UNIVERSIDADE DE LISBOA
FACULDADE DE MOTRICIDADE HUMANA

Analysis of goal scoring in football matches according to performance indicators and the context of competition

José Maria Dionísio Calado Pratas

Orientadora: Prof.Doutora Anna Georgievvna Volossovitch
Co-orientadora: Prof.Doutora Ana Isabel Andrade Dinis Carita

Tese especialmente elaborada para a obtenção do grau de Doutor em ramo de Motricidade Humana na especialidade de Treino Desportivo

Júri:
PRESIDENTE:
Doutor Francisco José Bessone Ferreira Alves
Professor Catedrático e Presidente do Conselho Científico
Faculdade de Motricidade Humana da Universidade de Lisboa

VOGAIS:
Doutor Duarte Fernando da Rosa Patronilho de Araújo
Professor Associado com Agregação
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor Hugo Miguel Cardinho Alexandre Folgado
Professor Auxiliar
Escola de Ciências e Tecnologia da Universidade de Évora

Doutor Bruno Filipe Rama Travassos
Professor Auxiliar
Departamento de Ciências do Desporto da Universidade da Beira Interior

Doutora Anna Georgievna Volossovitch
Professora Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor Ricardo Filipe Lima Duarte
Professor Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

2017

UNIVERSIDADE DE LISBOA

FACULDADE DE MOTRICIDADE HUMANA

Analysis of goal scoring in football matches according to performance indicators and the context of competition

José Maria Dionísio Calado Pratas

Orientadora: Prof.Doutora Anna Georgievvna Volossovitch
Co-orientadora: Prof.Doutora Ana Isabel Andrade Dinis Carita

Tese especialmente elaborada para a obtenção do grau de Doutor em ramo de Motricidade Humana na especialidade de Treino Desportivo

Júri:
PRESIDENTE:
Doutor Francisco José Bessone Ferreira Alves
Professor Catedrático e Presidente do Conselho Científico
Faculdade de Motricidade Humana da Universidade de Lisboa

VOGAIS:
Doutor Duarte Fernando da Rosa Patronilho de Araújo
Professor Associado com Agregação
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor Hugo Miguel Cardinho Alexandre Folgado
Professor Auxiliar
Escola de Ciências e Tecnologia da Universidade de Évora

Doutor Bruno Filipe Rama Travassos
Professor Auxiliar
Departamento de Ciências do Desporto da Universidade da Beira Interior

Doutora Anna Georgievna Volossovitch
Professora Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

Doutor Ricardo Filipe Lima Duarte
Professor Auxiliar
Faculdade de Motricidade Humana da Universidade de Lisboa

2017

The work presented in this dissertation was supported by the Foundation for Science and Technology (Portugal) grant SFRH/BD/80719/2011 awarded to José Maria Pratas.

"People in both fields operate with beliefs and biases. To the extent you can eliminate both and replace them with data, you gain a clear advantage."

Michael Lewis*, *Moneyball: The Art of Winning an Unfair Game

Acknowledgments

I would like to thank a number of people, without whom my PhD course would not have been so enjoyable and memorable.

Firstly, and most importantly, I want to thank my thesis tutor, **Prof. Anna Volossovitch**, for her strategic insight, guidance, patience and confidence in me. I am extremely grateful to her for my PhD training and I learned so much from her over recentyears. She is a wonderful tutor who not only teaches and inspires, but also cares about her students.

Secondly, I wish to thank to my thesis co-tutor, **Prof. Isabel Carita** who gave me valuable feedback and input on my research and advised me whenever I had questions about statistics and programming.

I also wish to thank to **Prof. Orlando Fernandes**, who always took the time to answer my queries despite his busy schedule, and gave me valuable input on my research. He is also a good friend: always there for me when I need help.

Many thanks go to **Prof. Michael Lewis**, for his quick response whenever I needed help in preparing for presentations for scientific meetings and also for reviewing my papers.

Next, I wish to thank all those at the **Faculdade Motricidade Humana** - all **department members** and many other fine graduate students of the **Spertlab Department** - for their helpful comments and suggestions throughout my academic career.

This research was funded by a PhD grant from the **Portuguese Foundation for Science and Technology**, which enabled me to focus my energy and efforts on research for my thesis.

Finally, I wish to thank my wife **Inês** and all my **family** for their love and support, without which I could not have got so far. I would also like to mention my dear son **Diogo** who was born during this research project.

Resumo

Esta tese teve como objetivo analisar os padrões temporais de marcação de golos no futebol profissional de acordo com a perspectiva dinâmica e em função de diferentes variáveis explicativas relacionadas com a performance das equipas e contexto competitivo. O trabalho inicia-se com a revisão sistemática dos estudos que analisaram a marcação de golos em ligas profissionais de futebol masculino. Em função dos métodos utilizados na recolha e tratamento de dados os estudos analisados foram classificados em dois grupos – realizados segundo a perspetiva estática ou uma perspetiva dinâmica. As vantagens e limitações de ambas as perspetivas no sentido de proporcionar informações relevantes sobre a performance das equipas são discutidas no capítulo.

Nos capítulos que se seguem foram analisados os fatores que influenciam os tempos de marcação do primeiro e segundo golos do jogo, tal como do golo-chave (key-goal) que define o resultado do jogo (vitória ou empate). Procedimentos de análise de sobrevivência (função de incidência cumulativa, modelos de regressão de Cox com covariáveis fixas e variáveis no tempo e modelos competitivos) foram desenvolvidos e aplicados para modelar o risco de marcação do primeiro, segundo e golo-chave ao longo do jogo em função de um conjunto de indicadores de performance e variáveis contextuais. Os resultados confirmaram que os indicadores de performance, como o diferencial de golos (diferença entre os golos marcados e sofridos durante o campeonato até ao jogo observado), remates à baliza, sanções disciplinares e substituições, permitem predizer o tempo em que o primeiro golo é marcado pelas equipas da casa. Os resultados também demonstraram que o primeiro golo do jogo marcado pelas equipas de casa tem uma associação positiva e significativa com a vitória final. A análise do segundo golo do jogo revelou que este é marcado com a maior frequência na segunda parte do jogo quando a equipa joga em casa. O modelo de Cox demonstrou que o tempo do primeiro golo tem um efeito positivo e significativo no tempo do segundo golo marcado pela equipa da casa, particularmente se o primeiro golo foi marcado na segunda parte do jogo. Para modelar o tempo do golo-chave foram utilizados modelos de riscos competitivos que demonstraram que as equipas da casa têm uma probabilidade maior de marcar o golo-chave associado à vitória, enquanto as equipas com um diferencial acumulado positivo de golos marcados e sofridos tiveram menor

probabilidade de marcar o golo-chave associado à vitória. As equipas da casa e as equipas com um maior diferencial positivo de golos marcados e sofridos demonstraram uma maior probabilidade de marcar o golo-chave associado a um empate como resultado final. O estudo abre novas perspetivas para a análise da performance coletiva no futebol de alto rendimento. A abordagem metodológica proposta no estudo toma em consideração a dinâmica de desempenho coletivo durante o jogo e permite fazer previsões de rendimento coletivo que facilitam a tomada de decisão dos treinadores na direção da equipa.

Palavras-chave: golos, futebol, probabilidade, análise de sobrevivência, indicadores de performance, variáveis situacionais, primeiro golo, segundo golo, golo-chave

Abstract

The aim of this thesis is to analyse goal-scoring patterns in professional football in accordance with the dynamic perspective, using different explanatory variables associated with team performance and game context. In the first section there is a systematic review of studies which analyse goal scoring in men's professional football leagues. In accordance with the methods used for data collection and analysis, studies were classified into two groups: studies carried out according to the static perspective and those carried out according to the dynamic perspective. The advantages and limitations of both perspectives in providing relevant information on team performance are also discussed.

In the following chapters there is analysis of the factors that influence the time at which the first and second goals are scored in a match, as well as the idea of the key-goal which defines the outcome of the game (a win or a draw). Survival analysis procedures (cumulative incidence function, Cox regression models with fixed and time varying covariates and competing models) were devised and applied in match analysis to estimate the probability of the first goal, second goal and the key-goal being scored during a match as a function of a set of explanatory variables regarding team performance and match context. The findings of this study confirmed that performance indicators, such as positive accumulated goal difference (the difference between the goals scored and conceded during the whole competition up until the match considered), shots on goal, sanctions and substitutions enable the time at which the first goal is scored by home teams to be predicted. They also revealed the importance of scoring the first goal for home teams to win a match. In addition, they showed that the first goal in a match scored by home teams was positively and significantly associated with a winning outcome. Analysis of the second goal of the match demonstrated that this is scored most often by home teams in the second half of the game. The Cox model showed that the time at which the first goal is scored in a match had a positive effect on the second goal being scored by the home team, particularly if the first goal was scored in the second half of the match. The competitive risk regressions model showed that there was a higher probability of home teams scoring the key-goal of match, while there was a lower probability of teams with a positive accumulated goal difference scoring the key-goal that led to victory. Home teams, and teams with a positive accumulated goal

difference, showed a higher probability of scoring the key-goal that tied the match. This study brings new perspectives for performance analysis in elite football by taking into account the dynamics of team performance during matches. The approach proposed in this study enables the prediction of scoreline evolution based on performance indicators and thus might facilitate decision-making by coaches regarding instructions given to teams during matches.

Keywords: goal scoring, football, probability, survival analysis, performance indicators, situational variables, first goal, second goal, key-goal

CONTENTS

Tables

Figures were numbered following chapter number

Figures

Figures were numbered following chapter number

CHAPTER 1

INTRODUCTION

1. Introduction

1.1. Scope and relevance of the study

The objective in football is simple: in order to win a match, one team has to score one more goal than the opposing team. However, in sports like football, where goals are inherently random and scarce, identifying which factors might predict or be associated with success has become an almost unattainable aim in game analysis (Wright, Carling & Collins, 2014). Goal scoring may be the only performance indicator, or at least the most valid such indicator, which determines team success. For this reason, current trends in football match analysis reveal a growing interest in the probabilistic modelling of goal scoring (Collet, 2013, Tenga, Holme, Ronglan & Bahr, 2010; Wright, Atkins, Polman, Jones & Sargeson, 2011).

Over the last ten years, our knowledge about individual and team performance indicators associated with goal scoring has increased greatly due to technological developments (Carling, Reilly & Williams, 2009). Much previous research attempted to analyse the likelihood of scoring using standard statistical methods (for example, the chi-square test, parametric tests of significance, non-parametric tests for comparison, linear and logit/probit models), which frequently do not take into account the dynamic nature of football matches and the influence of certain events on the subsequent course of the match (Tenga, Holme, Ronglan & Bahr, 2010a; Wright et al., 2011). Most descriptive and comparative studies on performance analysis in football used end-of-match accumulated statistics to look for a linear relationship between the frequency of players' actions and the match outcome. The problem is that this analysis discards contextual data and does not acknowledge that some variables might be closely correlated with time. For these reasons, static probability models are not suitable for analysing events over time and do not provide relevant or useful information about the match process (Castilla, 2007).

By contrast, a dynamic approach includes the time dimension in the analysis of team performance (Volossovitch, 2013) and allows for the identification of specific patterns of players' interactions, which emerge over time during the match (Passos, Araújo & Davids, 2013). Research conducted in accordance with this approach assumes that performance varies during the match and uses advanced analysis techniques (Duarte, Araújo, Correia, Davids, Marques & Richardson, 2013; Fernando, Wei, Fookes, Sridharan & Lucey, 2015), which estimate game outcome by taking into

account the evolution of performance over time. To evaluate this evolution it is necessary to understand how previous events influence subsequent events (Bar-Eli, Tenenbaum & Geister, 2006; Dixon & Robinson, 1998).

For this purpose, some recent studies on performance analysis in football have used survival analysis (Del Corral, Barros & Prieto-Rodríguez, 2008; Nevo & Ritov, 2012), which has some advantages compared with standard statistical methods. Survival analysis is a set of statistical tools that considers the temporal dimension of data. Survival techniques explicitly include the time variable in the functional form of the model, which is particularly relevant for the analysis of goal-scoring patterns in football, because it is not possible to identify these patterns without considering the time of the goal, and also other temporal transformations associated with multiple factors which influence the probability of a goal being scored.

One reason why using such survival models is more pertinent than using ordinary regression models (such as ordinary least square regression or logistic regression, among others) is that these techniques add information about the time at which different match actions occur and goals are scored. Such methods allow a relationship to be established between not only different match actions and goals but also between the times, which elapse between these actions and goals. Since a match has a fixed duration, this time information is extremely important for the recognition of goal-scoring patterns. In other words, unlike other techniques, survival analysis not only focuses on the event of interest (a goal, for instance), but also analyses the time to this event as a function of explanatory variables. Another particularity of survival analysis is the capability for including in the analysis time-varying covariates (that is, explanatory variables whose values change over time, such as time of ball possession and number of shots on goal), which is not possible with ordinary regressions. For these reasons, these survival models are often referred to as dynamic models (Aalen, Borgan & Gjessing, 2008; Castilla, 2007).

By examining the factors behind the occurrence of any type of event over time during a match, survival analysis helps us to predict score evolution during the match and thus anticipate the outcome and also to be better prepared for different game scenarios.

1.2. Aims

The aim of this study is to analyse goal-scoring patterns in football in accordance with a dynamic perspective, using different explanatory variables associated with team performance and match context. The different types of goal considered in the study were determined according to current match scoreline (the first or second goal of the match) and the impact, which this goal had on final match outcome (the key-goal of the match).

This study comprises six chapters and there were the following four main objectives of analysis of goal scoring in matches played in the Portuguese Premier League from the 2009 to 2015 seasons:

1) to conduct a systematic review of relevant studies on goal scoring in elite male football leagues, and classify them according to the approaches used for the collection of data and data analysis; to identify the key performance indicators most frequently associated with goal scoring, characterising the main methodologies used in studies and identifying potential questions for future research;

2) to identify the performance indicators that influence the time at which the first goal is scored in high-level football matches; the difference between the total number of goals scored and conceded during the competition up until the match considered, total ball possession time, shots on goal, set plays, disciplinary sanctions and substitutions were selected as predictor variables for the time at which the first goal was scored;

3) to analyse the association between the type of second goal scored in a match, match venue, and final match outcome; to examine the influence of the time at which the first goal is scored and the current scoreline on the probability of a second goal being scored in a match;

4) to investigate key-goal scoring in high-level football matches; the explanatory covariates used were team scoring, accumulated goal difference and scoreline.

1.3. Organisation of the thesis

The present thesis comprises four chapters (a review and three correlational articles), one of which has already been published as a separate paper, while another paper has been submitted, and a further two papers are currently under review for for publication in peer-reviewed ISI journals with impact. Each chapter is presented as an individual paper in accordance with the format indicated by the relevant journal with regard to sections, reference style, figures, tables and figures.

Chapter 1 – The **General Introduction** characterises the scope and relevance of the study, and presents the aims and organisation of this thesis.

In Chapter 2 the available literature on goal scoring in elite male football leagues is reviewed. For this purpose, a systematic search of two electronic databases (SPORTDiscus with Full Text and ISI Web Knowledge All Databases) was conducted.

Scoring the first goal in football could be crucial for winning a match. The aim of **Chapter 3** is to identify the performance indicators (the difference between the total number of goals scored and conceded during the competition up until the match considered, total ball possession time, shots on goal, set plays, disciplinary sanctions and substitutions) that influence the time the first goal is scored in high-level football matches and for this purpose a Cox proportional model with fixed and time-varying covariates was used.

Chapter 4 examines the second goal scored in football matches in accordance with a range of different match contexts. Cox regression analysis was used to predict the time the second goal is scored as a function of the time of the first goal is scored in the match and the scoreline.

The time at which the goal which defines match outcome is scored is analysed in **Chapter 5**. The cumulative incidence function is used to estimate the probability of each type of key-goal being scored during the match (with match outcome defined as a win or a draw) and the influence of covariates (that is, match venue, accumulated goal difference and scoreline) on cause-specific hazards was explored using survival analysis in conjunction with a competing risk model.

Finally, in **Chapter 6** a **General Discussion** of the main findings and limitations of this research is presented along with some final considerations regarding the contribution of this work to football match analysis.

1.4. References

Aalen, O.O., Borgan, Ø., & Gjessing, H.K. (2008). *Survival and Event History analysis. A process Point of View*. New York: Springer.

Bar-Eli, M., Tenenbaum, G., & Geister, S. (2006). Consequences of Player's Dismissal in Professional Soccer: A Crisis-Related Analysis of Group-Size Effects. *Journal of Sports Sciences*, 24(10), 1083-1094.

Carling, C., Reilly, T., & Williams. M. (2009). *Performance Assessment for Field Sports*. Routledge: London.

Castilla, E.J. (2007). *Dynamic Analysis in the Social Sciences*, 1st Edition. USA: Elsevier, Inc.

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.

Del Corral, J., Barros, C. P., & Prieto-Rodríguez, J. (2008). The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League. *Journal of Sports Economics, 9*(2), 160-172.

Dixon, M., & Robinson, M. (1998). A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician), 47*:523–538.

Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., & Richardson, M. (2013). Competing together: Assessing the dynamics of *team-team* and *player-team* synchrony in professional association football. *Human Movement Science*, *32*(4), 555–566.

Fernando, T., Wei, X., Fookes, C., Sridharan, S., & Lucey, P. (2015). *Discovering methods of scoring in soccer using tracking data*. Paper presented at the Large-Scale Sports Analytics, Sidney.

Lago-Peñas, C., & Gomez-Lopez, M. (2014). How important is it to score a goal? The influence of the scoreline on match performance in elite soccer. *Perceptual and Motor Skills, 119*(3), 774-784.

Nevo, R., & Ritov, Y. (2012). Around the goal: Examining the effect of the first goal on the second goal in soccer using survival analysis methods. *Journal of Quantitative Analysis in Sports, 9*(2), 65-177.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effect of playing tactics on

goal scoring in Norwegian professional soccer. *Journal of Sports Sciences*, *28*(3), 237-244.

Volossovitch, A. (2013). Handball. In T. McGarry, P.O'Donoghue, & J.Sampaio (Eds.) *Handbook of sports performance analysis*. (380-392). London: Routledge.

Volossovitch, A., & Ferreira, A.P. (2013). Da descrição estática à predição dinâmica. A evolução das perspectivas de análise da performance nos jogos desportivos coletivos. In A. Volossovitch & A. P. Ferreira (Ed.). *Fundamentos e aplicações em análise do jogo.* Cruz Quebrada, Portugal: Edições FMH.

Wright, C., Carling, C., & Collinsz, D. (2014). The wider context of performance analysis and it application in the football coaching process. *International of Performance Analysis in Sport, 14*,709-733.

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.

CHAPTER 2

GOAL SCORING IN ELITE MALE FOOTBALL: A SYSTEMATIC REVIEW[1]

[1]Submitted as:
Pratas, J. M., Volossovitch, A., Carita, A.I. (In Press). Goal scoring in elite male football: a systematic review. *Journal of Human Sport and Exercise.*

2. Goal scoring in elite male football: a systematic review

2.1. Abstract

The aim of this paper is to review the available literature on goal scoring in elite male football leagues. A systematic search of two electronic databases (SPORTDiscus with Full Text and ISI Web Knowledge All Databases) was conducted and of the 610 studies initially identified, 19 were fully analysed. Studies that fitted all the inclusion criteria were organised according to the research approach adopted (static or dynamic).

The majority of these studies were conducted in accordance with the static approach (n=15), where the data were collected without considering dynamic of performance during matches and were analysed using standard statistical methods for data analysis. They focused predominantly on a description of key performance indicators (technical and tactical). Meanwhile, in a few studies the dynamic approach (n=4) was adopted, where performance variables were recorded taking into account the chronological and sequential order in which they occurred. Different advanced analysis techniques for assessing performance evolution over time during the match were used in this second group of studies. The strengths and limitations of both approaches in terms of providing the meaningful information for coaches are discussed in the present study.

Keywords: game analysis; team sport; tactics; performance indicators; soccer, goal scoring

2.2. Introduction

Match performance in football can be characterised by the interaction of different technical, tactical, mental (Carling, Reilly & Williams, 2009) and physiological (Di Salvo, Pigozzi, González-Haro, Laughlin & De Witt, 2013) performance indicators recorded in different contextual conditions during a football match (Lago-Peñas & Gomez-Lopez, 2014). The main objective of a football game is for a team to score more goals than their opponents and this is why the identification of goal-scoring patterns and successful attacking strategies is one of the most pertinent issues in football match analysis.

Over the last ten years, our knowledge about individual and team performance indictors associated with goal scoring has been greatly increased thanks to technological developments (Carling, Reilly & Williams, 2009), which have provided coaches and players with information about appropriate strategic and tactical decisions according to game context (Collet, 2013; Kempe, Vogelbein, Memmert & Nopp, 2014). The recognition of key performance indicators related to goal scoring in football in general derives from two main approaches: the static and the dynamic (Pfeiffer & Perl, 2006; Sampaio, Ibáñez & Lorenzo, 2013; Volossovitch, 2013). In accordance with the static perspective, performance and key game events are recorded using notation systems that address final match statistics, with little or no regard to the context of the match at any given moment, while with the dynamic perspective, actions and key events are recorded in connection with the state of the match at each instant in a chronological and sequential order (Prieto, Gómez & Sampaio, 2015).

To the best of our knowledge, very few systematic reviews of football match analysis have been conducted (Mackenzie & Cushion, 2013; Sarmento, Marcelino, Anguera, Campaniço, Matos & Leitão, 2014) and there is a lack of research focused particularly on goal scoring in adult male domestic football leagues. Therefore, the aims of this study were: 1) to conduct a systematic review of relevant studies on goal scoring in elite male football leagues, classifying them according to the approaches used for collecting and analysing data; 2) to identify the key performance indicators most frequently associated with goal scoring; 3) to characterise the main methodologies used in studies carried out; and 4) to identify potential questions for future research .

2.3. Methods

A systematic search of two electronic databases (SPORTDiscus with Full Text and ISI Web Knowledge All Databases) was conducted from February to September 2016. Reference lists of all identified papers were analysed so that no relevant study would be omitted. Combinations of the following keywords were used: “soccer” and “football”, each associated with the terms: “match analysis”, “game analysis”, “performance indicators”, “goals scored”, “tactical behaviour”, “tactical analysis”, “video analysis”, “playing tactics”, and “collective variables”. With this study no application was made for ethics approval because a systematic review is based on published sources.

Inclusion criteria for studies were: (1) conducted by professional adult male football players in domestic leagues; and (2) written in English or Spanish. Studies were excluded if they were: (1) unpublished articles, dissertations, book chapters, or conference abstracts without a corresponding full text paper; (2) conducted by amateur male footballers, minors and females; and (3) data from World Cup and European tournaments. The eligibility of each paper was assessed independently by two researchers (JP and AV); any disagreement regarding the inclusion of articles was resolved in consultation with a third reviewer (AC).

2.4. Results

The bibliographical search initially allowed for 610 titles to be identified using reference manager software (Endnote X7, Thomson Reuters, Philadelphia, PA, USA), while duplicates (n=12) were eliminated. The remaining 598 articles were screened for relevance based on their title and abstract, resulting in another 573 studies being eliminated from the database. The full text of the remaining 25 articles was read and another six papers were rejected due to a lack of relevance for the purpose of this study. At the end of the screening procedure, fifteen articles adopting a static approach and only four articles adopting a dynamic approach remained as being eligible for inclusion in the systematic review. The reviewed published studies on goal scoring adopting static and dynamic approaches were collated, then the sample and statistics and/or methods of data collection and analysis used in each study were identified for classification and analysis purposes (see Table 1).

Table 1. Characteristics of included studies on goal scoring in football according to the approach used

	Author	Sample (Competition and number of matches)	Variables	Methods
Static approach	1. Redwood-Brown, A. (2008)	England - 120	Pass accuracy	Non-parametric tests (Spearman's correlations; Friedman and Wilcoxon
	2. Armatas et al. (2009)	Greece - 240	Temporal analysis; Scoring efficiency	Chi square
	3. Johnson & Murphy (2010)	Australia - 84	Number of passes	Descriptive statistics
	4. Lago-Ballesteros & Lago-Peñas (2010)	Spain - 380	Scoring efficiency	Anova; Bonferroni
	5. Tenga et al. (2010a)	Norway - 163	Types of passes; Zones in which possessions started; Playing style	Chi square; Logistic regression
	6. Tenga et al. (2010b)	Norway - 163	Types of passes; Zones in which possessions started; Playing style	Chi square; Logistic regression
	7. Lago-Peñas & Lago-Ballesteros (2011)	Spain - 380	Scoring efficiency	Univariate (t-test and Mann-Whitney U) and multivariate (discriminant analysis)
	8. Janković et al. (2011)	Serbia - 228	Pass accuracy; Areas from which goals were scored; Scoring efficiency	Kruskal Wallis; Mann-Whitney-U; Chi-square
	9. Tenga & Sigmundstad (2011)	Norway - 1324	Number of passes; Duration of possession; Zones in which possessions started; Playing style	Kruskal Wallis H; Bonferroni adjusted Mann-Whitney U
	10. Wright et al. (2011)	England	Number of passes; Zones in which possessions started; Areas from which goals were scored; Game situation	Logistic regressions model
	11. Gomez et al., (2012)	Spain - 1900	Types of passes; Zones in which possessions started; Areas from which goals were scored	Factor analysis
	12. Alberti et al.(2013)	England, France, Italy, Spain - 4560	Temporal analysis	Chi square
	13. Collet (2013)	England, Italy, France, Spain, Germany - 5478	Pass accuracy; Playing style	Regression analysis (ordered logit)
	14. Armatas & Pollard (2014)	Greece - 2160	Areas from which goals were scored	Factor analysis with principal components
	15. Kempe et al. (2014)	Germany - 676	Playing style	Anova
Dynamic approach	16. Nevo & Ritov (2012)	England - 760	First and next goals	Cox models
	17. Fernando et al. (2015)	380	Space-time coordination	Machine learning algorithms
	18. Cintia et al. (2015)	Germany, Spain England, Italy - 1446	Space-time coordination	Network analysis
	19. Pratas et al. (2016)	Portugal - 240	First goal	Cox regression

2.4.1. Goal scoring analysis adopting a Static Approach

2.4.1.1. Temporal analysis of goals scored

Several studies attempted to identify goal-scoring patterns in football through the temporal analysis of scoring occurrences. It was demonstrated that team performance during certain periods of a match has a greater impact on match outcome (Armatas, Yiannakos, Papadopoulou & Skoufas, 2009; Alberti, Iaia, Arcelli, Cavaggioni & Rampinini, 2013). The review of the literature showed that the frequency of goals scored during a match was time-dependent (Armatas et al., 2009). A greater number of goals was scored in the second half of matches, 15-min interval analysis revealed a significant increase in goals scored as the game progresses (Alberti et al., 2013) and in the last 15-minute periods of each half (Armatas et al., 2009, Alberti et al., 2013). Alberti et al. (2013), suggesting that goal-scoring patterns are not related to particular season or country-related style of play with regard to the English Premier League, the French Football League 1, the Italian Series A and the Spanish Football League.

2.4.1.2. Passes

Pass accuracy

There is not doubt that pass accuracy is an important performance indicator in football. Frequent and accurate passing were closely linked to goals scored (Redwood-Brown, 2008; Collet, 2013). For instance, it was shown that the pass accuracy was significantly higher five minutes before a goal was scored. Pass accuracy not only enables possession to be retained, but may also lead to scoring opportunities, while restricting the possession and scoring opportunities of the opposing team (Redwood-Brown, 2008; Janković, Leontijević, JelušIć, Pašić & Mićović, 2011).

Type of passes

Successful teams were able to execute a greater variety of passes, which were used to create goal scoring opportunities, rendering players' actions less predictable and thus allowing them to shoot more frequently (Oberstone, 2009).

In the Norwegian professional league, short passes (< 30 meters) were more effective in producing goals than long passes (Tenga, Holme, Ronglan & Bahr, 2010b). Penetrative passes into the 18-yard box and the 6-yard box in the Norwegian elite league (Tenga, Holme, Ronglan & Bahr, 2010a,b) and the Spanish professional football

league (Gomez, Gomez-Lopez, Lago & Sampaio, 2012) were also significantly associated with the shots on goal and consequently the number of goals scored (Tenga et al., 2010a,b; Gomez, Gomez-Lopez, Lago & Sampaio, 2012). This kind of pass was found to be more effective than a non-penetrative pass when playing against both a balanced and an unbalanced defence (Tenga et al., 2010a).

Number of passes and time of possession prior to goals being scored

In the elite English FA Premier League (Wright, Atkins, Polman, Jones & Sargeson, 2011) and in the Norwegian elite league (Tenga & Sigmundstad, 2011) a higher number of goals was scored from shorter passing sequences involving one to four passes. Moreover, Wright et al. (2011) showed that normalised attempts to goal ratios for shorter passing sequences 0-4 were more favourable (9.52 goals per 100 possessions) than longer passing sequences 5-8 (7.28 goals per 100 possessions).

At the same time, in the Australian A-League (Johnson & Murphy, 2010) and the Norwegian league (Tenga et al., 2010a,b), five or more passes led to the highest percentage of goals being scored, indicating a greater degree of 'possession play'.

The duration of possession prior to a goal being scored is highlighted in this review. (Tenga & Sigmundstad, 2011) showed that the top three teams scored 19.2±5.9 goals on average from 12-second possessions, while mid-table teams and the bottom three teams scored 15.0±3.8.1 and 10.8±3.7 goals, respectively. Moreover, the mean frequency values for goals scored by the top three teams were higher compared with mid-table teams and the bottom three teams in the three categories of possession duration (0-5, 6-11 and 12 seconds or more). These findings suggest that Norwegian teams are more likely to score goals following a possession play.

2.4.1.3. Regaining possession resulting in goals being scored

An important body of research is related to the zones in which possessions started. The proportion of goals scored after regaining possession in the attacking third of the field of play was very low, when compared with the midfield and defensive thirds.

In the English and Spanish leagues the highest number of goals was scored after possession had been regained in a team's own half of the field (Wright et al., 2011; Gomez et al., 2012), while in the Norwegian league, more goals are scored following regained possession in the midfield third (Tenga et al., 2010a,b; Tenga & Sigmundstad,

2011). Tenga & Sigmundstad (2011); Tenga et al., (2010a,b); Wright et al., (2011) and Gomez et al., (2012) considered zones 3, 4, 8 and 19, respectively.

2.4.1.4. Areas from where goals were scored

An important feature of goal-scoring is the area from which goals are scored. The studies reviewed show that in the English (Wright et al., 2011), Greek (Armatas & Pollard, 2014), Serbian (Janković et al., 2011) and Spanish (Gomez et al., 2012) leagues the majority of goals were scored from inside the 18-yard box. Furthermore, Gomez et al., (2012) demonstrated that winning Spanish teams worked the ball into goal-scoring areas closer to their opponents' goal; they had more shots and goals inside the 6-yard box than drawing and losing teams. The findings also suggest that teams recording worse results failed their attacks closer to the goal while shooting at greater distances from goal (Janković et al., 2011).

2.4.1.5. Shots on goal /goals scores (scoring efficiency)

With regard to game location and team quality, top teams and home teams presented significantly higher means for total shots and shots on goal (Lago-Ballesteros & Lago-Peñas, 2010; Lago-Peñas & Lago-Ballesteros, 2011). Successful teams showed a higher ratio between goals and shots than less successful teams. Top-of-the-table teams and mid-table teams were more effective goal-scorers, requiring a lower number of shots to score a goal than bottom-of-the-table teams (Armatas et al., 2009; Janković et al., 2011).

2.4.1.6. Playing style

The question of which attacking style is more effective in goal scoring is considered relevant in game analysis and has been addressed in several studies (Collet, 2013; Kempe et al., 2014). Research focused mainly on how goals were scored by successful and unsuccessful teams (Tenga et al., 2010a,b; Tenga and Sigmundstad, 2011).

An analysis of the playing style of Norwegian professional league teams suggested that counterattacks were more efficient than elaborate attacks in producing goals (Tenga et al., 2010a,b; Tenga and Sigmundstad , 2011). However, Kempe et al., (2014) demonstrated that the most successful teams of the Bundesliga preferred

possession play instead of direct play in 612 games played in the 2009/10 and 2010/11 seasons.

Collet (2013) found that ball possession was a direct predictor of goal scoring in domestic competitions in Europe, but only when such outcomes were not aggregated to the team level. When team quality was taken into account at the individual match level, the impact of greater possession was no longer positive in league games. Collet (2013) also showed that accurate passing and shooting were more closely related to team success than ball retention.

2.4.1.7. Game situation

Some research suggests that in English set plays contribute to approximately 35% (Wright et al., 2011) of all goals, while the highest percentage of goals were scored from open play in elite (65%; Wright et al., 2011). According to the literature, despite set plays being less frequent, teams which employed them improved considerably in terms of goal-scoring efficiency. Free kicks, followed by corners, are considered to be the most productive game situations and successful teams are far more efficient than their opponents at scoring from set plays (Wright et al., 2011).

2.4.2. Goal scoring analysis adopting a Dynamic Approach

2.4.2.1. First and next goal occurrence

Nevo & Ritov (2012) examined the interaction between two random goal scoring occurrences during a football match, using a survival analysis method (the Cox proportional hazard model). According to the authors, first goal occurrence could either expedite or impede the next goal being scored depending on the time at which the first goal was scored. The findings of their study demonstrated that there was no immediate effect when a goal was scored in the 52nd minute. If the goal was scored before the 52nd minute, the effect was negative, and if it was scored after the 52nd minute, the effect was positive. Additionally, when the first goal was scored, another goal was less likely to be scored, as compared to the situation when no goal had been scored.

This analysis served as the starting point for the study carried out by Pratas, Volossovitch & Carita (2016), who analysed the effect of performance indicators that influence the time the first goal is scored in high-level Portuguese football matches played in the 2009/10 season. Using a Cox model with time dependent covariates, the

authors identified performance indicators such as accumulated goal difference, shots on goal, sanctions and substitutions, enabling the time the first goal is scored by home teams to be predicted. A greater goal difference (goals scored minus goals conceded by team up until the time of the game observed) and a larger number of shots on goal had a significant positive influence on the time the first goal was scored in matches by home teams; at the same time disciplinary sanctions and substitutions had a significant negative effect on the time of the first goal (Pratas et al., 2016).

2.4.2.2. Playing patterns

The technological advances of recent decades have allowed for the collection of positional data about players and the ball during the match, which are now widely used in tactical analysis in football (Bialkowski, Lucey, Carr, Yue, Matthews, 2014; Memmert, Koen, Lemmink & Sampaio, 2016).

Fernando, Wei, Fookes, Sridharan & Lucey (2015) used an entire season's player-tracking data and ball-event data from a top-tier professional league consisting of 380 matches from Prozone. The authors presented a method which can be used to compare the scoring methods of soccer teams using fine-grained player-tracking and ball-event data. By using machine learning algorithms, based on game-position data, differentiated attacking plays across teams using cluster analysis of game sequences can be identified (scoring methods by clustering offensive and defensive player trajectories according to the game context). These groups were then further divided into those with good, average and poor scoring chances according to an expected goal-value model.

This method allowed for the identification and quantification of goal-scoring methods (corners, free-kicks and open-play) used by teams, and also the comparison of the "goal scoring styles" of teams.

Another emerging technique used for the analysis of team tactics is the network method, which describes team passing behaviour. Cintia, Pappalardo, Pedreschi, Giannotti & Malvaldi (2015) used the network approach to analyse team passing behaviour in the four major European football leagues (those of Germany, England, Spain, Italy in the 2013/2014 season). It was found that the passing activity of a team is related to its success during the competition, as teams with high level of passing activity tend to score more goals, have more goal-scoring opportunities, and win more points. In particular, the strongest European teams (the winners of national league championships

or teams with good performance during European cups, like Barcelona, Real Madrid, Manchester City and Bayern Munich) present a high average number of passes along with many goals scored, attempts made and points won. However, some teams do not follow these clear trends: they either present a low level of passing activity while achieving a good resultsas far as goal-scoring/attempts/point-winning is concerned (teams like Levante and Atlético Madrid) or they show a high level of passing activity but the results in terms of goal-scoring/attempts/point-winning are poor (for example, Swansea and Borussia Mönchengladbach).

Moreover, network analysis, which includes temporal and positional information about players and passes executed allowed for the prediction of game outcomes and the final ranking of the top teams using a K-Nearest Neighbor classifier.

2.5. Discussion

2.5.1. Goal scoring analysis adopting the static approach

The static approach represents the simplest form of complexity, mainly because it assumes that the structure of studied phenomenon does not change over time (Sampaio, Ibáñez & Lorenzo, 2013). In accordance with this approach, the relevant aspects of sports performance are captured at a given time (as in a photograph), extrapolated and applied to other competitive situations (Volossovitch & Ferreira, 2013).

Studies adopting this approach frequently associate certain team performance patterns with team success or failure. Performance patterns are configured by mean values of variables calculated from the data accumulated by match or championship. In most cases the data are recorded and analysed with little or no reference to match context (Volossovitch & Ferreira, 2013). The static approach to match analysis relates a range of performance indicators to performance outcomes, roughly identifying the components of teams' success without disclosing the process of how this performance has been achieved (Volossovitch, 2013). The main limitation of this approach is the inevitable loss of information, because the data collected account neither for changes in players' performance over time nor the complex interaction between team mates and opponents (Mackenzie & Cushion, 2013; Volossovitch & Ferreira, 2013).

Despite some limitations, the static approach has been that which is most widely used in football match analysis and studies published have identified different performance indicators related to goal scoring. The reasons for this are that notational

analysis systems are easy to design, cost-effective in terms of both time and resources, and provide coaches with valuable and objective quantitative information about the mid- and long-term performance of their teams. This information may be of use for the assessment of different teams' strategies and the planning of the training process.

The review of studies conducted in accordance with the static approach showed that more goals are scored as the match progresses, especially during the final few minutes of each half and in the second half.

The findings of several studies support the contention that the majority of goals are scored late in each half, probably on account of fatigue, both physical (Bangsboo, 1994; Krustrup, Mohr, Steensberg, Bencke, Kjaer & Bangsbo, 2006) and mental (Smith, Zeuwts, Lenoir, Hens, De Jong & Coutts, 2016), which accumulates as a match proceeds, leading to a greater number of technical failures (Russell, Benton & Kingsley, 2011), or due to the adoption of riskier attacking strategies in an attempt to change the current score (Ferguson, 2013; Njororai, 2014). It seems that the final fifteen minutes of the second half frequently represents a critical period in the match. Coaches should take this trend into account by reinforcing their team's defence, especially if the team does not enjoy a score advantage.

Several authors using pass performance indicators to explain goal scoring patterns reported that the percentage of accurate passes (Redwood-Brown, 2008; Collet, 2013), short passes (Tenga et al., 2010b), and penetrative passes (Tenga et al., 2010a,b; Gomez et al., 2012), as well as possessions with a duration of over 12 seconds (Tenga & Sigmundstad, 2011) were associated with scoring situations. Regarding the number of passes, the findings of several studies showed that shorter passing sequences involving one to four passes (Wright et al., 2011; Tenga & Sigmundstad, 2011) and long passing sequences involving five or more (Johnson & Murphy, 2010; Tenga et al., 2010a,b), led to the highest number of goals being scored. These inconsistent findings raise some questions about the use of possession variables for identifying the most effective offensive strategies.

In order to score a goal a team needs to have possession of the ball, but in some cases percentage of possession may not be a valid indicator of how good a team played (Lago-Peñas & Dellal, 2010). What really matters is scoring goals; possession can never be as important as scoring goals, even though it increases the likelihood of scoring. Possession alone is not the key for success (winning a match). It seems that there is

decisive possession, which leads to a goal scoring opportunity, rather than the mere prolonged possession (Lago-Peñas Lago-Ballesteros, Alexandre & Gomez, 2010). In order to differentiate more effective match strategies from those which are less so, it is necessary to contextualise the information regarding ball possession during the match. This means taking into consideration the quality of the opposition (weaker or stronger, or of the same level), scoreline evolution and playing time (how much time remains until the end of the match). The style of team attacking play may also influence time of possession. Direct play and possession play are the two most common attacking styles of playing football (Kempe et al., 2014). A direct attack (also known as a counter-attack or long-ball game) is characterized by the launching of a high-speed attack following ball recovery before the opposing team can regain its composure. Teams often use this style when they consider themselves as being inferior to their opponents, especially when playing away.

Possession football, also known as 'tiki-taka', is advocated by many teams and personalities (from Johan Cruyff to Louis Van Gaal, Arséne Wenger and Pep Guardiola) and is aimed at retaining possession of the ball longer than the opponents by employing many low risk passes and drawing the opposition out of shape before exploiting any gaps they leave, thus using a gradual pass and move strategy. Teams need technically and tactically well prepared players in order to use this attacking style.

Using data from five European leagues, Collet (2013) found that in the domestic league, possession time and passing were predictors of goal scoring as well as total points scored. When Champions League teams were removed from the analysis, the effect of possession time on goal scoring did not reach a standard level of statistical significance. For non-elite teams playing in domestic leagues, possession time was not related to the likelihood of scoring. In terms of evaluation of effectiveness, ball efficiency was more closely related to success than ball retention. Accurate shooting and accurate passing were strongly linked to goals scored and wins and therefore these indicators are essential for match control.

Tenga et al. (2010) analysed 163 of the 182 (90%) matches played in the Norwegian Men's Professional Football League in the 2004 season. Their findings showed that counterattacks were more effective than elaborate attacks for scoring goals, when teams played against an unbalanced defence. It was also found that the assessment of opponent interaction is critical for evaluating the effectiveness of offensive playing

tactics as regards the probability of scoring goals, and this assessment improves the validity of team match performance analysis in football.

Recently, Kempe et al. (2014) analysed a total of 676 official matches played in: the Bundesliga in 2009/2010; the Bundesliga in 2010/2011; and the FIFA World Cup in 2010. The authors suggested some new informative tactical measures: the Index of Offensive Behaviour, based on a number of passing parameters (passes per action, passing direction, and target player passes) and parameters of passing success (passing success rate and passing success rate in a forward direction), and the Index of Game Control, based on parameters of ball possession, the gaining possession and the quality of possession (mean passes per attack, game speed, mean time of attack, the gaining possession, distance covered per attack, relative ball possession rate). Teams which opt for employing "possession play" present a positive Index of Offensive Behaviour. These teams were more successful than teams which employed "direct play" which presented a negative Index of Offensive Behaviour. One of the most interesting findings of this study is that the teams which recorded the highest Index of Game Control were the most successful, regardless of their style of play. These findings suggest that variables which characterise ball possession and contribute to greater match control might have a great impact on match outcome.

One of the most robust findings in the research literature is the high level of offensive effectiveness of successful teams, which always demonstrate a higher percentage of shot efficiency (Collet, 2013). This allows one to conclude that an offensive playing style is not the determining factor for success, since top teams are usually able to play employing either style. The effectiveness of attacking actions and good teamwork has an overwhelming influence on a team's scoring ability.

With regard to the field zones in which possession of the ball was regained, the findings suggest a significant positive association with the number of goals scored after regaining possession in the midfield third (Tenga et al., 2010a,b; Tenga & Sigmundstad, 2011) and defensive third (Wright et al., 2011; Gomez et al., 2012). It seems that teams playing in different domestic leagues put defensive pressure on their opponents in order to promote ball recovery in specific zones of the field. On the one hand, the strategy for regaining possession of the ball in mid-defensive zones enables a team's defensive organisation to be preserved, but on the other hand, starting moves from longer distances from the opponents' goal requires skilled players who are able to get the ball

near the target goal.

One of the greatest challenges in football is to anticipate pitch occupation by opponents, and frequently this information is more relevant for understanding the opponents' tactical intentions than the frequencies of different players' actions. One of the methods used to calculate a player's and teams' dominant region is a Voronoi diagram. A region is considered as dominant when a player can reach it before any other player does (Lopes, Fonseca, Lese & Baca, 2015). The objective measurement of a player's region using Voronoi diagrams may be helpful in the analysis of player distribution on the pitch in specific game situations. This evaluation allows for the anticipation and reading of the tactical behaviour of opponents and the timely application of measures to counter this behaviour.

A common trend, which was found in different domestic leagues, is that majority of goals (about 81-83%) were scored from the 18-yard box (Janković et al., 2011, Wright et al., 2011). This is not surprising because the penalty area is adjacent to the goal and the likelihood of scoring from a position in this area is much higher when compared with positions further from the target goal.

Besides the highest percentage (≥ 65%) of goals being scored from open play, the importance of the preparation and planning of players' action in set plays (approximately 30%) was highlighted in several studies (Wright et al., 2011). Previous studies suggested that set plays account for a large percentage (27% to 32%) of scoring opportunities per match (Armatas & Yiannakos, 2010; Taylor, James & Mellalieu, 2005; Whright et al., 2011). A substantial body of research has suggested that set plays account for approximately 30% to 35% of all goals, irrespective of the type of competition involved (Armatas, Yiannakos & Hatzimanouil, 2007; Yiannakos & Armatas, 2006; Bangsbo & Peitersen, 2000; Wright et al., 2011). According to Wright et al., (2011), in the English Premier League the free-kick conversion rate was 1:12 (the conversion rate for corners was 1:11 and for throw-ins 1:12.6).

These findings highlight the importance of practicing set plays. Despite the relatively low occurrence of set plays in comparison with open-play opportunities, the potential productivity of set plays is quite high. Even in elite football, set plays still play a key role and frequently have a decisive impact on match outcome. Thus, set plays can be regarded as a powerful strategy for any team, regardless of a team's position in the league.

In sum, the analysis of key indicators associated with goal scoring in football demonstrated the evident importance of: number of shots on goal, passing accuracy and recovering possession in zones nearest to the opponents' goal. The relevance of this information in studies using the static approach raises some pertinent questions. Can this summarised data, collected out of match context and without considering the specific characteristics of teams' tactical behaviour, provide meaningful information for future competitive contexts? It is assumed that changes in scoreline and match outcome are dependent on the dynamic interactions between the actions of players of both teams. Thus, in order to provide a better understanding of the factors that influence the capability of teams for scoring more complex approaches are required, which include the information about offensive and defensive interactions and take into account the relationship between different match events in time.

2.5.2. Goal scoring analysis adopting the dynamic approach

Observational data collected and analysed in accordance with the dynamic approach (see Table 1) provide information which takes into account time evolution during the match (Volossovitch & Ferreira, 2013). These data aim to describe not only 'what events' occur, and 'when' and 'where', but also 'how' match outcome is achieved in the match process. The dynamic approach allows for the identification of specific patterns of players' interactions which emerge during the match (Passos, Araújo & Davids, 2013). The dynamic approach includes the time dimension in the analysis of team performance during the match: players' actions are recorded and related to the match process at each instant, in a chronological and sequential order (Volossovitch & Ferreira, 2013).

Studies on goal scoring conducted in accordance with to the dynamic approach can be classified into two groups: 1) studies based on the dynamic analysis of space-time coordination between players and teams; and 2) the analysis of scoring dynamics.

2.5.2.1. Dynamic analysis of space-time coordination between players and teams

Research based on position data is aimed at describing the dynamic patterns of interactive player behaviour during the match and forecasting teams' performance. Recently, two main tools have been used in match analysis to examine dynamic interactions among players in goal scoring situations: network analysis (Cintia,

Pappalardo, Pedreschi, Giannotti & Malvaldi, 2015); and machine learning algorithms (Fernando et al., 2015).

The results of network analysis showed that teams with a higher level of passing activity tend to score more goals and the number of passes can be used as a predictor of match outcome (Cintia et al., 2015). Network analysis is growing in importance in sport performance research. Network methods are used to assess group relationships in a football team through the analysis of passing behaviour in combination with spatial information. These methods provide relevant information about team tactical behaviour, and have been enhanced by the development of a wide range of mathematical tools for analysing networks. For example, it is a straightforward process to work out the most important nodes in the network using a measure known as centrality. In football, goalkeepers and forwards present the lowest level of centrality, while defenders and midfielders have the highest level. However, there are numerous different ways of measuring centrality and determining clusters, and it is not always clear why one method should be preferred over another. Therefore, it is necessary to systematically evaluate and compare these different methods to determine their utility and value in match analysis.

Tactical decision-making in elite football has also been investigated using machine learning algorithms based on game position data (Fernando et al. 2015; Knauf, 2016). Machine learning algorithms allow for the identification and comparison of teams' "goal scoring styles" and the description of team formations in different game situations (Fernando et al., 2015; Grunz, Memmert & Perl, 2012; Kempe, Grunz & Memmert, 2015).

Over the last ten years, the use of positional data has increased in match analysis (Duarte, Araujo, Correia & Davids, 2012; Duarte, Araujo, Freire, Folgado, Fernandes & Davids, 2012; Folgado, Koen, Lemmink, Frencken & Sampaio, 2014). The application of collective variables, such as geometrical centre, surface area, stretch index, team length and team width, among others, has enabled us to capture the spatial and temporal properties of players' displacement and accurately describe the interactive behaviour of a set of players in scoring situations (Frencken, Lemmink, Delleman & Visscher, 2011; Frencken & Lemmink, 2008). Dynamic tools, such as Voronoi diagrams, used to measure individual and team dominant regions and team spatial configurations, and Approximate entropy, used to identify movement patterns and inter-player coordination,

provide new opportunities for describing teams' spatial behavior and characterising players' spatial profiles for different game contexts, particularly in goal scoring situations. Despite considerable technological advances, the use of these dynamic tools in the analysis of real play remains scarce.

The majority of studies on coordination dynamics in football have analysed players' behaviour in small-sided games (Sampaio, Lago, Goncalves, Macas & Leite, 2014; Goncalves, Marcelino, Torres-Ronda, Torrents & Sampaio, 2016), and only a few studies have examined the 11x11 game (Fradua, Zubillaga, Caro, Fernandez-Garcia, Ruiz-Ruiz & Tenga, 2013; Goncalves, Figueira, Macas & Sampaio, 2014).

From a tactical perspective there are some pertinent questions related to goal-scoring analysis which should be addressed in future research. Firstly, it would be interesting to clarify what team formation leads most frequently to success (for example, 1-4-4-2, 1-3-5-2, and so on), and secondly, how teams use such formations (for example, retreating deep or pressing high) to create scoring situations.

2.5.2.2. Analysis of scoring dynamics

The second group of studies, carried out in accordance with the dynamic approach, includes the prediction of goal scoring based on score evolution analysis during the match (Nevo & Ritov, 2012) and the analysis of factors which influence the probability of scoring (Pratas et al., 2016).

Nevo & Ritov (2012) examined the effect of the first goal on the time the second goal is scored using survival analysis methods and concluded that first goal occurrence could either expedite or impede the next goal being scored, depending on the time at which it was scored. "Moreover, once a goal is scored, another goal becomes more and more likely whether the goal was scored or conceded" (Nevo & Ritov, 2012). Also, Dixon & Robinson (1998) argued that expectation of a goal is dependent on current score, and when an early away goal is scored, expectation of further goals is increased (more than the original expectation of goals before the match) with a bias towards the home team having their goal expectation increased. A recent study conducted by Pratas et al., (2016), which used proportional hazards regression models with time-dependent covariates, allowed for the identification of performance indicators (that is goal difference, shots on goal, disciplinary sanctions and substitutions) that influence the time at which the first goal is scored in high-level football matches.

The probability of a goal being scored is certainly dependent on the current score and there are variables which increase goal expectation, such as shots on goal (Pratas et al., 2016), an early red card (Bar-Eli, Tenenbaum & Geister, 2006) and an early away goal (Dixon & Robinson, 1998). These findings support the suggestion that in order to predict future performance and outcome based on past performances, it is not enough to analyse merely what happened in a match, but it is also important to know when events occurred. Thus, in order to ensure reliable results analysis should include the temporal dimension of game actions.

One of the challenges in match analysis is to progress from the retrospective analysis of performance, based on either discrete performance indicators or positional data collected continuously, to the prediction of future performances and outcomes. Either perspective (continuous or discrete) of analysis may provide relevant information about players' behaviour and the two should be seen as being complementary. The use of different kinds of data will provide a better understanding of the factors that influence the time of goal scoring in football. This information may be useful for predicting (for example, using duration probabilistic models) match scenarios involving a particular opponent and preparing for strategic planning for the following match.

2.6. Conclusions

Research studies conducting football match analysis have examined goal scoring patterns from two different perspectives: the static and the dynamic. Football's inherent randomness makes analysis even more impactful and what makes a difference perhaps is not the data itself but rather the capacity to use this data to formulate a theory explaining how a team increases its chances of winning matches. Certain significant performance indicators associated with goal scoring are highlighted in this review and may be important in a given game context but would probably be irrelevant in another, depending on a number of factors related to teams' quality and playing style, the importance of the match, among others. In football analytics the way a team plays dictates which performance indicators are significant. The challenge is to find out how the relevance of different performance indicators changes according to match context.

Further research should continue to explore dynamic methods of analysis with the focus on the time of goal scoring and the evaluation of the relationship between different match events. The standardisation of some tactical metrics (for example, index

of game control, passing activity) and collective variables (for instance, surface area, inter-player coordination) might provide more consistent results regarding the goal scoring process and provide a better understanding about individual and collective behaviours which influence this process. Goal scoring is the main objective of a football match and indeed perhaps it can be said to be its sole purpose and it is the most valid performance indicator determining team success. Thus, all events which take place on the field should be evaluated taking into account the principal objective of the game: to score more goals than one's opponents.

The static approach, based on discrete frequency data may be suitable for the description and comparison of performance profiles associated with team success, but it provides limited information about the game process and performance variability during the match. The dynamic approach, which involves the analysis of the space-time coordination between players and teams in association with goal scoring opportunities and examines the variety of factors whose interaction may influence scoring dynamics, provides the basis for a better understanding of the game process and provides information on how the game outcome develops over time.

Currently, a huge amount of data is collected during matches, but it is not always clear how these data should be processed with a view to providing coaches, match analysts and players with relevant information. Finding more suitable methods for data analysis is an urgent task but there can be no progress if game analysts are not prepared to formulate pertinent questions and put forward relevant issues which need to be addressed in the field of match analysis.

2.7. References

Alberti, Alberti, G., Iaia, F. M., Arcelli, E., Cavaggioni, L., & Rampinini, E. (2013). Goal scoring patterns in major European soccer leagues. *Sport Sciences for Health, 9*(3), 151-153.

Armatas, V., & Pollard, R. (2014). Home advantage in Greek football. *European Journal of Sport Science, 14*(2), 116-122.

Armatas, V. & Yiannakos, A. (2010). Analysis and evaluation of goals scored in 2006 World Cup. *Journal of Sport and Health Research, 2*(2),119-128.

Armatas, V., Yiannakos, A., & Hatzimanouil, D. (2007). Record and evaluation of set plays in European Football Championship in Portugal 2004. *Inquiries in Sport*

and Physical Education, 5(2), 302-307.

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

Bangsbo, J. (1994). The physiology of soccer - with special reference to intense intermittent exercise. *Acta Physiologica Scandinavian*, 151 (Suppl. 619): 1-155.

Bangsbo, J., & Peitersen, B. (2000). *Soccer systems and strategies*. Champaign, IL: Human Kinetics.

Bar-Eli, M., Tenenbaum, G., & Geister, S. (2006). Consequences of Player's Dismissal in Professional Soccer: A Crisis-Related Analysis of Group-Size Effects. *Journal of Sports Sciences*, 24(10), 1083-1094.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., & Matthews, I. (2014). "Win at home and draw away": Automatic formation analysis highlighting the differences in home and away team behaviors. In *MIT sloan sports analytics conference*, Boston.

Carling, C., Reilly, T., & Williams, A. (2009). *Performance assessment for field sports*. London: Routledge.

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.

Cintia, P., Pappalardo, L., Pedreschi, D., Giannotti, F. & Malvaldi, M. (2015). The harsh rule of the goals: data-driven performance indicators for football teams. In: IEEE international conference on paper presented at the data science and advanced analytics (DSAA), 19–21 October 2015.

Di Salvo, V., Pigozzi, F. González-Haro, C., Laughlin, M.S., & De Witt, J.K. (2013). Match Performance Comparison in Top English Soccer Leagues. *International Journal of Sports Medicine, 34*(6), 526-532.

Dixon, M., & Robinson, M. (1998). A birth process model for association football matches. *Journal of the Royal Statistical Society*: Series D (The Statistician), *47*(3), 523–538.

Duarte, R., Araujo, D., Correia, V., & Davids, K. (2012). Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Medicine, 42*(8), 633–642.

Duarte, R., Araujo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. *Human Movement Science, 31*(6), 1639–1651.

Ferguson, A. (2013). *Alex Ferguson: My Autobiography*. London: Hodder and toughton, pp.402.

Fernando, T., Wei, X., Fookes, C., Sridharan, S., & Lucey, P. (2015). Discovering methods of scoring in soccer using tracking data. Paper presented at the Large-Scale Sports Analytics, Sidney.

Folgado, H., Lemmink, K.A., Frencken, W., & Sampaio, J. (2014). Length, width and centroid distance as measures of teams tactical performance in youth football. *European Journal of Sport Science, 14*(Suppl 1), S487–S492.

Fradua, L., Zubillaga, A., Caro, O., Fernandez-Garcia, A.I., Ruiz-Ruiz, C., & Tenga, A. (2013). Designing small-sided games for training tactical aspects in soccer: extrapolating pitch sizes from full-size professional matches. *Journal of Sport Science, 31*(6), 573–581.

Frencken, W.G.P., & Lemmink, K.A.P.M. (2008). *'Team kinematics of small-sided soccer games: a systematic approach'*, in T. Reilly and F. Korkusuz (eds), Science and Football VI (pp. 161–6). London: Routledge.

Frencken, W., Lemmink, K., Delleman, N., & Visscher, C. (2011). Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science, 11*, 215-223.

Goncalves, B., Figueira, B.E., Macas, V., & Sampaio, J. (2014). Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of Sport Science, 32*(2), 191-199.

Goncalves, B., Marcelino, R., Torres-Ronda, L., Torrents, C., & Sampaio, J. (2016). Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games. *Journal of Sport Science, 34*(14), 1346-1354.

Gomez, M. A., Gomez-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.

Grunz, A., Memmert, D & Perl, J. (2012). Tactical pattern recognition in soc- cer games

by means of special self-organizing maps. *Human Movement Science, 31*(2), 334–343.

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. *Journal of Physical Education & Sport, 11*(1), 32-41.

Johnson, K., & Murphy, A. (2010). Passing and goal scoring characteristics in the Australian A-League. *Journal of Science and Medicine in Sport, 12, e118.*

Lago-Peñas, C. & Dellal, A. (2010). Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *Journal of Human Kinetics*, 25: 93-100.

Lago-Peñas, 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 and Medicine, 9*(2), 288-293.

Lago-Ballesteros, J., & Lago-Peñas, C. (2010). Performance in Team Sports: Identifying the Keys to Success in Soccer. *Journal of Human Kinetics*, 25, 85-91.

Lago-Peñas, 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-Peñas, C., & Gomez-Lopez, M. (2014). How important is it to score a goal? The influence of the scoreline on match performance in elite soccer. *Perceptual and Motor Skills, 119*(3), 774-784.

Lopes, A., Fonseca, S., Lese, R., & Baca, A. (2015). Using Voronoi diagrams to describe tactical behaviour in invasive team sports: an application in basketball. *Cuadernos de Psicología del Deporte, 15*(1), 123-130.

Kempe, M., Vogelbein, M., Memmert, D., & Nopp, S. (2014). Possession vs. Direct Play: Evaluating Tactical Behavior in Elite Soccer. *International Journal of Sports Science, 4*(6A), 35-41.

Kempe, M., Grunz, A., & Memmert, D. (2015). Detecting tactical patterns in basketball: comparison of merge self-organising maps and dynamic controlled neural networks. European Journal of Sport Science, 15(4):249-255

Krustrup, P., Mohr, M., Steensberg, A., Bencke, J., Kjaer, M. & Bangsbo, J. (2006). Muscle and blood metabolites during a soccer game: implications for sprint

performance. *Medicine and Science in Sports & Exercise*, 38(6), 1165–1174.

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

Memmert, D., Lemmink, A.P.M.K., & Sampaio, J. (2016). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine, 47*(1), 1-10.

Nevo, R., & Ritov, Y. (2012). Around the goal: Examining the effect of the first goal on the second goal in soccer using survival analysis methods. *Journal of Quantitative Analysis in Sports, 9*(2), 65-177.

Njororai, W. (2014). Timing of Goals Scored in Selected European and South American Soccer Leagues, FIFA and UEFA Tournaments and the Critical Phases of a Match. *International Journal of Sports Science*, 4(6A), 56-64

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), 1-27.

Passos, P., Araújo, D., & Davids, K. (2013). Dyadic systems as dynamic systems in individual and team sports. *Routhedge Handbook of Sports Performance Analysis*. London and New York: Routledge.

Pfeiffer, M., & Perl, J. (2006). Analysis of tactical structures in team handball by means of artificial neural networks. *International Journal of Computer Science in Sport, 5*(1), 4–14.

Pratas, J. M., Volossovitch, A., & Carita, A. I. (2016). The effect of performance indicators on the time the first goal is scored in football matches. *International Journal of Performance Analysis in Sport, 16*(1), 347-354

Prieto, J., Gómez, M.A., & Sampaio, J. (2015). From a Static to a Dynamic Perspective in Handball Match Analysis: a Systematic Review. *The Open Sports Sciences Journal, 8*, 25-34.

Redwood-Brown, A. (2008). Passing patterns before and after goal scoring in FA premier league soccer. *International Journal of Performance Analysis in Sport, 8*(3), 172-182.

Russell, M., Benton, D., & Kingsley, M. (2011). The Effects of Fatigue on Soccer Skills Performed During a Soccer Match Simulation. *International Journal of Sports Physiology and Performance, 6*(2), 221-233.

Sampaio, J., Ibáñez, S., & Lorenzo, L. (2013). Basketball. *Routhedge Handbook of Sports Performance Analysis*. London and New York: Routledge.

Sampaio, J., Lago, C., Goncalves, B., Macas, V.M., & Leite, N. (2014). Effects of pacing, status and unbalance in time motion variables, heart rate and tactical behaviour when playing 5-a-side football small-sided games. *Journal of Science and Medicine in Sport, 17*(2), 229–233.

Sarmento, H., Marcelino, R., Anguera, M.T., Campaniço, J., Matos, N., & Leitão, J.C. (2014). Match analysis in football: a systematic review. *Journal of Sports Sciences, 29*(20), 1831-1843.

Smith, M. R., Coutts, A. J., Merlini, M., Deprez, D., Lenoir, M., & Marcora, S. M. (2016). Mental fatigue impairs soccer-specific physical and technical performance. *Medicine & Science in Sports & Exercise, 48*(2), 267–276.

Taylor, J. B., James, N., & Mellalieu, S. D. (2005). *National analysis of corner kicks in English premier league soccer.* En T. Reilly, J. Cabri and D. Araujo (Eds.), Science and Football V, The procedings of the Fifht World Congress on Science and Football (pp. 225–230). Londres: Routlegde.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010a). Effect of playing tactics on goal scoring in Norwegian professional soccer. *Journal of Sports Sciences, 28*(3), 237-244.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010b). Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science, 10*(4), 269-277.

Tenga, A., & Sigmundstad, E. (2011). 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*, 545-552. doi:

Volossovitch, A. (2013). Handball. In T. McGarry, P.O'Donoghue, & J.Sampaio (Eds.) *Handbook of sports performance analysis*. (380-392). London: Routledge.

Volossovitch, A. & Ferreira, A.P. (2013). Da descrição estática à predição dinâmica. A evolução das perspectivas de análise da performance nos jogos desportivos

coletivos. In A. Volossovitch & A. P. Ferreira (Ed.). *Fundamentos e aplicações em análise do jogo.* Cruz Quebrada, Portugal: Edições FMH.

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.

Yiannakos, A., & Armatas, V. (2006). Evaluation of goal scoring patterns in the European Championship in Portugal 2004. *International Journal of Performance Analysis in Sport, 6*, 178-188.

CHAPTER 3

THE EFFECT OF PERFORMANCE INDICATORS ON THE TIME THE FIRST GOAL IS SCORED IN FOOTBALL MATCHES[2]

[2]Published as:
Pratas, J. M., Volossovitch, A., & Carita, A. I. (2016). The effect of performance indicators on the time the first goal is scored in football matches. *International Journal of Performance Analysis in Sport, 16*(1), 347-354.

3. The effect of performance indicators on the time the first goal is scored in football matches

3.1. Abstract

Scoring the first goal in football could be crucial for winning a match. The aim of this study was to identify the performance indicators that influence the time the first goal is scored in high-level football matches. A total of 240 matches of the Portuguese Premier League played in 2009/10 season were analysed. The difference between the total number of goals scored and conceded during the competition up until the match considered, total ball possession time, shots on goal, set plays, disciplinary sanctions and substitutions were selected as predictor variables for the time the first goal is scored and loaded on a Cox regression model with time-dependent covariates. A greater goal difference ($p < 0.05$) and a larger number of shots on goal ($p < 0.01$) had a positive significant influence on the time the first goal was scored in the match by home teams; at the same time disciplinary sanctions ($p < 0.01$) and substitutions ($p < 0.01$) had a negative significant effect on the time of the first goal.

Keywords: first goal, soccer, football, survival analysis

3.2. Introduction

In a low-scoring sport such as football, scoring the first goal in a match may be crucial winning it. For instance, Armatas, Yiannakos, Papadopoulou and Skoufas (2009) found that teams, which scored the first goal in a match, won 71.43% of matches played in the Greek Super League during the 2006/07 season. According to findings of Molinuevo and Bermejo (2012), the same phenomenon was evident in the Professional Spanish Football League from the 2005/06 to 2009-10 seasons, during which period 74.47% of matches were won by the team that scored the first goal in the match. The findings of Molinuevo and Bermejo (2012) also showed that a significantly high percentage of first goals were scored by home teams (56.96%). Home advantage for scoring the first goal in a match was also reported by Tenga (2012), who analysed the matches played in English Premier League (EPL) in the 2008/09 to 2009/10 seasons and in the Norwegian Top League (NTL) in the 2009/10 to 2010/11 seasons. In Tenga's study, the lowest percentage of first goals scored (58.88%) was recorded in the EPL in the 2008/09 season and the highest percentage of first goals scored (61.64%) was recorded in the NTL in the 2009/2010 season.

Different models and tools have been suggested for the analysis of the factors that influence goal scoring in football, e.g., chi-square test, linear and logit/probit regression (Michailidis, Michaildis and Primpa, 2013; Lago-Peñas and Gómez-López, 2014; Tenga, Holme, Ronglan and Bahr, 2010a,b). The use of the above-mentioned statistical methods provide a static view of performance indicators, while they do not take into account the dynamic nature of a football match and the influence of certain events on the subsequent course of the match. Some of these regression techniques may lead to inadequate modelling of the causal relationships between variables over time. In order to analyse changes of performance over time it is necessary to use play-by-play data and methods that take into account the way in which teams' actions at any given moment influence the course and the outcome of the match. One such method is survival analysis, also known as time-event analysis (Castilla, 2007). Because survival analysis explicitly incorporates time as a variable of interest, it is more flexible and suitable for extracting and using information from the match progresses than methods more commonly used in football performance analysis (Venturelli, Schena, Zanolla and Bishop, 2011).

Nevo and Ritov (2012) have used survival analysis (Cox model) to verify the relationship between the time of the first goal and the time of the second goal scored in a match. These authors concluded that first goal occurrence could either expedite or impede the scoring of the next goal, depending on the time it was scored.

Despite the first goal importance has been highlighted in several studies, the research in notation analysis in football still paid relatively little attention to this specific event. The first goal has been predominately analysed from the perspective of it impact on the match outcome (Armatas et al., 2009; Molinuevo and Bermejo, 2012; Michailidis et al., 2013). However, it would be interesting to evaluate not only the effect of performance indicators on the first goal being scored, but also to examine which performance indicators influence the time of the first goal being scored in different competitive contexts and how this influence occurs. This knowledge could help coaches to anticipate match status evolution and adopt in a timely manner the most appropriate attacking and defensive strategies for their team.

Thus, the aim of this study was to identify the performance indicators that influence the time the first goal is scored in high-level football matches in the Portuguese Premier League during the 2009-2010 season.

3.3. Methods

3.3.1. Sample

The sample consisted of 240 matches played in the Portuguese Premier League during the 2009/10 season. All data were obtained from the freely accessible website of the Portuguese Soccer League, “Officials Statistics Sagres League - LPFP/WTVision/Amisco" (www.lpfp.pt). Ethics approval for all study procedures was granted by the university’s ethics committee (Approval Number: 54/2015).

3.3.2. Procedure and data analysis

Survival analysis and Cox models with time-dependent covariates (Fisher and Lin, 1999) were used for modelling the time, which elapsed from the beginning of a match until the first goal was scored. The Cox model with time-dependent covariates, an extension of traditional Cox model (Cox, 1972), allows for the inclusion in modelling hazard function fixed covariates, whose values remain unchanged throughout the period of observation, and also time-dependent covariates, whose values change

over time. In the present study, the home team was considered as the reference team and the first goal of the match, when scored by the home team, was considered as the event of interest. If the away team scored the first goal, this was considered as a censored observation. The response variable in the Cox model was the time, which elapsed from the beginning of match to occurrence of the event of interest. Matches in which no goals were scored or goals were scored during stoppage time in the first or second half were not considered. Almost 12% of matches (28 out of 240) ended goal-less. Thus, data were collected on 212 matches, in which a first goal was scored either by the home team or the away team. It should be also mentioned that almost half of these matches (90) were censored because the away team scored the first goal in the match. Table 1 provides the definitions of variables.

Table 1. Definitions of variables.

Variable	Description
time	The first goal scored in the match, expressed as time in minutes elapsed since the start of the match.
event	A dummy variable: assumes a value of one if the home team scores the first goal of the match, and zero otherwise.
goal difference	Goal difference for home teams is expressed by the difference between goals scored and goals conceded during the competition until the observed match.
ball possession	Cumulative length or total time of ball possession expressed as time in minutes. The match period was divided into six periods of fifteen minutes each.
shots on goal	Number of shots on goal by home teams.
set plays	Number of set plays (corner kicks and free kicks) by home teams.
sanctions	A dummy variable: assumes a value of one if home team incurred disciplinary sanctions (yellow cards and red cards), and zero otherwise.
substitutions	A dummy variable: assumes a value of one if the home team made substitutions, and zero otherwise.

The univariate Cox regression model was used to investigate the effect of each covariate on the response variable. All variables that were shown to be significant in univariate analysis were then used in a final multivariate Cox regression model to evaluate survival predictors for first goal being scored by teams playing at home. Descriptive statistics and Student t-test were used to provide basic information about the variables in the dataset. To estimate all the models the R survival package version 2.37-

7 (Therneau, 2015) was used. All statistical analysis was performed using Software R, version 3.0.2

3.4. Results

Table 2 illustrates the outcomes for home and away teams that scored the first goal in a match. At least one goal occurred in 212 out of 240 matches observed (88.33%) and the team, which scored the first goal in the match won 70 % of matches. Home teams scored first in 57.5 % of matches and won 75% of matches, while away teams that scored the first goal achieved victory in 62% of matches. For home teams, the mean and median of overall survival time were 30.24 and 25.5 minutes, respectively; while for away teams the mean was 30.77 minutes and the median 24.0 minutes. No significant differences between home and away teams with regard to the mean time the first goal was scored ($t(210)=0.160$; $p=0.873$) were recorded.

Table 3 presents estimates of the parameters for the univariate and multivariate Cox regression models with time-dependent covariates. The effects of risk factors on the survival probability of the first goal being scored were analysed by using univariate models. A significant effect was found for goal difference, shots on goal, disciplinary sanctions and substitutions. All variables, which were found to be significant in the univariate analysis, were then used in a multivariate model, in which all variables remained significant. Home teams, presenting a better difference between the total number of goals scored and goals conceded in League matches than their opponents showed an increased risk (Hazard ratio (HR)=1.011) as far as scoring the first goal is concerned, i.e. for every goal contributing to difference the hazard increased by 1.1%. Moreover, for every shot on goal performed by home teams the hazard of scoring the first goal (HR=1.267) increased by 26.7%. At the same time, home teams that incurred disciplinary sanctions (HR=0.386) and made substitutions (HR=0.303) presented a decreased risk of scoring the first goal of the match. Total time of ball possession and set plays performed by home teams had no significant effect on the time the first goal was scored.

Table 2. Outcomes for home and away teams which scored the first goal in the match

FIRST GOAL	Final score	n	%	Total (N=212)
Home team 122 (57.54%)	Win	91	74.59	Win: 147 (70%) Draw: 39 (18%) Loss: 26 (12%)
	Draw	19	15.57	
	Loss	12	9.48	
Away team n= 90 (42.46%)	Win	56	62.22	
	Draw	20	22.22	
	Loss	14	15.56	

Table 3. Cox Regression models with time-dependent covariates

Variables	Univariate			Multivariate		
	Coefficent	Hazard Ratio	P-value	Coefficent	Hazard Ratio	P-value
Goal difference	0.012	1.013	0.018*	0.011	1.011	0.038 *
Ball possession	-0.003	0.996	0.225	/	/	/
Shots on goal	0.238	1.27	0.017*	0.237	1.267	0.002 **
Set plays	-0.119	0.887	0.333	/	/	/
Sanctions	-0.894	0.408	0.018*	-0.950	0.386	0.001 ***
Substitutions	-1.229	0.292	0.000***	-1.193	0.303	0.001 ***

Statistical significance: ***P<0.0001; **P<0.01; *P<0.05

3.5. Discussion

The findings of this study confirmed the importance of scoring the first goal for teams playing at home in matches of the Portuguese Premier League. Home teams scored first in 57.5 % of matches and won 75% of games. This is in keeping with results of Tenga (2012), who reported home advantage as significant for teams scoring the first goal. Armatas et al., (2009) and Molinuevo and Bermejo (2012) also concluded that teams, which scored, first presented a higher probability of winning a match, especially when playing at home.

Our results demonstrated that performance indicators such as goal difference, shots on goal, sanctions and substitutions, enable the time the first goal is scored by home teams to be predicted. The goal difference, i.e., the difference between the total number of goals scored and goals conceded during a competition is a measure of how well the team has performed over the season and can be used as an indicator of team

success or overall team strength. Successful teams are distinguished by their outstanding attacking and defensive capability. Our findings demonstrated the importance of total goal difference for scoring the first goal in a match. Teams with a higher total goal difference comparatively to their opponents had a higher risk of scoring first in a match. For each increase in goal difference by one goal, the hazard ratio of teams scoring the first goal increases by 1.1% (HR= 1.011). Therefore, total goal difference is the indicator that should be taken into account by coaches in opponent team analysis and defining strategy for the next match.

The results of this study demonstrate that shots on goal performed by home teams with a higher number of shots on goal showed a higher risk in term of hazard ratio for scoring the first goal than home teams with a fewer number of shots on goal (for every shot on goal performed by home teams the risk of scoring the first goal increased by 26.7%) (see Table 3). It means that teams, which performed more shots, scored the first goal earlier in the match, so a higher number of shots on goal accelerates the first goal scoring in a football match.

One of the interesting findings of this study is that there was no significant effect of the total time of ball possession or set plays on the time of the first goal was scored. Our findings provide further evidence for the importance of practicing open play situations. Nevertheless, set plays should not be totally discarded due to their potential productivity, even though the occurrence of set plays is less frequent. Previous studies (Lago-Peñas, Lago-Ballesteros, Dellal, and Gomez, 2010; Collet, 2013) showed that successful teams and teams with greater total time of possession demonstrated better shot and scoring performance. However, it is still not clear what type of attack (e.g. counter-attack or position play) leads more frequently to the first goal being scored when teams played against different types of defense (Tenga et al., 2010a,b).

According to Dobson and Goddard (2010) and Anders and Rotthoff (2011) the aggressive style of play can be advantageous for the teams and may influence positively the match outcome, but only if the teams are aggressive without being carded. Our study suggests that disciplinary sanctions have a significant effect on time the first goal is scored in a football match. It was demonstrated that home teams, which incurred disciplinary sanctions, presented a decreased risk (61.4%) of scoring the first goal of the match.

This is borne out by Bar-Eli, Tenenbaum and Geister (2006) and Anders and Rotthoff (2011), who also found that occurrences of disciplinary sanctions in matches of the German Bundesliga significantly decreased the probability of scoring chances and victory for the penalised team, while the scoring rate of the opposing team increased slightly. Moreover, Bar-Eli et al. (2006) found that when the score was tied and the home team had received a red card, the chance of the away team and home team scoring the first goal was 54.2% and 45.8%, respectively. Anders and Rotthoff (2011) reported that a yellow card shown to the home team significantly decreased their probability of winning.

Home teams that made substitutions, compared with those that made no substitution, presented a decreased risk of scoring the first goal of the match. For example, an HR of 0.303 in the model is equal to a 69.7% reduction of risk of scoring the first goal of the match for home teams. These results suggest that in the case of a tied score (0-0) a tactical intervention like a substitution, frequently carried out by a coach, does not allow for the improvement of the team's offensive performance. This means that coaches should consider the relatively greater importance of players who start the match and the relatively lesser importance of substitutes as far as scoring the first goal of a match is concerned.

Similar results were reported by Del Corral, Barros and Prieto-Rodríguez (2008), who analysed substitutions in matches of the BBVA Spanish Football League played in the 2004-2005 season. The authors found that the timing of the first substitution was based on goal difference, and teams that were trailing made substitutions earlier. It was also found that defensive substitutions occurred later in the match, while attacking substitutions occurred earlier.

It seems that when the match status is a draw, coaches tend to hold off from making substitutions until later in the match. However, from a tactical and strategic perspective it remains unclear whether or not the first substitution depends on goal difference and what type of substitution (defensive, neutral or attacking) has the greatest influence on the time the first goal is scored. Thus, further research is needed in this field to provide an explanation of how coaches can use substitutions to increase the probability of scoring the first goal of the match.

These findings may be useful for coaches when they devise attacking tactics and make decisions during the match. It may also be advantageous to conduct tactical play

exercises that enhance shot and score opportunities more often in training sessions and highlight the importance of defensive drills and fair play. Further research is needed to provide an understanding of how different performance indicators influence the time of scoring in football in different match scenarios, taking into account the dynamics of teams' performance during the match.

3.6. Conclusions

It was found that home teams presenting a greater difference between the total number of goals scored and goals conceded during the competition and a higher number of shots on goal had a higher risk of scoring the first goal in the match. Evidence was also found for the negative effect of substitutions and disciplinary sanctions on the time of the first goal was scored. Making substitutions and receiving disciplinary sanctions delays the scoring the first goal of the match by teams.

3.7. References

Anders, A., and Rotthoff K. W. (2011). Yellow Cards: Do They Matter? **Journal of Quantitative Analysis in Sports**, 7(1).

Armatas, V., Yiannakos, A., Papadopoulou, S., and Skoufas, D. (2009). Evaluation of goals scored in top ranking soccer matches: Greek "Superleague" 2006-07. **Serbian Journal of Sports Sciences**, 3(1-4), 39-43.

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

Castilla, E.J. (2007). **Dynamic Analysis in the Social Sciences**, 1st Edition. USA: Elsevier, Inc.

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.

Cox, D.R. (1972). Regression models and life tables. **Journal of the Royal Statistical Society**, 34(2), 187-220.

Del Corral, J., Barros, C. P., and Prieto-Rodríguez, J. (2008). The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League. **Journal of Sports Economics**, 9(2), 160-172.

Dobson, S., and Goddard, J. (2010). Optimizing strategic behaviour in a dynamic setting in professional team sports. **European Journal of Operational Research**, 205(3), 661-669.

Fisher, L.D., and Lin, D.Y. (1999). Time-dependent covariates in the cox proportional-hazards regression model. **Annual Reviews Public Health***, 20*, 145-57.

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

Lago-Peñas, C., and Gomez-Lopez, M. (2014). How important is it to score a goal? The influence of the scoreline on match performance in elite soccer. **Perceptual and Motor Skills**, 119(3), 774-784.

Michailidis, Y., Michaildis, C., and Primpa, E. (2013). Analysis of goals scored in European Championship 2012. **Journal of Human Sport and Exercise**, 8(2), 367-375.

Molinuevo, J. S., and Bermejo, J. P. (2012). El efecto de marcar primero y la ventaja de jugar en casa en la liga de fútbol y en la liga de fútbol sala de España. **Revista de Psicología del Deporte**, 21(2), 301-308.

Nevo, R., and Ritov, Y. (2012). Around the goal: Examining the effect of the first goal on the second goal in soccer using survival analysis methods. **Journal of Quantitative Analysis in Sports**, 9(2), 65-177.

Tenga, A., Holme, I., Ronglan, L. T., and Bahr, R. (2010a). Effects of match location on playing tactics for goal scoring in professional Norweigan soccer. **Journal of Sports Behavior**, 33(1), 89- 108.

Tenga, A., Holme, I., Ronglan, L. T., and Bahr, R. (2010b). Effect of playing tactics on goal scoring in Norwegian professional soccer. **Journal of Sports Science***,* 28(3), 237-244.

Tenga, A. (2012). Chapter: First goal and home advantage at different levels of play in professional soccer (pg. 47-51). **World Congress of Performance Analysis of Sport IX**, Publisher: London & New York: Routledge Taylor & Francis Group, Editors: D. Peters, P. G. O'Donoghue.

Therneau, T. (2015). A Package for Survival Analysis in S. R package version 2.37-7, **http://CRAN.R-project.org/package=survival**.

Venturelli, M., Schena, F., Zanolla, L., and Bishop, D. (2011). Injury risk factors in

young soccer players detected by a multivariate survival model. **Journal of Science and Medicine in Sport**, 14(4), 293–298.

CHAPTER 4

ANALYSIS OF SCORING SEQUENCES IN MATCHES OF PORTUGUESE PREMIER LEAGUE[3]

[3]Submitted as:
Pratas, J.M., Volossovitch, A., & Carita, A. I. (In Press). Analysis of scoring sequences in matches of Portuguese Premier league. *Journal of Human Kinectics*.

4. Analysis of scoring sequences in matches of Portuguese Premier league

4.1. Abstract

The aim of this study is to examine the sequences of the first two goals scored in football matches in accordance with a range of different match contexts. Data from 1506 matches played in the Portuguese Premier League during six consecutive competitive seasons (2009-10 to 2014-2015) was analysed using a descriptive statistics and the chi-square test in order to verify the association between variables and a Cox regression analysis was used to predict the time the second goal is scored in function of the time of the first goal scored in the match and the scoreline.

The results revealed a higher frequency of the second goals being scored in the second half of a match (58%) and in the last 5-minutes periods of each half. A positive association has been found for home teams and score-doubling goals (58%), as well as for away teams and score-equalizing goals (56%). For home and away teams the score-doubling goal of a match was strongly and positively associated with a win outcome for home (93%) and away teams (92%), while the score-equalizing goals were associated with a draw (home and away teams: 44%) and loss outcome (home: 33% and away teams: 32%). Finally, the Cox model showed that if the first goal was scored in the second half of the match the probability of second goal being scored is three times more higher compared to the first half.

Keywords: game analysis, soccer, survival analysis

4.2. Introduction

In football, it has been demonstrated that the performance of teams can be influenced by the scoreline (Lago-Peñas, 2012; Gómez, *et al*., 2013). Football players perform significantly less high-intensity activity when winning than when losing or the score is tied (Lago et al., 2010). Teams had longer periods of possession in matches when they were losing than when they were winning (Lago-Peñas and Dellal, 2010; Lago-Peñas and Gomez-Lopez, 2014), teams play more in the attack zone and defensive zones when the score was level (Lago, 2009) and final-third entries were greater when teams were 1 goal down than when they were 1 goal up (Lago-Peñas and Gomez-Lopez, 2014). Findings also showed that shots on goal decreased by 14.1% and 14.45% when teams were 1 goal up and when the scores were level, respectively (Lago-Peñas and Gomez-Lopez, 2014) and when a team was drawing or winning, the probability of reaching the goal decreased by 43 and 53 percent, respectively (Lago-Ballestero et al., 2012).

The score evolution in time is an important situational factor that may influence the team performance during the match and it also allows to identify the critical incidents/moments of the game (Leite, 2013; Njororai, 2014). In low scoring games, like a football, each goal can be considered a critical incident that influences the course of the game (Ferreira, 2013). Studies conducted with different professional football leagues have shown that the team which scores first in a match has a higher probability of winning, and the home team is more likely to score first than the opposing team (Molinuevo and Bermejo, 2012; Tenga, 2012; Pratas et al., 2016). Most of these studies focused only on the analysis of first goal effect on match outcome, and did not examine the role of goals subsequently scored. A recent study of Lago-Peñas et al. (2016) that analysed games played in the most important European domestic leagues (English FA Premier League, French Ligue 1, Spanish La Liga, Italian Serie A and German Bundesliga) in the 2014/2015 season demonstrated that home teams scored first in 57.8 % of games and went on the obtain 84.85% of points won in these games. Instead, when away team scored first, they obtained only 76.25% of subsequent points.

The average number of goals scored per game in each of the major European football leagues is not more than three goals per match. Moreover, scores of 2-0 and 1-1 are two of the five most frequently recorded full-time scores at big-5 league level (England, France, Germany, Italy and Spain) in recent seasons (Anderson and Sally,

2013). Interestingly, in season 2015-16, 1-1 was the most common result in all big-5 leagues (http://www.soccerstats.com). The same trend was observed in the Portuguese Premier League. In the 2015/16 season, 20% of matches ended with total of two goals (i.e. the sum of total goals scored per game by both teams): 10% finished 1-1 (home draw, 38 matches), 7% 2-0 (home win, 21 matches) and 3% 0-2 (home defeat, 10 matches). The home advantage effect was confirmed in these matches. The home teams of the Portuguese Premier League scored 80 goals, while away teams scored 58 goals.

The advantage of playing at home may be related to several factors reported in literature, such as crowd support (Wolfson et al., 2005), travel (Pollard et al., 2008), familiarity with the game environment (Loughead et al., 2003), referees (Brandão et al., 2011) and territoriality aspects (Sampedro and Prieto, 2011). The importance of early leading in score is attributed to greater excitement and involvement of public, when the local team scores first, or a spectators disconnection of the match, when the visiting team is ahead on the score (Courneya, 1990). From the socio-psychological point of view, the local public support is considered by own fans as decisive, which can reinforce the self-esteem of players and increase social identity that leads to improvement of team performance and a reduction of the negative effects of stress and anxiety levels (Wolfson et al., 2005). From the strategic and tactical point of view, scoring first and gaining a score advantage allow winning team to extend the range of tactical options, for instance creating more counterattack opportunities against opposite team, which can opt for more risky strategies looking for tying the game.

The second goal might have a decisive impact on match outcome, which is not less than the influence of the first goal (Anderson and Sally, 2013). The second goal may be the equalizer (0-1 to 1-1, or 1-0 to 1-1) or the goal, which enables a team to double their advantage (1-0 to 2-0, or 0-1 to 0-2). The anticipation of temporal localization of the second goal associated to its nature (i.e., goal, which creates advantage or recovers from score disadvantage) seems to be useful for a timely adaptation of the team's tactics. However, there is a lack of attention in the literature to the second goal of match. One of the few studies, which analysed this goal, has been conducted by Nevo and Ritov (2012). The interaction between two random goal-scoring times (of the first and second goals) during a soccer match has been examined in 760 games played from 2008 to 2010 (two full seasons) in the English Premier League. Using survival analysis methods the authors reported that the first goal

occurrence could either expedite or impede the next goal scoring, depending on the time it was scored.

In order to provide a better understanding of which factors influence the second goal being scored, it would be useful to examine how the time of the second goal is associated with a range of different match contexts. A football match is a complex dynamic process, in which certain events influence the subsequent course of the match, and this influence must be considered in analysis. In order to explain the relationship between different events it is necessary to extract information as regards match progress using time as a variable of interest (Venturelli et al., 2011). Several studies have suggested using survival (also known as time-event) analysis for this purpose (Castilla, 2007; Nevo and Ritov, 2012). This tool involves the use of regression methods for explaining the relationship between independent variables and the time of an event of interest and it is considered to be the most suitable means for characterizing and explaining match progress than other tools commonly used in football performance analysis (Del Corral et al., 2008; Barros et al., 2009).

Thus, the first aim of this study is to analyse the association between the type of second goal scored in a match (i.e. goal that increases the leading team's advantage or the goal that re-establishes restores equality in the scoreline) and the match venue, and final match outcome. Secondly, the study aims to examine the influence of the time when the first goal is scored and the current scoreline on the probability of a second goal being scored in a match. This knowledge could help coaches to anticipate match scenarios and adopt in a timely manner the most appropriate tactics for influencing scoreline.

4.3. Material and Methods

4.3.1. Sample

The sample consisted of 1506 matches played in the Portuguese Premier League during six consecutive competitive seasons (from 2009-10 to 2014-2015). All data were collected from the official League website (http://www. ligaportugal.pt).

4.3.2. Statistical Analysis

Descriptive and chi-square analysis was performed to examine second goals scored during game period and the association between the type of second goal being

scored (score-doubling or score-equalizing) and match venue and final match outcome. In 29% of matches (443 out of 1506) a second goal was not scored. About 9% of matches (135 out of 1506) ended goalless and in 20% of matches (308 out of 1506) just one goal was scored. Thus, data were collected from 1063 matches in which the second goal was scored by home team (586).

A Cox proportional hazards (PH) model was used to estimate the time of the second goal as a function of the time of the first goal in a match and the scoreline.

In survival analysis, the home team was considered as the reference team and the second goal of the match, when it was scored by the home team, was considered as the event of interest. If the away team scored the second goal of the match, this goal was considered as a censored observation. The response variable in the Cox Model was the time elapsed from the time when the first goal in a match was scored to the occurrence of the second goal, considered as event of interest. Second goals scored in additional time in the first and second halves were recorded at the 45th and 90th minutes, respectively. Matches in which no goals or only one goal was scored were not considered. Almost half of 1063 matches (477) were censored because the away team scored the second goal of the match.

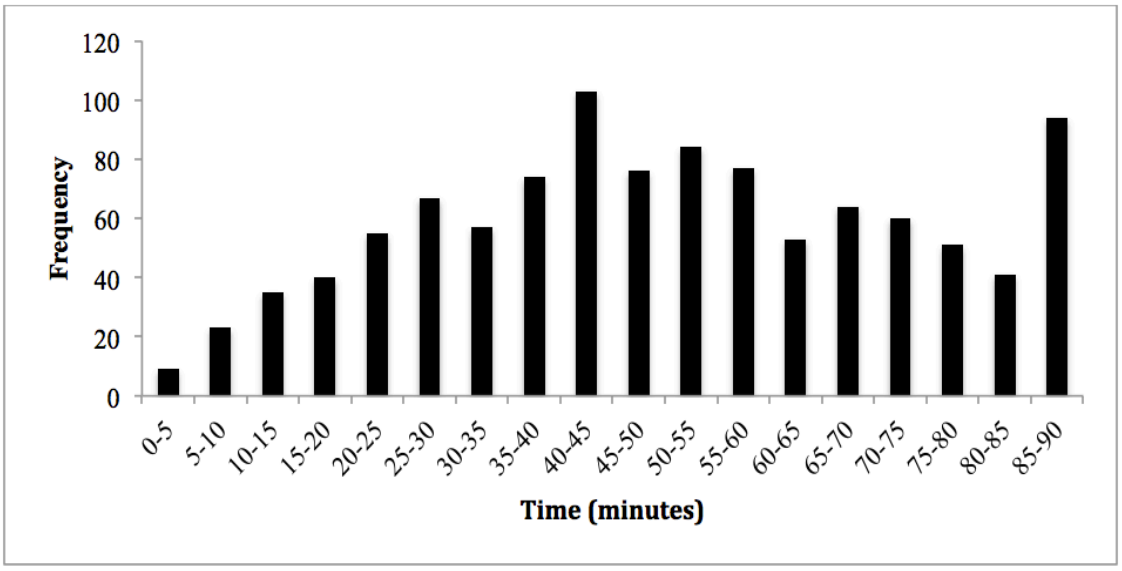
The Cox model relies on the assumption of the proportionality of hazards, implying that the factors analysed have a constant impact on the hazard over time (Broström, 2012). The proportionality assumption of data was checked to ensure a non-violation of the proportionality assumption. For this purpose the function cox.zph in the survival package, version 2.37-7 was used (Therneau, 2015). The p-value obtained in this function was significant ($p<0.05$) and indicated that the proportionality assumption was not met for the variable time of the first goal in a match. Since the time of the first goal is a continuous covariate, it was necessary to categorize it in order to ensure the proportionality assumption. First, the distribution of the time of first goals was checked using a histogram, which showed that the interval of time during which first goals are scored may reasonably be split into four equal-length intervals using the cut function. The time of the first goal in a match is expressed as time in minutes from the start of the match and categorized in accordance with four classes: first period [0-15], second period [16-30], third period [31-45] and second half [46-90]. The current scoreline was also considered as a categorical variable in accordance with two classes: a score-doubling goal (a second goal that doubles the advantage) and an equalizing.

All statistical analysis was performed using Software R, version 3.0.2. For all analyses, statistical significance was set at $p<0.05$.

4.4. Results

Data analyses showed that significantly more second goals were scored in the second half of matches ($\chi^2=28.81$, $p<0.05$) than the first half (58% v. 42%, respectively). The highest frequency of the second goal being scored was recorded in the last 5-min. periods of each half (see Figure 1).

Figure 1



Frequency of second goal of the match

As can be seen in Table I, 55% of second goals were scored by home teams and 45% by away teams. Home teams presented a greater percentage (58%) of second goals that doubled score advantage (1-0 to 2-0, or 0-1 to 0-2) as compared with away teams (44%). On the other hand, away teams scored a higher percentage (56%) of goals, which restored the score equality (1-0 to 1-1, or 0-1 to 1-1) in comparison with home teams (42%). The results of the chi-square test ($\chi2=20.23$, $p<0.05$) established a significant association between match venue and type of second goal scored.

Table 1

Frequency, percentage and chi-square value for second goal scored by home and away teams cross-referenced with type of second goal

Second goal of match	Type of second goal: Score-doubling 1-0 to 2-0, or 0-1 to 0-2	Type of second goal: Score-equalizing 1-0 to 1-1, or 0-1 to 1-1	n	χ^2
Home Team	338 (58%)[b]	248 (42%)[a]	586 (55%)	20.23*
Away Team	209 (44%)[a]	268 (56%)[b]	477 (45%)	
		Total	**1063**	

*Statistical significance: *P<0.05; [a] Negative association (adjusted residual <-1.96); [b] Positive association (adjusted residual >1.96)*

Table II shows a significant association between type of second goal scored and match outcome ($\chi 2=521.08$, $p<0.05$). The team, which scored the second goal, won the match in 59% of cases. The score-doubling goal was strongly positively associated with a win outcome and the equalizing goals with a draw and loss outcome.

Table 2

Frequency, percentage, and chi-square value for type of second goal cross-referenced with outcome of match

Type of second goal	Outcome: Win	Outcome: Draw	Outcome: Loss	Total	χ^2
Score-doubling	506 (93%)[b]	33 (6%)[a]	8 (1%)[a]	547	
Score-equalizing	123 (24%)[a]	225 (44%)[b]	168 (32%)[b]	516	521.08*
	629 (59%)	258 (24%)	176 (17%)	**1063**	

*Statistical significance: *P<0.05; [a] Negative association (adjusted residual <-1.96); [b] Positive association (adjusted residual >1.96)*

Table III displays a statistically significant association between to match outcome and type of second goal scored by home teams ($\chi 2=298.85$, $p<0.05$) as well as by away teams ($\chi 2=218.55$, $p<0.05$).

When home teams scored the score-doubling goal, they won 93% of matches, and when they scored the score-equalizing goal, they won 23%, drew 44% and lost 33% of matches. Away teams, when they scored the score-doubling goal, achieved victory in 92% of matches, and after scoring the equalizing goal they won 24%, drew 44% and lost 32% of matches.

Additionally, the frequency of the score-doubling goal of a match scored by home and away teams was positively associated with a win outcome and a negatively associated with draw and loss outcomes. Draw and loss outcomes were significantly and

positively associated with the equalizing goal of a match scored by both home and away teams and negatively associated with a win outcome.

Table 3

Frequency, percentage and chi-square value for type of second goal scored by home and away teams cross-referenced with match outcome

	Type of second goal		Outcome			Total	χ^2
			Win	Draw	Loss		
Home team	Score-doubling		314 (93%)[b]	17 (5%)[a]	7 (2%)[a]	338	
	Score-equalizing		58 (23%)[a]	108 (44%)[b]	82 (33%)[b]	248	298.85*
		Total	372	125	89	**586**	
Away team	Score-doubling		192 (92%)[b]	16 (7%)[a]	1 (1%)[a]	209	
	Score-equalizing		65 (24%)[a]	117 (44%)[b]	86 (32%)[b]	268	218.55*
		Total	257	133	87	**477**	
						1063	

*Statistical significance: *P<0.05; [a] Negative association (adjusted residual <-1.96); [b] Positive association (adjusted residual >1.96)*

The parameters estimated by using the Cox proportional hazard regression model (see Table IV) revealed a significant positive effect of the time of the first goal on the next goal occurrence, i.e. first goal occurrence in later game periods steadily increases the probability of the second goal in a match being scored. At the same time, for home teams the current scoreline had no significant effect on second goal occurrence in a match.

Table 4

Cox proportional hazard regression model

Variable		Coefficient	Hazard Ratio	P-value
Time of the first goal				
	0-15		1.00	
	16-30	0.349	1.419	0.001 *
	31-45	0.645	1.906	<0.001 *
	Second half	1.131	3.099	<0.001 *
Scoreline				
	Score-equalizing goal		1.00	
	Score-doubling goal	0.0001	1.00	0.998

*Statistical significance: *P<0.05*

4.5. Discussion

In the present study, the time of the second goal scored in football matches was analysed. As can be seen in Figure 1, in general, teams scored a larger number of second goals in the second half of matches, and the highest number of goals was scored during the last minutes of each half. Previous studies reported that the frequency of goals scored during a match was time-dependent, more goals were scored in the second half of matches and in the last 10- and 15-minute periods of each half (Armatas et al., 2009; Dobson and Goddard, 2010). The tendency for increasing scoring rates over the time of a match could be attributed to physiological and tactical factors that influence teams' performance. On the one hand, fatigue, which is greatest at the end of each half, leads to an increase in the number of technical and tactical errors, which may lead to more goal scoring opportunities. On the other hand, the little time remaining until the end of each half encourages players to use their last chances to attempt score a goal that also may influence match outcome.

In low-scoring games such as football, scoring the second goal in match could be crucial for winning. The results of this study confirm the effect of home advantage on the second goal being scored and subsequent victory in matches in the Portuguese professional Football League (see Table I). It was also found that match venue had an impact on the second type of goal scored. Score-doubling goals were positively associated with home matches, while equalizing goals were associated with away matches. Courneya and Carrón (1992) and Wolfson et al. (2005) suggested that an early lead in a match caused greater excitement and involvement of fans when the home team scored first. Score advantage reduces the negative effects of stress and anxiety and positively influences a team's performance, which could help home teams to enhance their score advantage.

The higher number of equalizing goals scored by away teams may be related to the defensive strategic behaviour adopted by home teams when they are winning and trying to maintain their score advantage. This tendency was reported by Dobson and Goddard (2010) for match played in the English Premier League and by Lago-Peñas and Gómez-López (2014) for the Spanish Professional League matches, where teams used defensive strategy when were winning and attacked more when were losing in order to maintain the score advantage or to change the current result, respectively. A

team that switches from defensive to attacking style usually increases its own and also opponent's probabilities to score (Armatas et al., 2009; Dobson and Goddard, 2010).

Home and away teams, which scored a score-doubling goal, won the majority of matches. Additionally, the score-doubling goals were positively associated with a winning match outcome. Since football is a low-scoring game, this result was expected. At the same time, the equalizing goals were positively associated with the draw and loss outcome; however, home teams and away teams both avoided defeat in approximately 68% of matches after scoring an equalizing goal, so this goal was extremely important.

Regardless of the type of goal (score-doubling or equalizing), scoring in football always has an impact on the subsequent course of the game, influencing players psychologically and frequently leading to tactical adjustments of a team's play. Survival analysis demonstrated that the first goal scored in match had a significant positive effect on the probability of the next goal being scored, and this probability was dependent on time, when the first goal of the match was scored (see Table 4). Cox model parameters revealed that the probability of the second goal of the match being scored by home gradually increased as a match progressed. When the first goal of the match was scored in the second or third 15-minute period of the first half, the probability of home teams scoring the second goal increased by 42% and 190.6% respectively, as compared with the situation when the first goal of match was scored in the first 15-minute period of the first half.

Moreover, if the first goal occurred only in the second half, the probability of the home team scoring the second goal increased threefold (309.9%) as compared with cases where the first goal of match was scored in the first 15 minutes. A similar tendency was reported for the Spanish Football League by Nevo and Ritov (2012), who suggested that, depending on the time when the first goal was scored, it could accelerate or delay the time when the next goal was scored. When a goal is scored before the 52nd minute, its effect on the probability of the next goal being scored is negative, while when the first goal is scored after that time the effect is positive. Furthermore, the results of the study demonstrate that once a goal is scored, another goal is less likely to be scored as compared with the situation where no goal has been scored yet (Nevo and Ritov, 2012).

This tendency can be explained by the influence of contextual factors on teams' performance during the match, which has been demonstrated by several studies (Collet,

2013; Pratas et al., 2016). For instance, Heuer and Rubner (2012) showed that during the last ten minutes of matches of the German Premier Football League (Bundesliga) players' behaviour depends significantly on the current score, representing an increasing offensive (or decreasing defensive) behavior (Heuer and Rubner (2012).

4.6. Practical application

The findings of the present study could prove useful for football coaches, as knowing that the occurrence of the first goal in later match periods steadily increases the probability the second goal of a match being scored by the home team, coaches may adjust their decisions related to substitutions or tactical options in terms of attack and defence. During practice football players should be prepared for scenarios of psychological pressure and physical fatigue, without reducing their self-efficacy in the last periods of training sessions.

4.7. Conclusions

The results of the present study demonstrate that the highest number of second goals (both score-doubling and equalizing) in a match was scored in the last 5-minute period of each half. Different types of second goal (score-doubling or equalizing) had differing types of impact on match outcome for home and away teams. The time of the first goal in a match had a significant effect on the time of the next goal scored by home teams. This paper clearly demonstrates the influence of match context (i.e. match venue, scoreline, and time of the first goal) on the probability of the second goal of a match being scored and also on the effect of the second goal on match outcome. Further research is needed to identify performance indicators, which are related to the time of the second goal being scored during football matches in different European football leagues.

4.8. References

Anderson C, Sally D. *The numbers game*. New York: Penguin Books; 2013.

Armatas, V., Yiannakos, A., Papadopoulou, S. and Skoufas, D. (2009). Evaluation of goals scored in top ranking soccer matches: Greek "Superleague" 2006-07. *Serbian J Sport Sci***,** 3:39-43.

Barros CP, Frick B, Passos J. Coaching for survival: the hazards of head coach careers in the German 'Bundesliga'. *Appl Econ*, 2009; 41:3303-3311.

Brandão R, Sidónio S, Krebs R, Araújo D, Machado AA. El significado de arbitrar: percepción de jueces de fútbol profesional. *Rev Psicol Deporte*, 2011; 20:275-

286.

Broström G. *Event History Analysis with R*. CRC Press , Taylor & Francis Group, 2012.

Castilla EJ. *Dynamic Analysis in the Social Sciences*, 1st Edition. USA: Elsevier, Inc; 2007.

Collet C. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007–2010. *J Sport Sci*, 2013; 31:123-136.

Courneya KS. Importance of game location and scoring first in college baseball. *Percept Motor Skill*, 1990; 71:624-626.

Courneya KS, Carrón AV. The home advantage in sport competitions: A literature review. *J Sport Exerc Psychol*, 1992; 14:13-27.

Cox DR. Regression models and life tables. *J Roy Stat Soc B*,1972; 34:187-220.

Del Corral J, Barros CP, Prieto-Rodríguez J. The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League. *J Sports Econ*, 2008; 9:160-172.

Dobson S, Goddard, J. Optimizing strategic behaviour in a dynamic setting in professional team sports. *Eur J Oper Res*, 2010; 205:661-669.

Heuer A, Rubner O. How Does the Past of a Soccer Match Influence Its Future? Concepts and Statistical Analysis. *PLoS ONE*, 2012; 7.

Lago, C. The influence of match location, quality of opposition, and match status on possession strategies in professional association football. *J Sport Sci*, 2009; 27:1463-1469.

Lago C, Casais L, Dominguez E, Sampaio J. The effects of situational variables on distance covered at various speeds in elite soccer. *Eur J Sport Sci*, 2010; 10:103-109.

Lago-Ballesteros J, Lago-Peñas C, Rey E. The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *J Sport Sci*, 2012; 30:1455-1461.

Lago-Peñas, C. The role of situational variables in analysing physical performance in soccer. *J Hum Kin*, 2012; 35:89-95.

Lago-Peñas, C, Dellal, A. Ball possession strategies in elite soccer according to the evolution of the match-score: the influence of situational variables. *J Hum Kin*, 2010; 25:93-100.

Lago-Peñas C, Gomez-Lopez M. How important is it to score a goal? The influence of the scoreline on match performance in elite soccer. *Percept Motor Skill*, 2014; 119:774-784.

Lago-Peñas, C., Gómez-Ruano, M., Megías-Navarro, D., et al. Home advantage in football: Examining the effect of scoring first on game outcome in the five major European leagues. *Int J Per An Sport*, 2016; 16(2):411-42.

Leite, W.S.S. Analysis of goals in soccer World Cups and the determination of the critical phase of the game. *Phys Educ Sport*, 2013; 11(3):247-253.

Loughead TM, Carrón AV, Bray SR, Kim A. (2003). Facility familiarity and the home advantage in professional sports. *Int J Sport Psychol Exerc Psychol,* **2003;** 1:264-274.

Molinuevo JS, Bermejo JP. The effect os scoring first and home advantage in professional spanish football and indoor soccer leagues. *Rev Psicol Deporte*, 2012; 21:301-8.

Nevo R, Ritov Y. Around the goal: Examining the effect of the first goal on the second goal in soccer using survival analysis methods. *J Quant An Sports*, 2012; 9:65-177.

Njororai, W. Timing of Goals Scored in Selected European and South American Soccer Leagues, FIFA and UEFA Tournaments and the Critical Phases of a Game. *Int J Sports Sci*, 2014; 4(6A):56-64.

Pollard R, da Silva CD, Nísio CM. Home advantage in football in Brazil: Differences between teams and the effect of distance traveled. *Brazilian J Soccer Sci*, 2008; 1:3-10.

Pratas JM, Volossovitch A, Carita AI. The effect of performance indicators on the time the first goal is scored in football matches. *Int J Per An Sports*, 2016; 16:347-54.

Sampedro J, Prieto J. La territorialidad como factor asociado a la ventaja de jugar en casa. Un estudio comparativo por regiones en la liga de fútbol y en la liga de fútbol sala de España. Motricidad. *Eur J Hum Movement*, 2011; 26:93-104.

Tenga A. First goal and home advantage at different levels of play in professional soccer. *World Congress of Performance Analysis of Sport IX*, London & New York: Routledge Taylor & Francis Group, Editors: D. Peters, P. G. O'Donoghue, 47-51; 2012.

Therneau T. A Package for Survival Analysis in S. R package version 2.37-7, 2015. Available at :http://CRAN.R-project.org/package=survival; accessed on 01.08.2016.

Venturelli M, Schena F, Zanolla L, Bishop D. Injury risk factors in young soccer players detected by a multivariate survival model. *J Sci Med Sport*, 2011; 14:293–298.

Wolfson S, Wakelin D, Lewis M. Football supporters' perceptions of their role in the home advantage. *J Sport Sci*, 2005; 23:365-374.

CHAPTER 5

PREDICTING KEY-GOAL SCORING IN FOOTBALL, BASED ON PERFORMANCE INDICATORS AND CONTEXTUAL FACTORS[4]

[4]Submitted as:
Pratas, J. M., Volossovitch, A., Carita, A.I. (In Press). Predicting key-goal scoring in football, based on performance indicators and contextual factors. *Journal of Frontiers in Psychology*.

5. Predicting key-goal scoring in football, based on performance indicators and contextual factors

5.1. Abstract

The aim of this article is to estimate the predictive model of time of key-goal scoring as a function of contextual variables in high-level football games of the Portuguese Premier League played during the 2015-2016 season. The cumulative incidence function was used to estimate the probability of each type of key-goal scored during the game. Game outcome was defined as win or draw, while the lost games were not considered. The influence of covariates (game location, accumulated goal difference and scoreline) on cause-specific hazards for scoring a key-goal was explored using survival analysis with a competing risk model.

The results suggest that for home teams probability of key-goal occurrence that leads to a win increases by 32%, while for teams with a positive accumulated goal difference during the competition up until the game considered probability decreases by 61%. For teams playing at home and those with a positive accumulated goal difference the probability of scoring a key-goal that tied the game increased by 65% and 43%, respectively. The findings and methods used in the present study could be useful for anticipating scenarios in games with similar competitive contexts and predicting players' and team performances.

Keywords: game analysis, soccer, survival analysis, key-goal, probability

5.2. Introduction

In order to create predictive models of performance in sports games, first of all it is necessary to describe the dynamics of team performance during the game and identify underlying factors. A review of the literature enabled the identification of two main phenomena, which were used to explain the variability of team performance during the game and determine the factors of this variability (Volossovitch, 2008). While one of these phenomenon, designated as momentum, is focused on the psychological effect of game events on individual and collective performance (Burke, Edwards, Weigand & Weinberg, 1997; Jones & Harwood, 2008), the other, designated as critical incident or critical moment, is associated with certain occurrences on the field which influence score dynamics and eventually the game outcome (Bar-Eli, Taoz, Levy-Kolker & Tenenbaum, 1992; Ferreira, 2013). In low-scoring games like football, each goal can be regarded as a critical incident that influences the course of the game.

The time/score relationship is a situational factor, which may influence consequent performance and thus should be considered as determinant information for identifying the critical incidents/moments of the game (Leite, 2013; Njororai, 2014). The analysis carried out on football attempted to evaluate the influences of various game events on scoring probabilities in different game periods.

Dixon & Coles (1997) were one of the first researchers to investigate the dependency between the scores of competing soccer teams. They showed that the joint distribution of the scores of neither team in a match can be well represented by the product of two independent marginal Poisson distributions for the home and away teams. The model proposed by Dixon & Coles (1997) includes team-specific attack and defence ability parameters and then uses independent Poisson distributions for numbers of goals scored. Therefore, the linear predictor for the number of goals scored by a specific team depends on both the parameters of the team itself and those of its competitor. Using data from Italian Serie A of the 1991-1992 season Karlis & Ntzoufras (2003) demonstrated that the number of goals in a football game involving two competing teams was subject to the bivariate Poission distribution. The models proposed allow for correlation between the two scores and the bivariate Poission distribution can improve model fit and prediction of the number of draws in football games. Later, in Karlis & Ntzoufras (2009), the same authors, using data from the English Premiership of the 2006-2007 season, rather than modelling the number of

goals directly, focused on the difference in the number of goals, and the application of Bayesian methodology for Skellam's distribution using covariates was discussed. Groll, Schauberger & Tutz (2015) pointed out that when using exactly the same model as Dixon & Coles (1997) with FIFA World Cup 2014 data, estimates of the attack and defence capabilities of the teams are negatively correlated. Therefore, although independent Poisson distributions are used for the score in one game, linear predictors and, accordingly, predicted outcomes are negatively correlated.

The time at which goals are scored predominantly focuses on score evolution during games. 15-min interval analysis highlighted a significant upward trend in the number of goals scored as time progressed (Dobson & Goddard, 2010). The number of goals scored was higher in the second half of the game and scoring rate was highest in the last 15-min of the game (Armatas, Yiannakos, Papadopoulou & Skoufas, 2009; Alberti, Iaia, Arcelli, Cavaggioni, Rampinini, 2013; Njororai, 2014). According to Dobson & Goddard (2010), a possible explanation for this is that players tend to experience increasing fatigue as games progress, increasing the likelihood of mistakes being made and therefore goals being scored. An alternative explanation is that scoring rate increases because of the rise in the probability of the two teams' scores being unequal. There is agreement among analysts, coaches and the football community at large that there are certain periods during a game during which the probability of a team scoring or conceding a goal is higher than other periods, while not all goals have equal value and meaning for the game.

Goal scoring in the last minute does not occur very often but according to Van Ours & Van Tuijl (2011), comparative to the home advantage, the effect of goal scoring in the last minute on the probability of a team winning a game is quite large. Using data from 1960 onwards, in full 'A' international games of eight national teams (Argentina, Brazil, England, France, Germany, Italy, the Netherlands, and Spain) they found a country-specific element in goal scoring in the last few minutes of football games in which the German national team were more likely to score in the "dying seconds" of a game than other national teams.

In a low-scoring sport such as football, scoring the first goal scored in a game may be crucial to winning it and an early goal raises the confidence level of a team as it gets to dictate the pace of the game knowing that it is ahead (Njororai, 2004; Michailidis, Michailidis & Primpa, 2013). For instance, Molinuevo & Bermejo (2012) state that in

the Professional Spanish Football League from the 2005/06 to 2009-10 seasons 74.47% of games were won by the team that scored the first goal in the game. A recent study conducted by Lago-Peñas, Gómez-Ruano, Megías-Navarro & Pollard (2016) showed the same phenomenon as evident in the most important European domestic leagues (the English FA Premier League, French Ligue 1, Spanish La Liga, Italian Serie A and German Bundesliga for season 2014/2015). It was found that home teams scored first in 57.8 % of games and went on to obtain 84.85% of the points won in these games. By contrast, when the away team scored first, they obtained only 76.25% of points won. These differences were further influenced by the game period in which the first goal was scored and teams that scored first ended games scoring an average of more 1.88 goals than their opponents. This illustrates the significance of scoring an early goal or being the first team to score.

With regard to goals, which in football are critical incidents by definition, there is no doubt that the most critical goal is that which enables a team to win or draw the match. Thus, the present study focuses on the goal which decides the result of the game (that is, a win or a draw), defined in this paper as "key-goal". The anticipation of temporal localisation of the key-goal associated with its nature (that is, a goal which in terms of the scoreline creates an advantage for a team or enables a team to recover from a disadvantage) seems to be useful for a timely change in team tactics. It seems that strategic behaviour at any stage of the game is found to be dependent on the current difference in scores and the amount of playing time that has elapsed (Dobson & Goddard, 2010; Van Ours & Van Tuijl, 2011). However, in order to predict which of the two teams in a match is more likely to win or lose the game, it is necessary to clarify what performance variables and game contexts positively influence the probability of a key-goal being scored.

To the best of our knowledge, this issue has not been addressed in performance-analysis research in football. Thus, the aim of present study was to estimate the predictive model of time of key-goal scoring as a function of contextual variables in high-level football games played in the Portuguese Premier League during the 2015-2016 season.

5.3. Methods

5.3.1. Sample

The sample consisted of 306 games played in the Portuguese Premier League in the 2015/16 season. Data were sourced from official overviews of games in the database at *http://www.zerozero.pt*, and confirmed using data available on another relevant website at *http://www.ligaportugal.pt*. Ethics approval for all study procedures was granted by the university's ethics committee (Approval Number: 54/2015).

5.3.2. Statistical Analysis

Survival analysis methods, i.e., competing risk models were used for modelling the time which elapsed from the beginning of a game until the time at which a key-goal was scored.

Competing risk models refer to the fact that a subject (for example, a football game) may involve the occurrence of an event (for example, a goal scored) for a number of reasons. Two methods can be used to model competing risks: (1) the cumulative incidence function; and (2) the cause-specific approach. The first of these is more descriptive, focusing on the probability of each event type occurring, while the second is more structural, focusing on the effect of covariates on the probability of each type of event occurring (De Wreede, Fiocco & Putter, 2011).

In the present study, the goal that defines the game outcome (that is, from the moment the result is established as a win or a draw) was considered as the event of interest. If at the end of the game no goals had been scored, the game was defined as 'censored', that is, fixed at 90 min, and therefore regarded as having terminated at the end of the 90^{th} min. Any goals scored in extra time of the first and second halves were considered as having been scored in the 45^{th} and 90^{th} minute of game, respectively. It should be mentioned that 7% of these games (21 out of 306) ended goal-less. Thus, data were collected on 285 matches, in wich a key goal was scored. In our study the period of time up until the key-goal was scored was considered. A key-goal may have either of two contrasting types of consequence: a goal that breaks a tie by creating a score advantage for one of the teams; or a goal that that leads to a draw result. Thus, the key-goal may lead to either of two outcomes: a win for one team; or a draw (see Figure 1.)

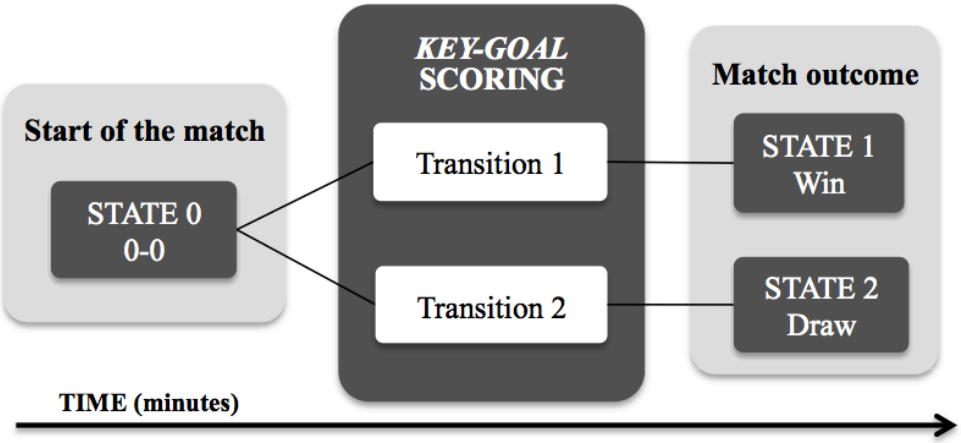


Figure 1. Competing risks model of key-goal scoring

The following covariates were considered in the analysis:game location, accumulated goal difference and scoreline. Table 1 provides the definitions of the variables used in the study.

Table 1. Definitions of variables used in the study

Variable	Description
id	The defining reference number of each game
time	The time when the key-goal was scored, expressed in minutes elapsed since the start of the game.
game location	A dummy variable which assumes a value of one if the home team scores the key-goal and zero if it was scored by the away team.
accumulated goal difference	The goal difference of the team which scored the key-goal, expressed as the difference between goals scored and goals conceded during the competition up until the observed game. This variable is categorical, assuming a value of one, when goal difference is positive and zero when goal difference is negative or null.
number of scoreline variations	This expresses the number of scoreline variations until the key-goal is scored. It is a categorical variable, assuming a value of one when the scoreline changes only once, and zero when the scoreline changes twice or more times.
event type	A categorical variable assuming a value of one if the key-goal leads to a win for a team, two when the key-goal leads to a draw, and zero (censored), if the game ends scoreless.

In this paper, the R statistical programming language (Version 3.1.2) and the survival (Version 2.37-7) and *cmprsk package* (Version 2.2–6) were used for all statistical analysis (R code for estimating the cumulative incidence function and the cause-specific hazard models).

5.4. Results

Descriptive statistics for study sample games are presented in Table 2.

Table 2. Frequencies of states and *key-goals* scored in different game contexts

States		N (%)
State	State 0	21 (7)
	State 1	230 (75)
	State 2	55 (18)
Variables		
Game location		
	Home team	157 (51)
	Away team	128 (42)
Accumulated goal difference		
	Negative or null	149 (49)
	Positive	136 (44)
Number of scoreline variations		
	One	177 (58)
	Two or more	108 (35)

The percentage of key-goals scored during each 15-min interval by home and away teams, as well as the game outcome are presented in Figure 2. Home and away teams scored the highest percentage of key-goals associated with wins in the first (0-15 min) and last (75-90 min) game periods. The proportion of key-goals scored which led to a draw gradually increased in each half as the game progressed; the highest proportion of key-goals scored by both home and away teams was registered in the last period (75-90) of the game.

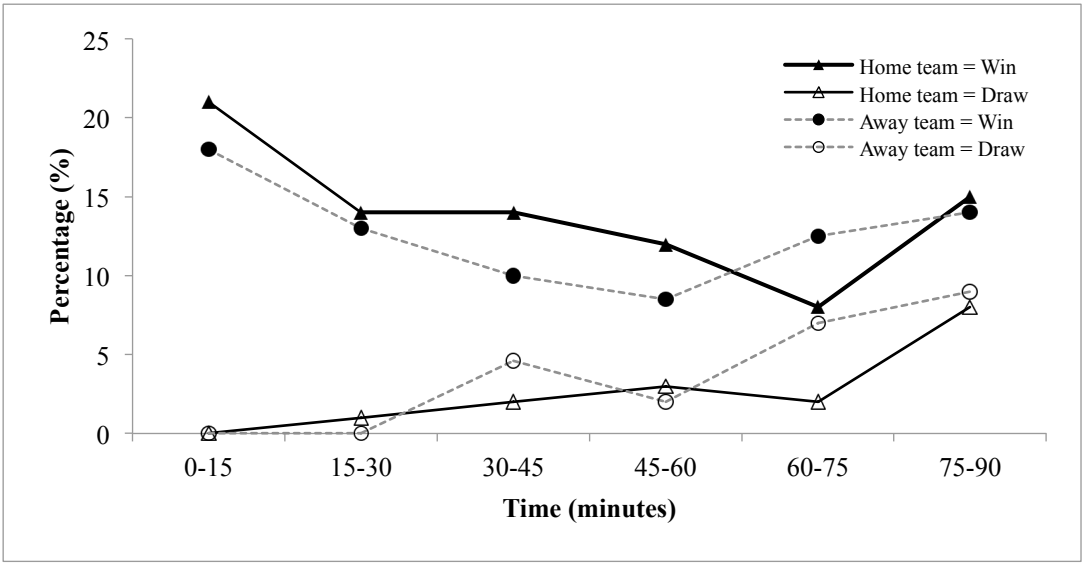


Figure 2. Percentage of key-goals scored by home and away teams and the respective game outcome in each of the six 15-min intervals of the games played in season 2015–2016

Table 3 shows the estimated marginal probability of each outcome (1=key-goal scored by the winning team, 2=key-goal that led to a draw) at the 20th , 40th , 60th and 80th minute. For example, the estimated marginal probability of scoring a key-goal being scored which gives a team a scoreline advantage at the 40th minute is 24.8%, and the estimated marginal probability of a key-goal being scored which produces a draw scoreline for both teams at the 40th minute is 16.6%. This data can be plotted on a graph to provide an overall picture of the cumulative incidence of each type of goal during the course of the game, which is given in Figure 3. Cumulative incidence curves describe the proportion of the total number of games in which a key-goal was scored at time *t*.

Table 3. Probability to score the key-goal during the game and the associated game outcome

Event type	Time (minutes)			
	20	**40**	**60**	**80**
Key-goal = Win	0.137	0.248	0.349	0.411
Key-goal = Draw	0.101	0.166	0.238	0.349

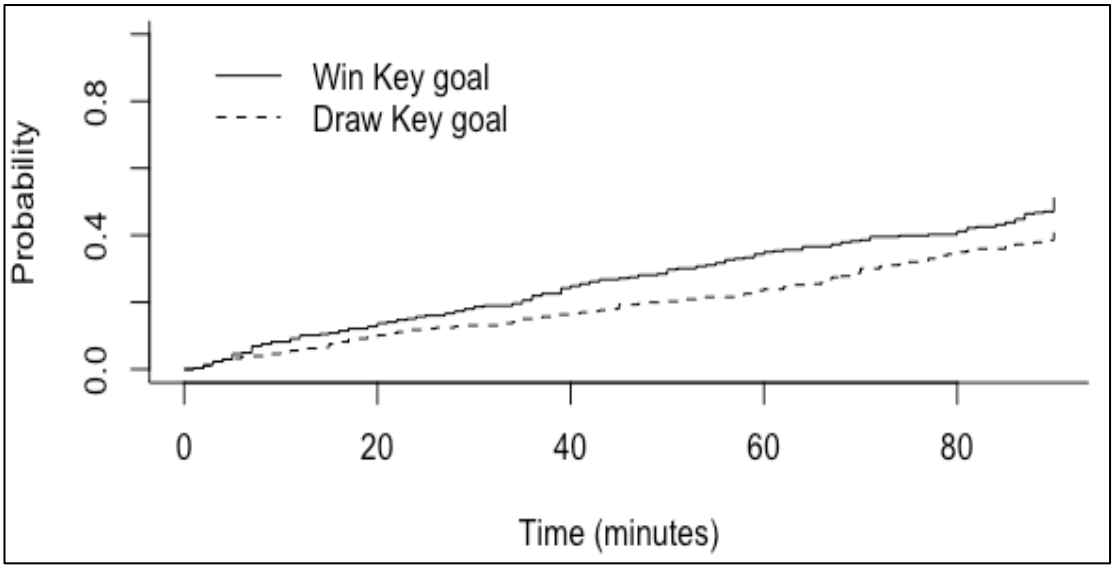


Figure 3. Cumulative incidence curves for key-goals scoring and the associated game outcome

In order to analyse the effect of covariates for each transition of state which was performed, a cause-specific hazard ratio (HR) was estimated. The results of the competing risks analysis are shown in Table 4.

The results demonstrated that game location and accumulated goal difference during the competition up until the time of the game considered had a significant effect on the probability of moving from the starting state (State 0 = 0-0) to a win state (State

1). Teams playing at home had a higher probability of moving to State 1 (that is, scoring the key-goal that gives them a win), as compared with the baseline group (away teams) with HRs equal to 1.32 (0.92–1.52). Teams with a positive accumulated goal difference experience a reducing effect, since the cause-specific hazard ratio for games involving teams with a positive accumulated goal difference is equal to 0.39 (0.51–1.06).

Game location and accumulated goal difference also had a significant effect on the transition to State 2, that is, when the game ended in a draw. Home teams and those with a positive accumulated goal difference presented a higher probability of moving to State 2 with HRs: 1.65 (1.41–1.93) and 2.43 (1.64–3.58), respectively.

The HR presented in transition for both states revealed no association between scoreline evolution throughout the game or probability of *key-goal* occurrences.

Table 4. Competing risks model: effect of covariates on each cause of key-goal scored

	State 0 to State 1 Game ended with victory			State 0 to State 2 Game ended tied		
	HR	**0.95CI**	**P-value**	**HR**	**0.95CI**	**P-value**
Game location						
Away	1			1		
Home	1.32	0.92 – 1.52	0.048*	1.65	1.41 – 1.93	<0.001**
Accumulated goal difference						
Negative or null	1			1		
Positive	0.39	0.51 – 1.06	<0.001**	2.43	1.64 – 3.58	<0.001**
Scoreline						
One	1			1		
Two or more	0.97	0.80 -1.03	0.151	1.01	0.89- 1.15	0.824

Statistically significant: **P<0.001*P<0.05; HR: Hazard Ratio; CI: Confidence Interval

5.5. Discussion

The present study examined the key-goal scoring characteristics in games played in the Portuguese Premier League. Figure 2 shows that the number of winning key-goals was higher in the first half of games, as compared with the second half. The highest proportion of winning key-goals was registered in the first (0-15min) and final (76-90 min) 15-minute periods of games for home and away teams. Drawing key-goals showed an upward trend as time progressed, with a higher proportion of goals being scored in the second half, particularly in the last 15-minute period (75-90 min) for both teams. It seems that in the first and final periods the critical events (key-goals) of the game occur most frequently.

These results are in accordance with many studies conducted in football that have

pointed out the influence on the final score of a goal being scored in the final minutes (Van Ours & Van Tuijl, 2011), an early goal being scored, or a team being the first to score (García-Rubio, Gómez, Lago-Peñas & Ibáñez, 2015; Lago-Peñas, Gómez-Ruano, Megías-Navarro & Pollard, 2016; Molinuevo & Bermejo, 2012; Olsen & Larsen, 1997). According to sources in the literature, the causes for more key-goals being scored in the final minutes could be attributed to physiological factors (Bangsboo, 1994), mental factors (Smith, Zeuwts, Lenoir, Hens, De Jong & Coutts, 2016), and tactical or strategic factors (Carling, Le Gall, McCall, Nedelec, & Dupont, 2015). In his autobiography, coach Alex Ferguson (2013) highlighted the last 15 minutes of a game as being a critical moment for taking risks. Thus, the tactical risks that coaches are willing to take and the execution of an all-out offensive contribute to a higher proportion of goals being scored in the last 15 minutes of games, making this a decisive period of the game (Njororai, 2014).

The estimated cumulative incidence function presented in Figure 3 shows an increase in cumulative key-goal scoring rates over time, with a higher probability of the winning key-goal being scored in all moments of the game, as compared with a drawing key-goal. The competitive risk model developed allowed for the identification of the effect of covariates on the probability of each type of event and the results support the notion that key-goal scoring is influenced by game location and the accumulated goal difference of a team during the competition up until the game considered, which express the team's scoring and defensive capability. The data confirm that the home advantage represents an important factor influencing the probability of key-goals being scored in football. Home teams presented a probability of scoring the winning key-goal which was 32% higher than that of away teams. The probability of home teams drawing the game by scoring the key-goal was 65% higher than that of away teams. These findings are in agreement with those of several studies of home advantage in football (Pollard & Gómez, 2014; Pollard & Pollard, 2005). The effect of home advantage was reported for teams scoring the first goal of game (Tenga, 2012), and those winning the game after scoring the first goal (García-Rubio, Gómez, Lago-Peñas & Ibáñez, 2015; Lago-Peñas, Gómez-Ruano, Megías-Navarro & Pollard, 2016; Molinuevo & Bermejo, 2012). Similar findings were also reported by Van Ours & Van Tuijl (2011) for the effect of last-minute goals, where it was found that the probability of a team scoring a goal in the final minutes and winning a game was almost 20% higher for home teams, while the

probability of home teams losing a game after their opponents had scored a goal in the last minute was about 15% lower than that of away teams.

According to sources in the literature, home advantage in football is influenced by many factors such as: players' familiarity with the venue, fans' support, referee bias, and travel distance favouring the home team either directly or indirectly (Armatas et al., 2009; Lago-Peñas & Lago-Ballesteros, 2011; Molinuevo & Bermejo, 2012; Armatas & Pollard, 2014).

As might be expected, the results also highlighted the effect of accumulated goal difference on key-goal scoring. Teams with a higher scoring and defensive capability, which also had a positive accumulated goal difference, took greater risks in order to change the scoreline. This suggests that losing teams play more attacking football in order to create more scoring opportunities. A positive relationship between a positive accumulated goal difference and game outcome would be expected; however, our findings showed that a positive accumulated goal difference had a negative effect on winning key-goals. Perhaps one of the reasons for this derives from the imprecise categorisation of the goal difference variable, which included only two classes (positive difference and negative difference). Further research should consider more detailed categorisation for this variable in order to improve the analysis of its effect on the key-goal scoring.

Accumulated goal difference is a factor that may affect team performance and our findings indicate that teams with better scoring and defensive capabilities demonstrated more stable patterns of play, independently of the evolving scoreline. Accumulated goal difference is an interesting indicator, which can be used to predict future success, as it reflects the quality of the team and its capability for scoring and avoiding conceding goals.

With regard to the limitations of this study, the following aspects should be noted. Firstly, additional situational variables and performance indicators which may influence goal scoring in football should be included in the analysis. Secondly, assessments of opponent interactions (Tenga, Holme, Ronglan & Bahr, 2010) are of critical importance for the evaluation of the key-goal scoring patterns and improving the validity of analysis outcomes (Carling et al., 2015). Thirdly, models in which transitions to intermediate states are possible and events may occur during the game more than once (for example, to evaluate different types of goals scored) should be estimated.

Despite these limitations, in comparison with previous studies on scoring prediction in football, the present study adopts a novel approach, not only focusing on the outcome itself, but also analysing the time which elapses to the occurrence of a crucial event which led to this outcome. Further research should consider including time-varying covariates (that is, explanatory variables whose values change over time, such as ball possession), in analysis. There are a number of limitations in attempting to predict future success based on past performance. Competitive risk models are often referred to as dynamic and may provide be a useful option for predicting events based on current performances.

5.6. Conclusions

This paper highlights the effect of game location and accumulated goal difference on key-goal scoring in games played in the Portuguese Professional Football League. Our findings suggest that survival modelling can provide a useful tool for performance analysis, enabling game analysts and coaches to predict team performances. We encourage researchers to take full advantage of the range of statistical methods for the analysis of survival data that have been developed in this paper. Investigators need to be aware of the presence of competing risks and their potential effect on statistical analysis.

The model presented covers the simplest competing risk situations, with two competing events. It could be extended to cover an arbitrary number of competing events and could even be applied to multi-state models, in which transitions to intermediate states are possible and events may also occur during the game more than once (for example, taking into account all goals scored up until key-goal occurrence). These models allow for the description of scoreline progression in a more accurate way and thus could provide additional insights for understanding the diversity of types of goals scored during the game and also provide information about which games may be more likely to experience more repeated events (for example, goals that establish a lead, goals that tie the game, and so on) than others.

5.7. References

Armatas, V., & Pollard, R. (2014). Home advantage in Greek football. *European Journal of Sport Science, 14*(2), 116-122.

Armatas, V., Yiannakos, A., Papadopoulou, S., & Skoufas, D. (2009). Evaluation of goals scored in top ranking soccer games: Greek "Superleague" 2006-07. *Serbian Journal of Sports Sciences, 3*(1-4), 39-43.

Bar-Eli, M., Taoz, E., Levy-Kolker, N., & Tenenbaum, G. (1992). Performance quality and behavioral violations as crisis indicators in competition. *International Journal of Sport Psychology, 3*, 325-342.

Burke, K.L., Edwards, T., Weigand, D., & Weinberg, R. (1997). Momentum in sport: a real or illusionary phenomenon. *International Journal of Sport Psychology, 28*(1), 79-96.

Carling, C., Le Gall, F., McCall, A., Nedelec, M., & Dupont, G. (2015). Squad management, injury and game performance in a professional soccer team over a championship-winning season. *European Journal of Sport Science, 15*(7), 573-582.

De Wreede, C., Fiocco, M., & Putter, H. (2011). *mstate*: an R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software*, *38*(7).

Dixon, M.J., & Coles, S.G. (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *J R Stat Soc: Series C (Applied Statistics), 46*(2): 265-280.

Ferguson, A. (2013). *Alex Ferguson: My Autobiography*. London: Hodder and toughton, p. 402.

Ferreira, A. (2013). From game momentum to criticality of game situations. In T. McGary, P. O'Donoghue & J. Sampaio (Ed.). *Routledge Handbook of Sports Performance Analysis* (270–282). New York, USA: Routledge.

García-Rubio, J., Gómez, M.Á., Lago-Peñas, C., & Ibáñez, J. S.(2015). Effect of game venue, scoring first and quality of opposition on game outcome in the UEFA Champions League. *International Journal of Performance Analysis in Sport, 15*(2): 527-39.

Groll, A., Schauberger, G., & Tutz, G. (2015). Prediction of major international soccer tournaments based on team-specific regularized Poisson regression: an application to the FIFA World Cup 2014. *Journal of Quantitative Analysis in Sports, 11*(2), 97-115.

Jones, M. & Harwood, C. (2008). Psychological *momentum* within competitive soccer: Player's perspectives. *Journal of Applied Sport Psychology, 20*(1), 57-72.

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson

models. *The Statistician 52*, 381–393.

Karlis, D., & Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics, 20*(2), 133-145.

Lago, C. (2009) The influence of game location, quality of opposition, and game status on possession strategies in professional association football. Journal of Sports Sciences, *27*, 1463-1469.

Lago-Peñas, C., & Dellal, A. (2010). Ball Possession Strategies in Elite Soccer According to the Evolution of the Game-Score: the Influence of Situational Variables, *Journal of Human Kinetics, 25*, 93-100.

Lago-Peñas, 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-Peñas, C., Gómez-Ruano, M., Megías-Navarro, D., & Pollard, R. (2016). Home advantage in football: Examining the effect of scoring first on game outcome in the five major European leagues. *International Journal of Performance Analysis in Sport, 16*(2), 411-42.

Leite, W.S.S. (2013). Analysis of goals in soccer World Cups and the determination of the critical phase of the game. *Physical Education and Sport, 11*(3), 247-253.

Michailidis, Y., Michailidis, C., & Primpa, E. (2013). Analysis of goals scored in European Championship 2012. *Journal of Human Sport and Exercise, 8*(2): 367-375.

Molinuevo, J. S., & Bermejo, J. P. (2012). The effect of scoring first and home advantage in professional Spanish football and indoor soccer leagues. *Revista De Psicologia Del Deporte, 21*(2), 301-308.

Njororai, W. (2004). Analysis of the goals scored at the 17th World Cup Soccer Tournament in South Korea- Japan 2002. *African Journal for Physical Activity and Health Sciences, 10*(4), 326-332.

Njororai, W. (2014). Timing of Goals Scored in Selected European and South American Soccer Leagues, FIFA and UEFA Tournaments and the Critical Phases of a Game. *International Journal of Sports Science, 4*(6A), 56-64.

Njororai, W. (2014). Timing of Goals Scored in Selected European and South American Soccer Leagues, FIFA and UEFA Tournaments and the Critical Phases of a

Game. *International Journal of Sports Science, 4*(6A), 56-64.

Olsen, E., & Larsen, O. (1997) 'Use of game analysis by coaches', in T. Reilly, J. Bangsbo and M. Hughes (eds), *Science and Football III* (pp. 209–20). London: E. and F.N. Spon.

Pollard, R., & Pollard, G. (2005). Home advantage in soccer: A review of its existence and causes. *International Journal of Soccer and Science Journal, 3*, 25-33.

Pollard, R., & Gómez, M. A. (2014). Comparison of home advantage in men's and women's football leagues in Europe. *European Journal of Sport Science*, *14*:sup1, S77- S83.

Pratas, J. M., Volossovitch, A., & Carita, A. I. (2016). The effect of performance indicators on the time the first goal is scored in football games. *International Journal of Performance Analysis in Sport, 16*(1), 347-354.

Tenga, A. (2012). Chapter: First goal and home advantage at different levels of play in professional soccer (pp. 47-51). *World Congress of Performance Analysis of Sport IX*, Publisher: London & New York: Routledge Taylor & Francis Group, Editors: D. Peters, P. G. O'Donoghue.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010). Effect of playing tactics on goal scoring in Norwegian professional soccer. *Journal of Sports Sciences*, *28*(3), 237-244

Van Ours, J. C., & Van Tuijl, M. A. (2011). Country-specific goal scoring in the "dying seconds". *International Journal of Sport Finance*, *6*, 138-154.

Volossovitch, A. (2008). Dynamic analysis of handball game. Study of the factors that influence the probability of scoring (Unpublished doctoral thesis). Faculdade de Motricidade Humana, Universidade Técnica de Lisboa, Lisboa.

CHAPTER 6

GENERAL DISCUSSION

6. General Discussion

The aim of this thesis was to analyse time patterns of goal scoring in football using different explanatory variables associated with team performance and match context. In order to accomplish this, a systematic review of relevant studies on goal scoring was conducted, and three correlational studies were carried out using survival methods to analyse the first and second goal of the match, as well as the goal which defined match outcome. In this final chapter, an overview of the main findings of all the studies which were carried out are presented, their limitations indicated, and suggestions made for further research.

6.1. Overview of main findings

In **Chapter 2** there was a review of research on football match analysis, which examined goal scoring patterns in accordance with two different approaches: the static and the dynamic. In this systematic review, the advantages and limitations of each approach, as well as their respective contributions towards providing an understanding of the factors associated with goal scoring in elite male football matches played in domestic leagues are summarised. It demonstrated that the static approach (based on game-related statistics and outcomes of player and team actions) was the approach most widely used in football match analysis. However, recently the authors of some studies have demonstrated a growing interest in dynamic analysis, focusing on the emergent and co-adaptive behaviours of interacting players and teams, using positional data to analyse space-time coordination at the player and team levels (Duarte, Araújo, Correia & Davids, 2012; Memmert, Lemmink, & Sampaio, 2016; Pratas, Volossovitch & Ferreira, 2012; Fonseca, Milho, Travassos, & Araújo, 2012). The dynamic approach was also used in the analysis of scoring dynamics, that is, the analysis of score evolution and the prediction of goal scoring during the match depending on factors associated with team performance and match context.

The static approach is more suitable for result-description purposes and enables the identification of the principal factors producing success in football, but it is limited when there is a need to understand the process which led to the final outcome. The dynamic approach addresses the time dimension and provides for the interpretation of performance taking into account the chronological and sequential order in which different player actions or game events occur. While dynamic analysis can be regarded

as a more detailed method for analysing goal scoring performance, this does not necessarily mean that it should automatically be preferred to the static approach in every situation where a choice emerges; rather, the two approaches - the static and the dynamic - should be seen as complementary.

A quasi-dynamic perspective for analysing the influence of different factors on goal scoring in football was used in **Chapter 3**. In this study, the performance indicators that influence the time at which the first goal is scored in football was investigated. Our results demonstrated that home teams with a greater difference between the total number of goals scored and goals conceded during the competition and those with a higher number of shots on goal presented a higher probability of scoring the first goal in the match, while making substitutions and receiving disciplinary sanctions had the effect of delaying the time at which the first goal was scored. The duration of ball possession by home teams had no significant effect on the time at which the first goal was scored. These findings demonstrate that game venue and a high level of offensive effectiveness are the determining factors that influence the time at which the first goal is scored in a match. Our results suggest that these two factors are more important than the style of play employed by teams (either ball possession or direct play) (Collet, 2013). Some studies have shown that an aggressive defensive style of play based on putting constant pressure on opponents who have possession of the ball, can be advantgeous for the teams and may influence goal scoring positively, but only if teams manage to apply this pressure without players being carded (Dobson & Goddard, 2010; Anders & Rotthoff, 2011). For instance, Bar-Eli, Tenenbaum & Geister, (2006) reported that the chances of a team which is sanctioned (for example, by the dismissal of a player) scoring a goal or winning the match were substantially reduced following the sanction. Our findings also demonstrated the negative effect of sanctions on the first goal being scored. Home teams which incurred disciplinary sanctions presented a decreased probability (61.4%) of scoring the first goal of the match.

An interesting finding concerned substitutions. Home teams that made substitutions, as compared with those which made no substitutions, presented a decreased probability (69.7%) of scoring the first goal of the match. This suggests that in the case of a 0-0 draw a tactical intervention like a substitution does not necessarily lead to an increase in scoring opportunities. According to some previous studies, substitution strategies depend on specific match context (Bradley, Lago & Rey, 2014;

Del Corral, Barros, Prieto, 2008; Gomez, Lago & Owen, 2016; Rey, Lago-Ballesteros & Padron, 2015). The majority of substitutions occur during the last 30 minutes of matches (Del Corral et al., 2008; Gomez et al., 2016). Coaches tend to make substitutions earlier than this when their teams are losing or drawing, and when they are playing at home and/or against weaker opponents (Del Corral et al., 2008; Gomez et al., 2016). The results of our study suggest that coaches should not place too much expectation in the value of substitutions when their teams find it difficult to score the first goal of the match. It seems that changing tactics before the first goal of the match is scored is not the most productive strategy for home teams. A substitution made in this context may demonstrate that the team has been forced to change its tactics because it was not able to impose its style on the game and gain and dominate its opponents. Further research is needed to provide a better understanding of the effect of substitutions on the team-scoring performance, and such analysis should take into account the time and type of substitution (offensive or defensive), scoreline, and the quality of the opposition. Del Corral et al. (2008) demonstrated that the quality of the team that made the substitution was not a significant factor, while the quality of the opposing team had a significant positive effect on the time of substitution. When playing against better opponents coaches made the substitutions later. Thus, it would be useful to analyse how the quality of opposition is related to the time of substitution and the time at which the first goal is scored in a match.

To sum up, the variables that positively influence the time at which the first goal is scored in the match can be considered as key-performance indicators, which should be taken into account by coaches in drawing up a strategy for the following match.

In **Chapter 4** the dynamic analysis of scoring patterns was conducted on the second goal scored in a match. The association between the type of second goal scored in a match (that is, the goal that increases the leading team's advantage, or the goal that restores parity to the scoreline) and match venue, as well as final match outcome were analysed. The influence of the time at which the first goal was scored and that of the current scoreline on the probability of a second goal being scored in a match was also examined. Our findings confirmed the significant influence of match context (that is, match venue, scoreline, and time at which the first goal was scored) on the probability of a second goal being scored in a match and also the significant influence of the second goal on the match outcome. The highest number of second goals (both score-doubling

and equalising) scored in a match occurred in the last 5-minute period of each half. Home teams presented a higher percentage (58%) of second goals that doubled score advantage (1-0 to 2-0, or 0-1 to 0-2) as compared with away teams (44%). At the same time, away teams scored a higher percentage (56%) of goals, which restored the scoreline to parity (1-0 to 1-1, or 0-1 to 1-1) in comparison with home teams (42%). Overall, the team which scored the second goal won the match in 59% of cases and a score-doubling goal was strongly positively associated with a win outcome, and equalising goals with a draw and loss outcome. In addition, first goal occurrence in later game periods steadily increases the probability of the second goal in a match being scored by home teams. As might be expected, the results of this study highlighted the decisive impact of the second goal of the match on match outcome. The findings also revealed the significant relationship between the time of the first and second goals scored in the match. Several studies have clearly demonstrated that teams modified their tactics in response to scoreline changes. For example, Lago-Ballesteros, Lago-Peñas & Rey (2012) reported that teams often use a more defensive strategy when winning than when losing, and vice versa. When a team was drawing or winning, the probability of reaching their opponents' goal area decreased by 43% and 53%, respectively. It was also reported that scoreline influenced different performance indicators, such as ball possession and distance covered by players (Lago, 2009; Lago, Casais, Dominguez & Sampaio, 2010). When losing, teams showed a tendency to increase ball possession in midpitch offensive zones (Barreira, Garganta & Anguera, 2011). When winning, teams frequently performed defence/attack transition behaviours in order to get closer to their opponents' goal (Barreira, Garganta & Anguera, 2011). When drawing, teams tended to vary their offensive methods with a view to scoring a goal (Barreira, Garganta, Guimarães, Machado & Anguera, 2014). The most frequent position from which shooting occurred was the central zone. Winning teams showed different profiles from drawing and losing teams in terms of recovering the ball in their own half (the central area) and using penetrative passes into the opponents penalty box (Gomez, Gomez-Lopez, Lago, & Sampaio, 2012).

Thus, coaches should be aware of the effect of the timing of the first goal on the probability of the second goal of a match being scored. This knowledge may help them to adapt their decisions regarding substitutions and tactical options during the match in

accordance with the state of the game, characterised by the combination of the current scoreline and time remaining (Hibbs & O´Donoghue, 2013; Heuer & Rubner, 2012).

For the correlational study described in **Chapter 5,** the time at which a key-goal (a goal which decided the match outcome) was scored, was modelled as a function of contextual variables, such as match venue and scoreline. The results of this study suggest that key-goals, which can be regarded as critical events in a match, occur more frequently in the first (0-15 min) and the final (75-90 min) periods. Our findings showed that a higher proportion of winning key-goals scored by home and away teams was recorded in the first (0-15 min) and final (76-90 min) periods. Drawing key-goals presented an upward trend as time progressed, with the highest number of key-goals scored in the last period for both teams. A winning key-goal had a higher probability of being scored at any moments in the match, as compared with a drawing key-goal. The probability of a key-goal being scored in a match was analysed in accordance with game venue and the indicator of offensive and defensive team efficacy, which was evaluated by using the accumulated goal difference during the competition up until the match considered. The results showed that the probability of home teams scoring the winning key-goal increased by 32%, while for teams with a positive accumulated goal difference it decreased by 61%. For teams playing at home and those with a positive accumulated goal difference the probability of scoring the drawing key-goal increased by 65% and 43%, respectively.

Overall, the findings of this study confirmed the importance of home advantage for scoring a winning or drawing key-goal. This is in line with those of Pollard & Gómez (2014), who reported that home advantage in the major Portuguese domestic league had a probability of 58.71% (from the 2006/2007 to 2011/2012 seasons) of producing a win for a team, although the decreasing trend for home advantage has been observed during recent years in Portuguese and other European leagues (Pollard, 2006; Page & Page, 2007; Pollard & Gómez, 2009).

Our study also showed that the critical events (key-goals) of a game occur most frequently in the first (0-15 min) and the final (75-90 min) 15-min game periods. In football, each goal scored may be regarded as a critical incident that influences the course of the game. In addition to technique and fitness, tactics and strategies can play a fundamental role in producing a higher key-goal scoring frequency in the initial and final stages of a match. This phenomenon could be explained in the following way:

when a match starts, both teams play in accordance with pre-established tactics and strategies. If no goal is scored or an early goal is scored, it is not expected that either team will make substantial changes to their tactics because there is still enough time to score a goal. As the game progresses, the need to score a goal increases and hence tactics become more adventurous (Abt, Dickson & Mummery, 2002).

Accumulated goal difference (the number of goals scored minus the number of goals conceded at the time of the match) is an indicator rarely discussed in football. In the short term it can be a vital tie-breaker, but in the long term it is one of the most accessible indicators of the long-term sustainability of a team's results. The relationship between goal difference and points is transparent and intuitive: better teams win more points, score more goals and concede fewer. Our findings show that the rate at which a key-goal is scored depends on the accumulated goal difference up until the occasion of the match in question.

6.2. Methodological contributions

Current research trends in football match analysis reveal a growing interest in probabilistic modelling of game performance focusing on the prediction of amount of time elapsing to multiple endpoints, the number of events in a given time period, updating predictions as events take place, and so on. Therefore, while time is gaining increasing prominence in football performance analysis, time modelling poses some methodological challenges that were addressed in this study.

Survival analysis methods were used in this study to analyse changes in performance over time as they enable the temporal component of the dynamic process under investigation to be described and are highly recommended when the dominant component of the data is temporal (Castilla, 2007).

The semi-parametric Cox proportional model with fixed and time varying covariates (Cox, 1972), used in Chapter 3 and 4 was shown to be a convenient tool for estimating the time at which the first and second goals are scored. Chapter 5 contains an example of regression analysis using competing risks models. It was considered that key-goal may occur in two game scenários, when the team scores a goal that breaks tie creating a score advantage, or a goal that that ties a game.

These two scenarios were treated differently from each other, allowing for the comparison of the hazard function for a key-goal being scored for a range of competing risks (scoreline states).

6.3. Limitations

There are certain limitations to the present study, which provide important pointers for further research.

1. A larger sample is required for goal-scoring analysis in order to provide meaningful results and more robust predictive models.
2. The public source data (archive data: official football statistics) used in the study provided a limited set of explanatory variables, which does not fully cover the factors influencing goal-scoring in football.
3. The interactive effects of situational variables were not considered.
4. The modelling process for goal scoring did not take into account some important events that can affect goal-scoring probability, for example, player dismissal, early goal scoring, and changes to the tactical system employed in a match.

6.4. Suggestions for future research

During this research program some issues were raised that might constitute interesting topics for future goal-scoring research in football, such as:

1. The need for analysing events which may occur during the match more than once (for example, taking into consideration all goals scored up until the end of the match, goals that give one of the teams the lead, goals that break ties, and goals that increase or decrease the advantage for one of the teams). Modelling recurrent events (using a shared frailty model) may help to identify team and player behaviours associated with each game state, characterised by the scoreline and time remaining.
2. The need for expanding the range of explanatory variables of goal scoring, including those which can only be determined during the course of the game.
3. The need for using multi-state models for event prediction in which transitions to intermediate states are possible (for example, scoreline evolution during the match).

4. The need for complementing game analysis associated with the scoring of goals with new variables, especially those associated with the defensive process (Tenga, Holme, Ronglan & Bahr, 2010a,b) in order to improve the validity of analysis (Carling, Le Gall, McCall, Nedelec & Dupont, 2015).
5. The need for extending the analysis of goal-scoring patterns to other European leagues, women's competitions and lower-level leagues.
6. There is a need for examining performance in football using different scales of analysis. In addition to the study of score evolution (macro analysis), it is necessary to perform the analysis of the spatial-temporal coordination between players and teams (micro analysis) based on positional data collected in situations that lead to goal scoring (Gudmundsson & Horton, 2017).
7. There are a number of limitations in attempting to predict future success based on past performance, especially if such analysis fails to take into account current form, the difficulty of fixtures and possible injuries to key players (Wright, Carling & Collins, 2014), thus future work should take these aspects into account and attempts should be made to predict successful performance based on current performance.

6.5. Conclusions

This research program enabled the identification of how some of performance indicators influence the time at which goals are scored in different game scenarios in high-level football.

1. The first aim of the thesis was to identify the performance indicators (the difference between the total number of goals scored and conceded during the competition up until the match considered, total ball possession time, shots on goal, set plays, disciplinary sanctions and substitutions) that influence the time at which the first goal is scored. The study produced the following findings:

1.1 The team which scored the first goal in a match won 70 % of matches. Home teams scored first in 57.5 % of matches and won 75% of matches, while away teams that scored the first goal won 62% of matches.

1.2 No significant differences were observed between home and away teams with regard to the mean time at which a first goal was scored.

1.3 Increasing the difference between the total number of goals scored and conceded during the competition by one goal enhanced the probability of a team scoring the first goal of a match by 1.1%.

1.4 Every shot on goal by home teams increased the probability of its scoring the first goal by 26.7%.

1.5 Home teams that incurred disciplinary sanctions or made substitutions presented a probability of scoring the first goal of the match which decreased by 38.6% and 30.3%, respectively.

2. With regard to the second aim of the thesis, the association between the type of second goal scored in a match (that is, a goal that increases a team's advantage, or a goal that restores parity to the scoreline), match venue and final match outcome were analysed. The influence of the time at which the first goal was scored and current scoreline on the probability of a second goal being scored in a match was also examined. The study produced the following findings:

2.1 The probability of a second goal being scored in the second half was significantly higher than such a goal being scored in the first half of matches (58% and 42%, respectively). The highest probability of a second goal being scored occurred in the last 5-min period of each half.

2.2 A significant association between match venue and type of second goal scored was found. Home teams presented a greater percentage (58%) of second goals that doubled score advantage (1-0 to 2-0, or 0-1 to 0-2). Away teams scored a higher percentage (56%) of goals, which restored scoreline parity (1-0 to 1-1, or 0-1 to 1-1) in comparison with home teams.

2.3 When both home teams and away teams scored a score-doubling goal, they won 93% and 92% of matches, respectively. A score-doubling goal in a match was positively associated with a win outcome; draw and loss outcomes were significantly positively associated with an equalising goal in a match scored by both home and away teams.

2.4 First-goal occurrence in later game periods steadily increased the probability of the second goal in a match being scored.

3. To address the third aim of the thesis, a predictive model of the time at which key-goal scoring occurred as a function of contextual variables (game location, scoreline, and accumulated goal difference up until the match took place) was estimated. The study produced following findings:

3.1. Key-goals occurred more frequently in the first (0-15 min) and final (75-90 min) periods of a match.

3.2. It was found that there was a higher probability of the winning key-goal being scored throughout the game, as compared with a drawing key-goal being scored.

3.3. A home advantage effect associated with key-goals being scored was observed. Home teams presented a probability of scoring a winning key-goal which was 32% greater than that of away teams. The probability of scoring the drawing key-goal for home teams was 65% higher than that of away teams.

3.4. Teams with a positive accumulated goal difference had a lower probability (39%) of scoring a winning key-goal and a higher probability (more than two times higher) of scoring a drawing key-goal.

6.6. Practical applications

The competitive contexts associated with goal scoring which were identified in this study could be relevant for team management in the context of a particular game. In football analytics, the way a team plays dictates which performance indicators are likely to be significant. The challenge is to find out how the relevance of different performance indicators changes in accordance with match context. Thus, in this study an attempt was made to explain how game scoreline evolves in different match situations. Such information may be useful for the tactical decision-making by coaches, either during the game or in preparation for a match.

6.7. References

Abt, G., Dickson, G. & Mummery,W.K. (2002). *Goal Scoring Patterns Over the Course of a Match: An Analysis of the Australian National Soccer League*, in: Sprinks, T., Reilly, T. and Murphy, A., eds., Science and Football IV. Routledge, London and New York, 106-111.

Anders, A., & Rotthoff K. W. (2011). Yellow Cards: Do They Matter? *Journal of Quantitative Analysis in Sports, 7*(1).

Bar-Eli, M., Tenenbaum, G., & Geister, S. (2006). Consequences of Player's Dismissal in Professional Soccer: A Crisis-Related Analysis of Group-Size Effects. *Journal of Sports Sciences*, 24(10), 1083-1094.

Barreira, D., Garganta,. J., & Anguera,. T. (2011) *In search of nexus between attacking game-patterns, match status and type of ball recovery in European Soccer Championship 2008*, in Research Methods and Performance Analysis, M. Hughes, et al., Editors, University of West Hungary, pp. 226-237.

Barreira, D., Garganta, J., Guimarães, P., Machado, J., & Anguera, M.T. (2014). Ball recovery patterns as a performance indicator in elite soccer. *Journal of Sports Engineering and Technology, 228*(1), 61-72.

Bradley, P.S., Lago-Peñas, C., & Rey, E. (2014). Evaluation of the Match Performances of Substitution Players in Elite Soccer. International Journal of Sports Physiology and Performance, 9(3), 415-424.

Carling, C., Le Gall, F., McCall, A., Nedelec, M., & Dupont, G. (2015). Squad management, injury and game performance in a professional soccer team over a championship-winning season. *European Journal of Sport Science, 15*(7), 573-582.

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.

Del Corral, J., Barros, C. P., & Prieto-Rodríguez, J. (2008). The Determinants of Soccer Player Substitutions: A Survival Analysis of the Spanish Soccer League. *Journal of Sports Economics, 9*(2), 160-172.

Dobson, S., & Goddard, J. (2010). Optimizing strategic behaviour in a dynamic setting in professional team sports. *European Journal of Operational Research, 205*(3), 661-669.

Duarte, R., Araujo, D., Correia, V., & Davids, K. (2012). Sports teams as superorganisms: implications of sociobiological models of behaviour for research and practice in team sports performance analysis. *Sports Medicine, 42*(8), 633–642.

Fonseca, S., Milho, J., Travassos, B., & Araújo, D. (2012). Spatial dynamics of team

sports exposed by Voronoi diagrams. *Human Movement Science, 31*, 1652-1659.

Gomez, M-A., Lago-Peñas, C., & Owen, A.L. (2016). The influence of substitutions on elite soccer teams' performance. *International Journal of Performance Analysis in Sport., 16*(2), 553-68.

Gómez, M.A., 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), 393-8.

Heuer, A., & Rubner, O. (2012). How Does the Past of a Soccer Match Influence Its Future? Concepts and Statistical Analysis. *PLoS ONE, 7*.

Hibbs, A., & O'Donoghue, P. (2013). Strategy and tactics in sports performance. In T. McGarry, P. O'Donoghue, & J. Sampaio (Eds.), *Handbook of sport performance analysis*. London, UK: Routledge. pp. 239-249.

Hughes, M., Caudrelier, T., James, N., Redwood-Brown, A., Donnelly, I., Kirkbride, A. & Duschesne, C. (2012). Moneyball and Soccer - an analysis of the key performance indicators of elite male soccer players by position. *Journal of Human Sports & Exercise, 7*(2),402-412.

Lago, C., The influence of match location, quality of opposition, and match status on possession strategies in professional association football. Journal of Sports Sciences, 2009. 27(13): 1463-1469.

Lago, C., Casais, L., Dominguez, E., & Sampaio, J. (2010). The effects of situational variables on distance covered at various speeds in elite soccer. *European Journal of Sport Science, 10*(2): 103-109.

Lago-Ballesteros, J., Lago-Peñas, C., & Rey, E. (2012). The effect of playing tactics and situational variables on achieving score-box possessions in a professional soccer team. *Journal of Science and Medicine in Sport*, *30*(14), 37–41.

Memmert, D., Lemmink, A.P.M.K., & Sampaio, J.(2016). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine, 47*(1), 1-10.

Pollard, R. (2006). Worldwide regional variations in home advantage in association football. *Journal of Sports Sciences, 24*(3), 231-240.

Page, L., & Page, K. (2007). The second leg home advantage: Evidence from European football club competitions. *Journal of Sports Sciences, 25*(14), 1547-1556.

Pollard, R., & Gómez, M. Á. (2009). Home advantage in football in South-West

Europe: Long-term trends and regional variation. *European Journal of Sport Science, 9*(6), 341-352.

Pollard, R., & Gómez, M.A. (2014). Comparison of home advantage in men's and women's football leagues in Europe. *European Journal of Sport Science, 14*(sup1):S77-S83.

Pratas, J., Volossovitch, A., & Ferreira, A.P. (2012). The effect of situational variables on teams' performance in offensive sequences ending in a shot on goal. A case study. *The Open Sports Sciences Journal, 5*, 193-199.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010a). Effect of playing tactics on goal scoring in Norwegian professional soccer. *Journal of Sports Sciences, 28*(3), 237-244.

Tenga, A., Holme, I., Ronglan, L. T., & Bahr, R. (2010b). Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science, 10*(4).

Rey, E., Lago-Ballesteros, J., & Padrón-Cabo, A. (2015). Timing and tactical analysis of player substitutions in the UEFA Champions League. *International Journal of Performance Analysis in Sport, 15*(3), 840-50.

Wright, C., Carling, C., & Collinsz, D. (2014). The wider context of performance analysis and it application in the football coaching process. *International of Performance Analysis in Sport, 14*,709-733.

APPENDIX

R Programming

```
#########################################################################
# Paper 1.
#  The effect of performance indicators on the time the first goal is scored in football matches
#########################################################################
library(survival)

firstgoal <- read.delim(pipe('pbpaste'))
summary(firstgoal)

# Define variables
time <- timefirstgoal
event <- event
x <- cbind(goaldifference, ballpossession, shotsongoal, setplays, sanctions, substitutions )

# Kaplan-Meier non-parametric analysis
kmsurvival <- survfit(Surv(timefirstgoal,event) ~ 1)
summary(kmsurvival)
plot(kmsurvival, xlab="timefirtsgoal" , ylab="Survival Probability")

#########################################################
# Cox regression model, univariate and multivariate
#########################################################

CoxModel.1 <- coxph(Surv(start,end,event) ~ x, method="breslow", data=firstgoal)

#########################################################################
# Paper 2.
# Analysis of the second goal scored in Portuguese football matches: 2009-2015
#########################################################################
library(survival)

secondgoal <- read.delim(pipe('pbpaste'))
summary(secondgoal)

# Define variables
time <- timesecondgoal
event <- event
x <- cbind(timefirstgoal, scoreline)

# Kaplan-Meier non-parametric analysis
kmsurvival <- survfit(Surv(timefirstgoal,event) ~ 1)
summary(kmsurvival)
plot(kmsurvival, xlab="timefirtsgoal" , ylab="Survival Probability")

#########################################################
# Cox regression model
#########################################################

CoxModel.1 <- coxph(Surv(start,end,event) ~ x, method="breslow", data=firstgoal)
summary(coxModel.1)


#########################################################
# Figure 1: Plot survival function for first and second goal scoring
#########################################################

ggsurv <- function(s, CI = 'def', plot.cens = T, surv.col = 'gg.def',
```

```
                cens.col = 'red', lty.est = 1, lty.ci = 2,
                cens.shape = 3, back.white = F, xlab = 'Time',
                ylab = 'Survival', main = ''){

library(ggplot2)
strata <- ifelse(is.null(s$strata) ==T, 1, length(s$strata))
stopifnot(length(surv.col) == 1 | length(surv.col) == strata)
stopifnot(length(lty.est) == 1 | length(lty.est) == strata)

ggsurv.s <- function(s, CI = 'def', plot.cens = T, surv.col = 'gg.def',
                     cens.col = 'red', lty.est = 1, lty.ci = 2,
                     cens.shape = 3, back.white = F, xlab = 'Time',
                     ylab = 'Survival', main = ''){

  dat <- data.frame(time = c(0, s$time),
                    surv = c(1, s$surv),
                    up = c(1, s$upper),
                    low = c(1, s$lower),
                    cens = c(0, s$n.censor))
  dat.cens <- subset(dat, cens != 0)

  col <- ifelse(surv.col == 'gg.def', 'black', surv.col)

  pl <- ggplot(dat, aes(x = time, y = surv)) +
    xlab(xlab) + ylab(ylab) + ggtitle(main) +
    geom_step(col = col, lty = lty.est)

  pl <- if(CI == T | CI == 'def') {
    pl + geom_step(aes(y = up), color = col, lty = lty.ci) +
      geom_step(aes(y = low), color = col, lty = lty.ci)
  } else (pl)

  pl <- if(plot.cens == T & length(dat.cens) > 0){
    pl + geom_point(data = dat.cens, aes(y = surv), shape = cens.shape,
                    col = cens.col)
  } else if (plot.cens == T & length(dat.cens) == 0){
    stop ('There are no censored observations')
  } else(pl)

  pl <- if(back.white == T) {pl + theme_bw()
  } else (pl)
  pl
}

ggsurv.m <- function(s, CI = 'def', plot.cens = T, surv.col = 'gg.def',
                     cens.col = 'red', lty.est = 1, lty.ci = 2,
                     cens.shape = 3, back.white = F, xlab = 'Time',
                     ylab = 'Survival', main = '') {
  n <- s$strata

  groups <- factor(unlist(strsplit(names
                                   (s$strata), '='))[seq(2, 2*strata, by = 2)])
  gr.name <-  unlist(strsplit(names(s$strata), '='))[1]
  gr.df <- vector('list', strata)
  ind <- vector('list', strata)
  n.ind <- c(0,n); n.ind <- cumsum(n.ind)
  for(i in 1:strata) ind[[i]] <- (n.ind[i]+1):n.ind[i+1]

  for(i in 1:strata){
```

```
  gr.df[[i]] <- data.frame(
    time = c(0, s$time[ ind[[i]] ]),
    surv = c(1, s$surv[ ind[[i]] ]),
    up = c(1, s$upper[ ind[[i]] ]),
    low = c(1, s$lower[ ind[[i]] ]),
    cens = c(0, s$n.censor[ ind[[i]] ]),
    group = rep(groups[i], n[i] + 1))
 }

 dat <- do.call(rbind, gr.df)
 dat.cens <- subset(dat, cens != 0)

 pl <- ggplot(dat, aes(x = time, y = surv, group = group)) +
  xlab(xlab) + ylab(ylab) + ggtitle(main) +
  geom_step(aes(col = group, lty = group))

 col <- if(length(surv.col == 1)){
  scale_colour_manual(name = gr.name, values = rep(surv.col, strata))
 } else{
  scale_colour_manual(name = gr.name, values = surv.col)
 }

 pl <- if(surv.col[1] != 'gg.def'){
  pl + col
 } else {pl + scale_colour_discrete(name = gr.name)}

 line <- if(length(lty.est) == 1){
  scale_linetype_manual(name = gr.name, values = rep(lty.est, strata))
 } else {scale_linetype_manual(name = gr.name, values = lty.est)}

 pl <- pl + line

 pl <- if(CI == T) {
  if(length(surv.col) > 1 && length(lty.est) > 1){
   stop('Either surv.col or lty.est should be of length 1 in order
       to plot 95% CI with multiple strata')
  }else if((length(surv.col) > 1 | surv.col == 'gg.def')[1]){
   pl + geom_step(aes(y = up, color = group), lty = lty.ci) +
    geom_step(aes(y = low, color = group), lty = lty.ci)
  } else{pl +  geom_step(aes(y = up, lty = group), col = surv.col) +
        geom_step(aes(y = low,lty = group), col = surv.col)}
 } else {pl}


 pl <- if(plot.cens == T & length(dat.cens) > 0){
  pl + geom_point(data = dat.cens, aes(y = surv), shape = cens.shape,
            col = cens.col)
 } else if (plot.cens == T & length(dat.cens) == 0){
  stop ('There are no censored observations')
 } else(pl)

 pl <- if(back.white == T) {pl + theme_bw()
 } else (pl)
 pl
}
pl <- if(strata == 1) {ggsurv.s(s, CI , plot.cens, surv.col ,
                        cens.col, lty.est, lty.ci,
                        cens.shape, back.white, xlab,
                        ylab, main)
} else {ggsurv.m(s, CI, plot.cens, surv.col ,
```

```
            cens.col, lty.est, lty.ci,
            cens.shape, back.white, xlab,
            ylab, main)}
  pl
}

#########################################################################
# Paper 3.
# Predicting key-goal scoring in football, based on performance indicators and contextual factors
#########################################################################
library(survival)
library(cmprsk)
library(mstate)

msrprep(keygoal)

keygoal <- read.delim(pipe('pbpaste'))
summary(keygoal)

# Define variables
time <- time
event <- eventtype
x <- cbind(scoreline, goaldifference, gamelocation)

keygoal<- data.frame(read.table("keygoal.txt",header=T))

# time denotes the survival time: time to occurrence of # the first event.

# event.type is the event type indicator:
# 1: Win Key Goal.
# 2: Draw Key goal.
# 0: Censored observation: matches ended goal less at end of follow-up.

keygoal$wintransition <- ifelse(keygoal$eventtype==1,1,0)
# Create variable denoting occurrence of win key goal.

keygoal$drawtransition <- ifelse(keygoal$eventype==2,1,0)
# Create variable denoting occurrence of draw key goal.

keygoal$event <- ifelse(keygoal$event.type > 0,1,0)
# Create event indicator for any type of event.

attach(keygoal)

#########################################################
#Prepare transition matrix with 3 stages
#########################################################
tmat <- trans.comprisk(2, names=c("0", "1","2"))
tmat # look at transition matrix

#create WKG variable
keygoal$stat1 <- as.numeric(keygoal$eventtype == 1)

#create DKG variable
keygoal$stat2 <- as.numeric(keygoal$eventtype == 2)

#prepare the long data using msprep
keygoal <- msprep (time= c(NA, "time"), status= c(NA,"stat1","stat2"), data= keygoal, keep =
"matchlocation", trans = tmat)
```

```
keygoallong <- msprep (data = keygoal, trans = tmat, time= c(), eventtype=c(NA,"stat1", "stat2"),

events(keygoal)

#########################################################
# Figure 1: Plot cumulative incidence functions for Win Key Goal and Draw Key Goal in the combined sample.
#########################################################
CI.overall <- cuminc(ftime = keygoal$time, fstatus = keygoal$eventtype, cencode=0)
CI.overall

plot(CI.overall, curvlab = c("Win Key goal", "Draw Key goal"), xlab = "Minutes")

postscript("CIF.figure.ps",horizontal=T,paper="letter")

cif1 <- cuminc(ftime = keygoal$time, fstatus = keygoal$eventtype,cencode=0)

plot(cif1, fun="eventtype", bty="l", conf.int=FALSE, col=1, lty=1, lwd=1, ylim=c(0, 1),
    xlab="time", ylab="Cumulative incidence", mark.time=TRUE)

plot(cif1$"1 1"$time,cif1$"1 1"$est,type="l",
    ylim=c(0,0.7),
    xlab="Time (minutes)",ylab="Probability",
    lty=1,col="red")
title("Figure 1. Cumulative Incidence functions")
lines(cif1$"1 2"$Time,cif1$"1 2"$est,type="1",
     lty=1,col="blue")
legend("topleft",
legend = c("Win Key Goal (WKG)","Draw Key Goal (DKG)","All cause (1-KM)/Sum of two CIFs"),
      lty = c(1,1,1),
      col = c("red","blue","black"),
      bty="n")
km.composite <- survfit(Surv(time,eventtype) ~ 1)
lines(km.composite$Time,1- km.composite$surv,type="l",lty=1,col="black")

#########################################################
# Cause-specific hazard models.
#########################################################

# Cause-specific hazard for Win Key Goal
cox.1 <- coxph(Surv(time,eventtype=1) ~  Scoreline.variation + Goal.difference + Game.location; data =
keygoal, method="breslow")

# Cause-specific hazard for Draw Key Goal
cox.2 <- coxph(Surv(survival.time,eventype=2) ~ Scoreline.variation + Goal.difference + Game.location;
data = keygoal, method="breslow")

summary(cox.1)
summary(cox.2)
```