundesliga as well as the World Cups, ball possession was not a significant factor. These results are in line with other studies which used a variety of variables (Collet, 2013; Liu et al., 2015). Collet (2013) stated that ball possession seems to loose influence if the study controls for other variables, especially team quality and offensive factors.

In addition, the variable distance, which has been discussed in past studies, did not show a significant influence in neither the Bundesliga nor the World Cups, and in contrast to earlier findings by Schauberger et al. (2017). They showed that distance is connected to match outcome, but they only analyzed eight variables. In contrast, Yang et al. (2018) showed that total distance without ball possession has no significant influence on winning. Moreover, Hoppe et

al. (2015) focused on match running performance, and showed that distance with ball possession is a significant predictor for accumulated points. Despite having shown that distance is not significant in this thesis, there needs to be further research to identify possible interactions, like distance with and without ball possession or difference in vertical distance and horizontal distance.

Lastly, the age of the starting formation did not exert a significant influence. This appears counterintuitive at first, but it can be explained by a well-distributed age structure (see Table 12 and Table 17). For example, a team could start with eleven players at the age of 25, which would result in an average age of 25. The opponent could also start with an average age of 25, but the age structure could be more divers. Therefore, the results show that the average age of the starting formation did not significantly impact winning or losing in the Bundesliga or World Cups, but it does not oppose the results of previous research (Baker & Tang, 2010).

5.4. Practical implications

The results can have several impacts for football teams and their coaches. The tendency that accuracy is a critical success factor may lead to a stronger emphasis on accuracy rather than on quantity (number of game actions such as shots, passes and tackles). Additionally, coordination, accuracy of shots, and tactical and physical ability to get into a favorable position (e.g., shot from within 6-yard box) are trainable skills that all increase goal efficiency. Defensive errors also have shown high influence on winning, underlining the importance of defensive actions. Furthermore, the research indicates that fitness (both physically and mentally) needs to be well trained to reduce the risk of errors (see also Njororai (2012) and (Njororai, 2013)). Substitutions to accommodate for physical and mental fatigue of the starting formation can also contribute to a lower error rate, especially in the defense. Shots from counterattacks are another important success factor, and match preparation should consider possible benefits of counterattacks. This could be relevant, for example, when playing against a team, which favors ball possession, or while playing against stronger opponents. The play against an imbalanced defense can lead to more scoring opportunities, especially if played at a faster pace (Almeida, 2019). Based on the findings it also seems indicated that coaches work on the quality of the team actions and do not focus solely on quantity. Finally, it is important to know for coaches and managers that external factors, like the market value and the venue of the games, must be considered to explain success. In this regard, it is also noteworthy that the differences in success factors between home and away teams can lead to different tactics, for example a stronger emphasis on duels during the training week before playing away.

5.5. Limitations and implications for future research

Limitations

By interpreting the results of this thesis, at least six limitations must be considered. First, the variables passes (Chapter 3) and short passes (Chapter 3 and 4) were removed due to multicollinearity. Therefore, no conclusion about those variables or their interactions was possible. The decision to drop those variables was also driven by the fact that pass accuracy remained in the analysis as a dimension of the accuracy of the passes played. Second, the data on market value is not a standardized factor, which can be easily assessed or counted. Third, the average age of the starting formation can be the same for two teams, but their age structure can be very different. Hence, an influence of age cannot be ruled out completely. Fourth, during the analyzed seasons of the Bundesliga, Bayern Munich was the dominating team of the Bundesliga and had the highest market value. This could have led to undisclosed interactions. Fifth, in the study on World Cups the sample size was limited to 128 matches. Therefore, the possible generalization of the results is limited. Sixth, the sample of the World Cups consisted of matches from the group stages and knock out stages. The tactics used in the different stages could have interfered with the results.

Future research

This thesis also revealed some topics that should be addressed in future research. First, it seems that defensive variables are more important in World Cups. This finding needs to be confirmed by future research, especially whether this is the cases for national teams in general. If so, the underlying reasons also need to be identified. Additionally, market value showed no significant effect for the World Cups despite having a substantial effect in the Bundesliga. The influence of ball possession needs to be analyzed in future research as well. In this thesis, no significant influence was revealed which is in agreement with some previous research (e.g. Collet, 2013). However, other recent studies found a significant positive effect of ball possession (e.g. Dufour et al., 2017; Lago et al., 2011; Liu et al., 2015; Schauberger et al., 2017). Moreover, future research needs to analyze the effects of the distance covered as well. Distance covered was identified as the most influential variable in recent studies on the German Bundesliga (Hoppe et al., 2015; Schauberger et al., 2017). This variable could be also be analyzed in more detail by considering whether the distance is covered with or without ball, how fast the distance was

covered, and whether the distance was covered vertically or horizontally. In addition, the negative impact of crosses should be analyzed. It needs to be determined when crosses are a negative predictor and in which cases they are not. The accuracy of the crosses could be an indicator for that. Finally, no significant effect of the average age of the starting formation was found in this thesis. However, as stated earlier this does not imply that age has no effect at all. Future research should analyze possible effect of the age structure of football teams, which could help coaches in the composition of their teams.

Methodologically, predictive analyses are the methods of choice. In both empirical studies, we have shown the superiority of those methods. Not only does it allow for a more sophisticated analysis, it also provides results that can be used to predict future performance. Furthermore, we overcame the issue of not fulfilling all assumptions of the method by using a generalized ordered logit approach. Nevertheless, alternative methodological approaches such as social network analysis (Wäsche et al., 2017) should be considered. Social network analysis already revealed new insights (Gonçalves et al., 2017; Grund, 2012; Mclean et al., 2018; Pina et al., 2017). Additionally, new variables like packing can provide new insights into the nature of success in football (Steiner et al., 2017). Finally, the availability of numerous data about football matches and players and the growing field of artificial intelligence can also lead to new discoveries in terms of success factors.

5.6. Conclusions

Our research revealed four novel insights. First, of the 29 variables examined, goal efficiency, shots from counter attack and clearances were found to have a positive effect on winning, whereas defensive errors and crosses had a negative influence. Second, significant variables had different effect sizes, ranging from < 1 % to > 10 %. It therefore seems that accuracy and quality is more important than quantity. Third, some variables were only significant predictors for winning by either the home or away team. Fourth, some variables only exhibited significant effects either in the Bundesliga or during the World Cup matches.

Overall, my thesis contributes to a deeper understanding of success factors in football and provides new insights into previously unobserved variables. A novel methodological approach has been utilized to identify the significant performance factors in the German Bundesliga as well as in national teams during World Cups. In both circumstances, it was the first time this method was applied to a vast set of performance factors.

6. References

Abt, G. A., Dickson, G., & Mummery, W. K. (2001). Goal Scoring Patterns Over the Course of a Match: An Analysis of the Australian National Soccer League. In W. Spinks, T. Reilly, & A. Murphy (Eds.), *Science and football IV* (Vol. 4, pp. 106–111). Psychology Press.

Almeida, C. H. (2019). Comparison of successful offensive sequences in the group stage of 2018 FIFA World Cup: Eliminated vs. qualified teams. *Science and Medicine in Football*, *1*(23), 1–7. https://doi.org/10.1080/24733938.2019.1613557

Apter, M. J. (1984). Reversal theory and personality: A review. *Journal of Research in Personality*, *18*(3), 265–288.

Armatas, V., & Pollard, R. (2014). Home advantage in Greek football. *European Journal of Sport Science*, *14*(2), 116–122. https://doi.org/10.1080/17461391.2012.736537

Armatas, V., Yiannakos, A., Papadopoulou, S., & Skoufas, D. (2009). Evaluation of goals scored in top ranking soccer matches: Greek "Super League" 2006-07. *Serbian Journal of Sports Sciences*, *3*(1), 39–43.

Armatas, V., Yiannakos, A., Zaggelidis, G., Skoufas, D., Papadopoulou, S., & Fragkos, N. (2009). Differences in offensive actions between top and last teams in Greek first soccer division. A retrospective study 1998-2008. *Citius Altius Fortius*, *23*(2), 40–45.

Baker, A. B., & Tang, Y. Q. (2010). Aging performance for masters records in athletics, swimming, rowing, cycling, triathlon, and weightlifting. *Experimental Aging Research*, *36*(4), 453–477. https://doi.org/10.1080/0361073X.2010.507433

Bar-Eli, M., Tenenbaum, G., & Geister, S. (2006). Consequences of players' dismissal in professional soccer: A crisis-related analysis of group-size effects: A crisis-related analysis of group-size effects. *Journal of Sports Sciences*, *24*(10), 1083–1094. https://doi.org/10.1080/02640410500432599

Bekris, E., Gioldasis, A., Gissis, I., Komsis, S., & Alipasali, F. (2014). Winners and losers in top level soccer. How do they differ? *Journal of Physical Education and Sport*, *14*(3), 398.

Bekris, E., Mylonis, E., Sarakinos, A., Gissis, I., Gioldasis, A., & Sotiropoulos, A. (2013). Offense and defense statistical indicators that determine the Greek Superleague teams placement on the table 2011-12. *Journal of Physical Education and Sport*, *13*(3), 338–347. https://doi.org/10.7752/jpes.2013.03055

References

Bishop, D. (2008). An applied research model for the sport sciences. *Sports Medicine*, *38*(3), 253–263. https://doi.org/10.2165/00007256-200838030-00005

Borrie, A., Jonsson, G. K., & Magnusson, M. S. (2002). Temporal pattern analysis and its applicability in sport: An explanation and exemplar data: An explanation and exemplar data. *Journal of Sports Sciences*, *20*(10), 845–852. https://doi.org/10.1080/026404102320675675

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 1171–1178.

Broich, H., Mester, J., Seifriz, F., & Yue, Z. (2014). Statistical Analysis for the First Bundesliga in the Current Soccer Season. *Progress in Applied Mathematics*, *7*(2), 1–8. https://doi.org/10.3968/4886

Cameron, A. C., & Trivedi, P. K. (2010). *Microeconometrics using stata*. Stata press College Station, TX.

Carling, C., Williams, M., & Reilly, T. (2005). *Handbook of soccer match analysis: A systematic approach to improving performance*. Routledge. http://www.carlocapelli.it/ftp/didattica/Specialistica/Bibliografia/Sprint_running/Osgnach%20et%20al_MSSE_2010.pdf

Carmichael, F., & Thomas, D. (2005). Home-Field Effect and Team Performance. *Journal of Sports Economics*, *6*(3), 264–281. https://doi.org/10.1177/1527002504266154

Carmichael, F., Thomas, D., & Ward, R. (2000). Team performance: The case of English Premiership football. *Managerial and Decision Economics*, *21*(1), 31–45. https://doi.org/10.1002/1099-1468(200001/02)21:1<31::AID-MDE963>3.0.CO;2-Q

Carron, A. V., Bray, S. R., & Eys, M. A. (2002). Team cohesion and team success in sport. *Journal of Sports Sciences*, *20*(2), 119–126. https://doi.org/10.1080/026404102317200828

Castellano, J., Casamichana, D., & Lago, C. (2012). The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. *Journal of Human Kinetics*, *31*, 139–147. https://doi.org/10.2478/v10078-012-0015-7

Clarke, S. R., & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *The Statistician*, 509–521.

Clemente, F. M. (2012). Study of successful teams on FIFA World Cup 2010 through notational analysis. *Pamukkale Journal of Sport Sciences*, *3*(3), 90–103.

Coates, D., Frick, B., & Jewell, T. (2016). Superstar Salaries and Soccer Success. *Journal of Sports Economics*, *17*(7), 716–735. https://doi.org/10.1177/1527002514547297

Collet, C. (2013). The possession game? A comparative analysis of ball retention and team success in European and international football, 2007-2010. *Journal of Sports Sciences*, *31*(2), 123–136. https://doi.org/10.1080/02640414.2012.727455

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, *14*(3), 391–403. https://doi.org/10.1081/QEN-120001878

Delgado-Bordonau, J. L., Domenech-Monforte, C., Guzmán, J. F., & Méndez-Villanueva, A. (2013). Offensive and defensive team performance: Relation to successful and unsuccessful participation in the 2010 Soccer World Cup. *Journal of Human Sport and Exercise*, *8*(4), 894–904. https://doi.org/10.4100/jhse.2013.84.02

Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *46*(2), 265–280.

Drust, B. (2019). Applied science and soccer: a personal perspective on the past, present and future of a discipline. *Sport Performance & Science Reports*, *56*. https://sportperf-sci.com/wp-content/uploads/2019/04/SPSR63_Drust_190326_final.pdf

Drust, B., Favero, T., & Dawson, B. (Eds.). (2015). *International research in science and soccer*. Routledge.

Duch, J., Waitzman, J. S., & Amaral, L. A. N. (2010). Quantifying the performance of individual players in a team activity. *PloS One*, *5*(6), e10937. https://doi.org/10.1371/journal.pone.0010937

Dufour, M., Phillips, J., & Ernwein, V. (2017). What makes the difference? Analysis of the 2014 World Cup. *Journal of Human Sport and Exercise*, *12*(3). https://doi.org/10.14198/jhse.2017.123.06

Dufour W. (1993). Computer-assisted scouting in soccer. In T. Reilly, J. Clarys, & A. Stibbe (Eds.), *Science and Football II: Proceedings of the Second World Congress of Science and Football, Eindhoven, Netherlands, 22nd-25th May, 1991* (1st ed., pp. 160–166). Taylor & Francis.

Eagly, A. H., & Wood, W. (1994). Using research syntheses to plan future research. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis* (pp. 485–500). Russell Sage Foundation.

References

Eaves, J. S. (2017). A history of sports notational analysis: A journey into the nineteenth century. *International Journal of Performance Analysis in Sport*, *15*(3), 1160–1176. https://doi.org/10.1080/24748668.2015.11868859

Ender, P. B. (2010). Collin: Collinearity diagnostics. *Institute for Digital Research and Education, University of California, Los Angeles*. http://www.ats.ucla.edu/stat/stata/ado/analysis/default.htm

FIFA. (2007). *History of Football - The Origins*. https://www.fifa.com/about-fifa/who-we-are/the-game/index.html

FIFA. (2014). *FIFA world ranking: 05. June 2014*. https://www.fifa.com/fifa-world-ranking/ranking-table/men/rank=239/index.html

FIFA. (2015). *2014 FIFA World Cup Brazil. Televison Audience Report.* FIFA. https://resources.fifa.com/mm/document/affederation/tv/02/74/55/57/2014fwcbraziltvaudiencereport(draft5)(issuedate14.12.15)_neutral.pdf

FIFA. (2018). *FIFA world ranking*. https://www.fifa.com/fifa-world-ranking/ranking-table/men/rank=287/index.html

FIFA Communications Divisions. (2007). *FIFA Big Count 2006: 270 million people active in football*. http://www.fifa.com/mm/document/fifafacts/bcoffsurv/bigcount.statspackage_7024.pdf

Frick, B. (2011). Performance, Salaries and Contract Length: Empirical Evidence from German Soccer. *International Journal of Sport Finance*, *6*(2), 87–118.

Fullerton, A. S. (2009). A Conceptual Framework for Ordered Logistic Regression Models. *Sociological Methods & Research*, *38*(2), 306–347. https://doi.org/10.1177/0049124109346162

García-Rubio, J., Gómez, M. Á., Lago, C., & Ibáñez, J. S. (2017). Effect of match venue, scoring first and quality of opposition on match outcome in the UEFA Champions League. *International Journal of Performance Analysis in Sport*, *15*(2), 527–539. https://doi.org/10.1080/24748668.2015.11868811

Gásquez, R., & Royuela, V. (2016). The Determinants of International Football Success: A Panel Data Analysis of the Elo Rating *Social Science Quarterly*, *97*(2), 125–141. https://doi.org/10.1111/ssqu.12262

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, *21*(2), 331–340. https://doi.org/10.1016/j.ijforecast.2004.08.002

Göke, S., Prinz, J., & Weimar, D. (2014). Diamonds are Forever: Job-Matching and Career Success of Young Workers. *Journal of Economics and Statistics (Jahrbuecher Fuer Nationaloekonomie Und Statistik)*, *234*(4), 450–473.

Gómez, M. Á., DelaSerna, A., Lupo, C., & Sampaio, J. (2014). Effects of Situational Variables and Starting Quarter Score in the outcome of elite women's water polo game quarters. *International Journal of Performance Analysis in Sport*, *14*(1), 73–83. https://doi.org/10.1080/24748668.2014.11868704

Gómez, M. Á., Gasperi, L., & Lupo, C. (2017). Performance analysis of game dynamics during the 4 th game quarter of NBA close games. *International Journal of Performance Analysis in Sport*, *16*(1), 249–263. https://doi.org/10.1080/24748668.2016.11868884

Gómez, M. Á., Gómez-Lopez, M., Lago, C., & Sampaio, J. (2012). Effects of game location and final outcome on game-related statistics in each zone of the pitch in professional football. *European Journal of Sport Science*, *12*(5), 393–398. https://doi.org/10.1080/17461391.2011.566373

Gonçalves, B., Coutinho, D., Santos, S., Lago, C., Jiménez, S., & Sampaio, J. (2017). Exploring team passing networks and player movement dynamics in youth association football. *PloS One*, *12*(1), e0171156.

Goumas, C. (2014a). Home advantage and referee bias in European football. *European Journal of Sport Science*, *14 Suppl 1*, S243-9. https://doi.org/10.1080/17461391.2012.686062

Goumas, C. (2014b). Home advantage in Australian soccer. *Journal of Science and Medicine in Sport*, *17*(1), 119–123. https://doi.org/10.1016/j.jsams.2013.02.014

Goumas, C. (2014c). Tyranny of distance: Home advantage and travel in international club football. *International Journal of Performance Analysis in Sport*, *14*(1), 1–13. https://doi.org/10.1080/24748668.2014.11868698

Goumas, C. (2015). Modelling home advantage for individual teams in UEFA Champions League football. *Journal of Sport and Health Science*, *6*(3), 321–326. https://doi.org/10.1016/j.jshs.2015.12.008

Graham, M. H. (2003). Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, *84*(11), 2809–2815. https://doi.org/10.1890/02-3114

References

Greene, W. H. (2011). *Econometric analysis* (Seventh edition). Pearson.

Grund, T. U. (2012). Network structure and team performance: The case of English Premier League soccer teams. *Social Networks*, *34*(4), 682–690. https://doi.org/10.1016/j.soc-net.2012.08.004

Hall, S., Szymanski, S., & Zimbalist, A. S. (2016). Testing Causality Between Team Performance and Payroll. *Journal of Sports Economics*, *3*(2), 149–168. https://doi.org/10.1177/152700250200300204

Hanau, T. von, Wicker, P., & Soebbing, B. P. (2014). Determinants of expected vs. actual match outcome: An examination of the German Bundesliga. *Soccer & Society*, *16*(1), 63–75. https://doi.org/10.1080/14660970.2014.882823

Harrop, K., & Nevill, A. (2017). Performance indicators that predict success in an English professional League One soccer team. *International Journal of Performance Analysis in Sport*, *14*(3), 907–920. https://doi.org/10.1080/24748668.2014.11868767

Higham, D. G., Hopkins, W. G., Pyne, D. B., & Anson, J. M. (2014). Performance indicators related to points scoring and winning in international rugby sevens. *Journal of Sports Science & Medicine*, *13*(2), 358.

Hopkins, W. G., Marshall, S. W., Batterham, A. M., & Hanin, J. (2009). Progressive statistics for studies in sports medicine and exercise science. *Medicine and Science in Sports and Exercise*, *41*(1), 3–13. https://doi.org/10.1249/MSS.0b013e31818cb278

Hoppe, M. W., Slomka, M., Baumgart, C., Weber, H., & Freiwald, J. (2015). Match Running Performance and Success Across a Season in German Bundesliga Soccer Teams. *International Journal of Sports Medicine*, *36*(07), 563–566. https://doi.org/10.1055/s-0034-1398578

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary versus secondary. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 593–599). Elsevier Academic Press.

Hughes, M. (1988). Computerized notation analysis in field games. *Ergonomics*, *31*(11), 1585–1592. https://doi.org/10.1080/00140138808966808

Hughes, M. (2003). Notational analysis. In T. Reilly & M. Williams (Eds.), *Science and soccer* (pp. 343–361). Routledge.

Hughes, M., & Bartlett, R. M. (2002). The use of performance indicators in performance analysis. *Journal of Sports Sciences*, *20*(10), 739–754. https://doi.org/10.1080/026404102320675602

Hughes, M., & Franks, I. (2004). Notational analysis: A review of the literature. In M. Hughes & I. Franks (Eds.), *Notational analysis of sport: Systems for better coaching and performance in sport* (pp. 57–102). Psychology Press.

Hughes, M., & Franks, I. (2005). Analysis of passing sequences, shots and goals in soccer. *Journal of Sports Sciences*, *23*(5), 509–514. https://doi.org/10.1080/02640410410001716779

Hughes, M., Hughes, M. T., & Behan, H. (2007). The evolution of computerised notational analysis through the example of racket sports. *International Journal of Sports Science and Engineering*, *1*(1), 3–28.

James, N. (2012). Predicting performance over time using a case study in real tennis. *Journal of Human Sport and Exercise*, *7*(2), 421–433. https://doi.org/10.4100/jhse.2012.72.08

Janković, A., Leontijević, B., Jelušić, V., Pašić, M., & Mićović, B. (2011). Influence of tactics efficiency on results in serbian soccer super league in season 2009/2010. *Citius Altius Fortius*, *11*(1).

Janković, A., Leontijević, B., Pašić, M., & Jelušić, V. (2011). Influence of certain tactical attacking patterns on the result achieved by the teams participants of the 2010 FIFA World Cup in South Africa. *Fizička Kultura*, *65*(1), 34–45.

Kapidžić, A., Bečirović, E., & Imamovic, J. (2009). Situational efficiency analysis of the teams that participated in 2008 European football championship. *Sport Scientific & Practical Aspects*, *6*(1).

Kapidžić, A., Mejremić, E., Bilalić, J., & Bečirović, E. (2010). Differences in some parameters of situation efficiency between winning and defeated teams at two levels of competition. *Sport Scientific & Practical Aspects*, *7*(2).

Kleinbaum, D. G., & Klein, M. (2010). *Logistic Regression*. Springer New York. https://doi.org/10.1007/978-1-4419-1742-3

Kringstad, M., & Olsen, T.-E. (2016). Can sporting success in Norwegian football be predicted from budgeted revenues? *European Sport Management Quarterly*, *16*(1), 20–37. https://doi.org/10.1080/16184742.2015.1061032

References

Lago, C. (2007). Are winners different from losers? Performance and chance in the FIFA World Cup Germany 2006. *International Journal of Performance Analysis in Sport*, *7*(2), 36–47. https://doi.org/10.1080/24748668.2007.11868395

Lago, C., Gómez-Ruano, M., Megías-Navarro, D., & Pollard, R. (2016). Home advantage in football: Examining the effect of scoring first on match outcome in the five major European leagues. *International Journal of Performance Analysis in Sport*, *16*(2), 411–421. https://doi.org/10.1080/24748668.2016.11868897

Lago, C., & Lago-Ballesteros, J. (2011). Game location and team quality effects on performance profiles in professional soccer. *Journal of Sports Science & Medicine*, *10*(3), 465–471.

Lago, C., Lago-Ballesteros, J., Dellal, A., & Gómez, M. (2010). Game-related statistics that discriminated winning, drawing and losing teams from the Spanish soccer league. *Journal of Sports Science & Medicine*, *9*(2), 288–293.

Lago, C., Lago-Ballesteros, J., & Rey, E. (2011). Differences in performance indicators between winning and losing teams in the UEFA Champions League. *Journal of Human Kinetics*, *27*(1), 137–148. https://doi.org/10.2478/v10078-011-0011-3

Lago, C., & Martín, R. (2007). Determinants of possession of the ball in soccer. *Journal of Sports Sciences*, *25*(9), 969–974. https://doi.org/10.1080/02640410600944626

Lago-Ballesteros, J., & Lago, C. (2010). Performance in Team Sports: Identifying the Keys to Success in Soccer. *Journal of Human Kinetics*, *25*(1), 85–91. https://doi.org/10.2478/v10078-010-0035-0

Lago-Ballesteros, J., Lago, C., & Rey, E. (2012). The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *Journal of Sports Sciences*, *30*(14), 1455–1461. https://doi.org/10.1080/02640414.2012.712715

Lepschy, H., Wäsche, H., & Woll, A. (2018). How to be Successful in Football: A Systematic Review. *The Open Sports Sciences Journal*, *11*(1), 3–23. https://doi.org/10.2174/1875399X01811010003

Lepschy, H., Wäsche, H., & Woll, A. (2020). Success factors in football: An Analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport.* Advance online publication. https://doi.org/10.1080/24748668.2020.1726157

Liu, H., Gómez, M. Á., Lago, C., & Sampaio, J. (2015). Match statistics related to winning in the group stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, *33*(12), 1205–1213. https://doi.org/10.1080/02640414.2015.1022578

Liu, H., Hopkins, W. G., & Gómez, M. Á. (2016). Modelling relationships between match events and match outcome in elite football. *European Journal of Sport Science*, *16*(5), 516–525. https://doi.org/10.1080/17461391.2015.1042527

Liu, H., Hopkins, W. G., Gómez, M. Á., & Molinuevo, S. J. (2013). Inter-operator reliability of live football match statistics from OPTA Sportsdata. *International Journal of Performance Analysis in Sport*, *13*(3), 803–821. https://doi.org/10.1080/24748668.2013.11868690

Low, B., Coutinho, D., Gonçalves, B., Rein, R., Memmert, D., & Sampaio, J. (2019). A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Medicine.* Advance online publication. https://doi.org/10.1007/s40279-019-01194-7

Luhtanen, P., Belinskij, A., Häyrinen, M., & Vänttinen, T. (2001). A comparative tournament analysis between the EURO 1996 and 2000 in soccer. *International Journal of Performance Analysis in Sport*, *1*(1), 74–82. https://doi.org/10.1080/24748668.2001.11868250

Lupo, C., Condello, G., Capranica, L., & Tessitore, A. (2014). Women's water polo World Championships: Technical and tactical aspects of winning and losing teams in close and unbalanced games. *The Journal of Strength & Conditioning Research*, *28*(1), 210–222.

Lupo, C., & Tessitore, A. (2016). How Important is the Final Outcome to Interpret Match Analysis Data: The Influence of Scoring a Goal, and Difference Between Close and Balance Games in Elite Soccer: Comment on Lago-Penas and Gomez-Lopez (2014): The Influence of Scoring a Goal, and Difference Between Close and Balance Games in Elite Soccer: Comment on Lago-Penas and Gomez-Lopez (2014). *Perceptual and Motor Skills*, *122*(1), 280–285. https://doi.org/10.1177/0031512515626629

Mackenzie, R., & Cushion, C. (2013). Performance analysis in football: A critical review and implications for future research: A critical review and implications for future research. *Journal of Sports Sciences*, *31*(6), 639–676. https://doi.org/10.1080/02640414.2012.746720

Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118.

References

Mao, L., Peng, Z., Liu, H., & Gómez, M. Á. (2016). Identifying keys to win in the Chinese professional soccer league. *International Journal of Performance Analysis in Sport*, *16*(3), 935–947. https://doi.org/10.1080/24748668.2016.11868940

Marcelino, R., Sampaio, J., & Mesquita, I. (2011). Investigação centrada na análise do jogo: Da modelação estática à modelação dinâmica. *Revista Portuguesa De Ciências Do Desporto*, *11*(1), 125–152. https://doi.org/10.5628/rpcd.11.01.125

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *42*(2), 109–127. https://doi.org/10.1111/j.2517-6161.1980.tb01109.x

McFadden, D. (1977). *Quantitative methods for analyzing travel behavior of individuals: Some recent developments*. Institute of Transportation Studies, University of California.

Mclean, S., Salmon, P. M., Gorman, A. D., Stevens, N. J., & Solomon, C. (2018). A social network analysis of the goal scoring passing networks of the 2016 European Football Championships. *Human Movement Science*, *57*, 400–408. https://doi.org/10.1016/j.humov.2017.10.001

Mechtel, M., Bäker, A., Brändle, T., & Vetter, K. (2011). Red cards: Not such bad news for penalized guest teams. *Journal of Sports Economics*, *12*(6), 621–646. https://doi.org/10.1177/1527002510388478

Memmert, D., & Rein, R. (2018). Match Analysis, Big Data and Tactics: Current Trends in Elite Soccer. *German Journal of Sports Medicine/Deutsche Zeitschrift Fur Sportmedizin*, *69*(3).

Mitrotasios, M., Gonzalez-Rodenas, J., Armatas, V., & Aranda, R. (2019). The creation of goal scoring opportunities in professional soccer. Tactical differences between Spanish La Liga, English Premier League, German Bundesliga and Italian Serie A. *International Journal of Performance Analysis in Sport*, *19*(3), 452–465. https://doi.org/10.1080/24748668.2019.1618568

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement: The PRISMA statement. *PLoS Medicine*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moura, F. A., Martins, L. E. B., & Cunha, S. A. (2014). Analysis of football game-related statistics using multivariate techniques. *Journal of Sports Sciences*, *32*(20), 1881–1887. https://doi.org/10.1080/02640414.2013.853130

Myers, R. H. (1990). *Classical and modern regression with applications*. Duxbury press.

Nash, C., & Collins, D. (2006). Tacit Knowledge in Expert Coaching: Science or Art? *Quest*, *58*(4), 465–477. https://doi.org/10.1080/00336297.2006.10491894

Njororai, W. W.S. (2012). Physical demands of soccer: Lessons from team USA and Ghana matches in the 2010 FIFA WORLD CUP. *Journal of Physical Education and Sport*, *12*(4), 407.

Njororai, W. W.S. (2013). Analysis of goals scored in the 2010 world cup soccer tournament held in South Africa. *Journal of Physical Education and Sport*, *13*(1), 6.

O'Donoghue, P. (2007). Reliability Issues in Performance Analysis. *International Journal of Performance Analysis in Sport*, *7*(1), 35–48. https://doi.org/10.1080/24748668.2007.11868386

O'Donoghue, P. (2009). *Research Methods for Sports Performance Analysis*. Routledge. https://doi.org/10.4324/9780203878309

Oberstone, J. (2009). Differentiating the Top English Premier League Football Clubs from the Rest of the Pack: Identifying the Keys to Success. *Journal of Quantitative Analysis in Sports*, *5*(3). https://doi.org/10.2202/1559-0410.1183

Opta. (2018). *Opta's event definitions*. https://www.optasports.com/news/opta-s-event-definitions/

Papahristodoulou, C. (2007). An analysis of Champions League match statistics. *International Journal of Applied Sports Sciences*, *20*(1), 67–93.

Pappalardo, L., & Cintia, P. (2018). Quantifiying the relation between Performance and success in soccer. *Advances in Complex Systems*, *21*(03n04), 1750014. https://doi.org/10.1142/S021952591750014X

Pina, T. J., Paulo, A., & Araújo, D. (2017). Network Characteristics of Successful Performance in Association Football. A Study on the UEFA Champions League. *Frontiers in Psychology*, *8*, 534. https://doi.org/10.3389/fpsyg.2017.01173

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis: A retrospective analysis. *Journal of Sports Sciences*, *4*(3), 237–248. https://doi.org/10.1080/02640418608732122

Pollard, R. (2006). Worldwide regional variations in home advantage in association football. *Journal of Sports Sciences*, *24*(3), 231–240. https://doi.org/10.1080/02640410500141836

References

Pollard, R. (2008). Home Advantage in Football: A Current Review of an Unsolved Puzzle. *The Open Sports Sciences Journal*, *1*(1), 12–14. https://doi.org/10.2174/1875399X00801010012

Pollard, R. (2019). Invalid Interpretation of Passing Sequence Data to Assess Team Performance in Football: Repairing the Tarnished Legacy of Charles Reep. *The Open Sports Sciences Journal*, *12*(1), 17–21. https://doi.org/10.2174/1875399X01912010017

Pollard, R., & Gómez, M. Á. (2009). Home advantage in football in South-West Europe: Long-term trends, regional variation, and team differences. *European Journal of Sport Science*, *9*(6), 341–352. https://doi.org/10.1080/17461390903009133

Pollard, R., & Pollard, G. (2005). Long-term trends in home advantage in professional team sports in North America and England (1876-2003). *Journal of Sports Sciences*, *23*(4), 337–350. https://doi.org/10.1080/02640410400021559

Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *46*(4), 541–550. https://doi.org/10.1111/1467-9884.00108

Pollard, R., Silva, C. D., & Medeiros, N. C. (2008). Home advantage in football in Brazil: Differences between teams and the effects of distance traveled. *The Brazilian Journal of Soccer Science*, *1*(1), 3–10.

Poulter, D. R. (2009). Home advantage and player nationality in international club football. *Journal of Sports Sciences*, *27*(8), 797–805. https://doi.org/10.1080/02640410902893364

Rampinini, E., Impellizzeri, F. M., Castagna, C., Coutts, A. J., & Wisløff, U. (2009). Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, *12*(1), 227–233. https://doi.org/10.1016/j.jsams.2007.10.002

Read, B., & Edwards, P. (1992). Teaching children to play games. *Leeds: White Line Publishing*.

Reep, C., & Benjamin, B. (1968). Skill and Chance in Association Football. *Journal of the Royal Statistical Society. Series a (General)*, *131*(4), 581. https://doi.org/10.2307/2343726

Reilly, T., Lees, A., Davids, K., & Murphy, W. J. (Eds.). (2011). *Routledge revivals. Science and football: Proceedings of the first World Congress of Science and Football Liverpool, 13-17th April 1987*. Routledge.

Reilly, T., & Thomas, V. (1976). A motion analysis of work-rate in different positional roles in professional football match-play. *Journal of Human Movement Studies*, *2*, 87–97.

Reilly, T., & Williams, M. (2003). Introduction to science and soccer. In T. Reilly & M. Williams (Eds.), *Science and soccer* (pp. 1–6). Routledge.

Reis, M. A. M. d., Vasconcellos, F. V. d. A., & Almeida, M. B. d. (2017). Analysis of the effectiveness of long distance passes in 2014 Brazil FIFA World Cup. *Revista Brasileira De Cineantropometria & Desempenho Humano*, *19*(6), 676–685.

Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (the Statistician)*, *49*(3), 399–418.

Ruiz-Ruiz, C., Fradua, L., Fernández-García, A., & Zubillaga, A. (2013). Analysis of entries into the penalty area as a performance indicator in soccer. *European Journal of Sport Science*, *13*(3), 241–248. https://doi.org/10.1080/17461391.2011.606834

Saavedra García, M., Gutiérrez Aguilar, O., Fernández Romero, J. J., & Sa Marques, P. (2015). Measuring Home Advantage In Spanish Football (1928-2011). *Revista Internacional De Medicina Y Ciencias De La Actividad Física Y Del Deporte*, *57*, 181–194. https://doi.org/10.15366/rimcafd2015.57.010

Sampaio, J., Lago, C., Casais, L., & Leite, N. (2010). Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *European Journal of Sport Science*, *10*(6), 391–396.

Sánchez, P. A., García-Calvo, T., Leo, F. M., Pollard, R., & Gómez, M. Á. (2009). An analysis of home advantage in the top two Spanish professional football leagues. *Perceptual and Motor Skills*, *108*(3), 789–797. https://doi.org/10.2466/PMS.108.3.789-797

Sarmento, H., Marcelino, R., Anguera, M. T., Campanico, J., Matos, N., & Leitao, J. C. (2014). Match analysis in football: A systematic review: A systematic review. *Journal of Sports Sciences*, *32*(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852

Schauberger, G., Groll, A., & Tutz, G. (2017). Analysis of the importance of on-field covariates in the German Bundesliga. *Journal of Applied Statistics*, *11*, 1–18. https://doi.org/10.1080/02664763.2017.1383370

Seçkin, A., & Pollard, R. (2008). Home advantage in Turkish professional soccer. *Perceptual and Motor Skills*, *107*(1), 51–54. https://doi.org/10.2466/pms.107.1.51-54

References

Simiyu, W. W. N. (2014). Timing of Goals Scored in Selected European and South American Football Leagues, FIFA and UEFA Tournaments and the Critical Phases of a Match. *International Journal of Sports Science*, *4*(6A), 56–64. https://doi.org/10.5923/s.sports.201401.08

Steiner, S., Rauh, S., Rumo, M., Emery, N., Sonderegger, K., & Seiler, R. (2017). *Packing in football: A differential ecological perspective on passes*. Essen. https://doi.org/10.7892/BORIS.111947

Sternberg, R. J. (2003). *Wisdom, Intelligence, and Creativity Synthesized*. Cambridge University Press. https://doi.org/10.1017/CBO9780511509612

Szwarc, A. (2007). Efficacy of successful and unsuccessful soccer teams taking part in finals of Champions League. *Research Yearbook*, *13*(2), 221–225.

Tenga, A., Holme, I., Ronglan, L., & Bahr, R. (2010). Effect of playing tactics on goal scoring in Norwegian professional soccer. *Journal of Sports Sciences*, *28*(3), 237–244. https://doi.org/10.1080/02640410903502774

Tenga, A., Ronglan, L., & Bahr, R. (2010). Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science*, *10*(4), 269–277. https://doi.org/10.1080/17461390903515170

Tenga, A., & Sigmundstad, E. (2017). Characteristics of goal-scoring possessions in open play: Comparing the top, in-between and bottom teams from professional soccer league. *International Journal of Performance Analysis in Sport*, *11*(3), 545–552. https://doi.org/10.1080/24748668.2011.11868572

Thomas, S., Reeves, C., & Davies, S. (2004). An analysis of home advantage in the English Football Premiership. *Perceptual and Motor Skills*, *99*(3 Pt 2), 1212–1216. https://doi.org/10.2466/pms.99.3f.1212-1216

Torgler, B. (2004). The Economics of the FIFA Football Worldcup. *Kyklos*, *57*(2), 287–300. https://doi.org/10.1111/j.0023-5962.2004.00255.x

Torgler, B., & Schmidt, S. L. (2007). What shapes player performance in soccer? Empirical findings from a panel analysis. *Applied Economics*, *39*(18), 2355–2369. https://doi.org/10.1080/00036840600660739

Transfermarkt.de. (2017). *Marktorientierung und Definition*. https://www.transfermarkt.de/neu-marktorientierung-und-definition/thread/forum/67/thread_id/235745

UEFA. (2015). *Berlin final captures the world's imagination*. http://www.uefa.com/uefachampionsleague/news/newsid=2255318.html

UEFA. (2019). *Association club coefficients - Country coefficients.* UEFA. https://www.uefa.com/memberassociations/uefarankings/country/#/yr/2020

Vaz, L., van Rooyen, M., & Sampaio, J. (2010). Rugby game-related statistics that discriminate between winning and losing teams in IRB and Super twelve close games. *Journal of Sports Science & Medicine*, *9*(1), 51–55.

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), 1076–1083. https://doi.org/10.1080/02640414.2013.879671

Wäsche, H., Dickson, G., Woll, A., & Brandes, U. (2017). Social network analysis in sport research: An emerging paradigm. *European Journal for Sport and Society*, *14*(2), 138–165. https://doi.org/10.1080/16138171.2017.1318198

Weiss, A. A. (1997). Specification tests in ordered logit and probit models. *Econometric Reviews*, *16*(4), 361–391.

Whoscored.com. (2018). *Glossary*. https://www.whoscored.com/Glossary

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, *6*(1), 58.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata Journal*, *12*(2), 308.

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *The Journal of Mathematical Sociology*, *40*(1), 7–20.

Wright, C., Atkins, S., Polman, R., Jones, B., & Sargeson, L. (2011). Factors Associated with Goals and Goal Scoring Opportunities in Professional Soccer. *International Journal of Performance Analysis in Sport*, *11*(3), 438–449. https://doi.org/10.1080/24748668.2011.11868563

Yang, G., Leicht, A. S., Lago, C., & Gómez, M. Á. (2018). Key team physical and technical performance indicators indicative of team quality in the soccer Chinese super league. *Research in Sports Medicine*, *26*(2), 158–167. https://doi.org/10.1080/15438627.2018.1431539

References

Yue, Z., Broich, H., & Mester, J. (2014). Statistical Analysis for the Soccer Matches of the First Bundesliga. *International Journal of Sports Science & Coaching*, *9*(3), 553–560. https://doi.org/10.1260/1747-9541.9.3.553

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14. https://doi.org/10.1111/j.2041-210X.2009.00001.x