The right man in the right place? Substitutions and goal-scoring in soccer

Reference:
Amez Simon, Neyt Brecht, Van Nuffel Frederik, Baert Stijn.- The right man in the right place? Substitutions and goal-scoring in soccer
Psychology of sport and exercise - ISSN 1878-5476 - 54(2021), 101898
Full text (Publisher's DOI): https://doi.org/10.1016/J.PSYCHSPORT.2021.101898
To cite this reference: https://hdl.handle.net/10067/1774090151162165141

The right man in the right place? Substitutions and goal-scoring in soccer

Abstract

Objectives. We investigate how the goal-scoring probability in international club soccer evolves after player substitutions.

Design. To this end, we analyse rich data concerning 2,025 recent soccer games played in the two most prestigious club soccer competitions, i.e. the UEFA Champions League and the UEFA Europa League.

Method. As first in the literature, we control for within-game dynamics by applying a minute-by-minute bivariate probit approach.

Results/Conclusions. We find that teams experience increased goal-scoring probabilities after their first and second substitution and a decreased probability of scoring after the three substitutions made by their opponent. This association is less distinct during the first three minutes after the substitution, which is consistent with difficulties to adapt to (i) the game intensity by the substitute player or (ii) tactical changes by the entire team. Furthermore, we find that the change in the goal-scoring probability is substantially bigger if the team is losing at the moment of the substitution.

Keywords: soccer, substitution, goal-scoring, bivariate probit model, UEFA Champions League, UEFA Europa League

1 Introduction

On May 26th 2018, the Spanish football club Real Madrid won the UEFA Champions League for the 7th time by beating its English opponent Liverpool FC (UEFA, 2018b). Although, this final will always be remembered for the missteps of Liverpool's goalkeeper Loris Karius causing two goals of Real Madrid, it can be argued that it was Madrid's coach—Zinédine Zidane—who had a decisive impact on the outcome of this final game. With only 30 minutes to play and heading to a tied game, he brought Gareth Bale into the field. Only two minutes later, Real Madrid was leading the game after a bicycle kick of the Welshman. Additionally, following a misstep of the opposing goalkeeper, substitute player Bale made his second goal of the game, which ensured the win for Real Madrid.

For soccer coaches like Zinédine Zidane, making a substitution is the most direct way of influencing the game, once the game has started. According to Law 3 of the FIFA regulation (IFAB, 2019) a substitution means that a player is brought into the field to replace another player of his team. Once a player has been substituted, he cannot re-enter the field. Limited to only three substitutions (IFAB, 2019), soccer coaches are forced to allocate them very carefully (Myers, 2012). In peer-reviewed scientific literature, three main motivations for making a substitution are proposed: (i) to replace an injured or warned player, (ii) to replace tired players, or (iii) to apply tactical changes (Hirotsu & Wright, 2002; Ascari & Gagnepein, 2006; Del Corral, Barros & Prieto-Rodriguez, 2008; Bradley, Lago-Peñas & Rey, 2014; Rey, Lago-Ballesteros & Padrón-Cabo, 2015).

Surprisingly, the empirical research to date investigating the association between substitutions and the later outcomes of a soccer game is rather limited. To our knowledge, four studies have examined this association. First, Gomez, Lago-Peñas, and Owen (2016) analysed the substitutions made during 50 randomly selected games played in the Spanish first division during the 2014/2015 season. For all observed substitutions, they examined how the final outcome for the team changes after a substitution, conditional on its timing. They found that a first substitution of a losing team made between the 56th and the 70th minute was associated with an improved probability of ending the game with a draw. Additionally, a second substitution for these losing teams in the 70th to 75th minute of the game was associated with a higher chance of levelling the score. Substitutions made in other time intervals were not significantly associated with a better outcome. Second, Rey et al. (2015) analysed all 677 substitutions made in the 2013/2014 UEFA Champions League and concluded that substitutions of losing teams were overall associated with higher chances on an improved game outcome. Using a decision tree approach, they found that losing teams improved their outcome in 43% of the cases when the coach made his first substitution before the 53rd minute. A second substitution prior to the 71st minute was associated with a draw or a win in 34% of the cases. For drawing games and for the team leading the game no significant timings were found to be associated with the final outcome of the game. Third, instead of focusing on the final outcome of the games in terms of winning, drawing or losing, Myers (2012) investigated the association between substitutions and later changes in the goal difference. A decision tree approach based on 485 substitutions made in the English Premier League, the Italian Serie A, La Liga in Spain, MLS in the USA, and the 2010 World Cup, showed similar results as Rey et al. (2015). Again, no significant associations were found for winning teams or in a tied game. For losing teams, by contrast, substitutions were associated with an improved goal difference if they were made prior to the 58th (1st substitution), 73rd (2nd substitution), and 79th (3rd substitution) minute. Myers (2012) consequently argued that soccer coaches should interpret these findings as a decision rule. However, fourth, Silva and Swartz (2016) raised some crucial concerns with respect to the practical implementation of this decision rule because Myers (2012) use the current match status at a critical moment to provide guidelines on how coaches should have made substitutions prior to that time. To test their refined definition of the substitution guidelines, they applied a Bayesian logistic regression approach on an augmented version[1] of the dataset used by Myers (2012). Silva and Swartz (2016) found no specific timing of a substitution during the second half of the game to be more beneficial compared to other moments. Consequently, they concluded that coaches should substitute whenever they notice tired players. In summary, the current empirical literature does not provide consistent evidence on the association between substitutions and later game outcomes in soccer.

This lack of consensus in the literature to date might be because the aforementioned studies did not take into account that substitutions are endogenous decisions. That is, they correlate

[1] Silva and Swartz (2016) captured all games considered in Myers (2012) except the 2010 World Cup matches with incomplete information. In addition, the dataset was expanded with the games played in the English Premier League during the seasons 2010/2011, 2011/2012, and 2012/2013.

with other within-game dynamics that may also affect the (later) game outcomes, such as red cards and penalty kicks. As a consequence, when not controlling for within-game dynamics, the measured association may reflect the effect of those dynamics rather than the substitution itself. Therefore, in the present study, we apply a bivariate probit regression approach in which the goal-scoring probability of home and away teams is explained minute-by-minute by earlier substitutions while controlling for other within-game dynamics, examples of which are summed up above. This minute-by-minute bivariate probit regression approach has been applied to investigate related research questions in soccer economics such as: (i) home advantage (Buraimo, Forrest, & Simmons, 2010) and (ii) the impact of additional referees (Albanese, Baert, & Verstraeten, 2020). This statistical framework is applied to the most extended dataset in this literature to date (*infra*, Section 2). By means of our analyses, we answer the following research question in the context of European international clubs soccer.

- **Research question 1 (R1).** Are a soccer team's substitutions associated with later higher goal-scoring probabilities after controlling for important other game dynamics?

By making a substitution, the coach might change the formation of the team (Hirotsu & Wright, 2006). Therefore, one could argue that the team might need some time to implement the new tactical formation on the field. Even when the team's formation does not change after the substitution, i.e. the substitute player plays the exact position as the substituted player, it may take some time for the substitute player to adapt to his role within the team and the intensity of the game.[2] Therefore, it may be hypothesised that the change in the goal-scoring probability is smaller immediately following the substitution. We investigate this hypothesis by answering the following research question in particular.

- **Research question 2 (R2).** Is the association between a soccer team's substitutions and its later goal-scoring probabilities smaller during the three minutes following the substitution?

As coaches can have different motivations to make a substitution (*supra*), it may be that the change in goal-scoring probability after a substitution is heterogeneous by the underlying

[2] An alternative explanation might be that the flow of the game is interrupted for a short period of time due to the substitution.

motivation of the substitution. First, in line with coaches' decision to substitute and given that players typically get more tired later on in the game, we could hypothesise that substituting later in the game may change the goal-scoring probability differently as suggested by previous studies (see e.g. Myers, 2012; Rey et al., 2015). Second, a substitution might alter the tactical formation of the team (Hirotsu & Wright, 2002). One could argue that an offensive orientated substitution may increase the team's goal-scoring probability. Simultaneously, the more offensive formation can increase the goal-scoring probability of the opponent. Finally, Schneemann and Deutscher (2016) suggest that substitute players reduce their effort when the team is losing the game. As such, we might expect that the change in goal-scoring probability of the substituting team is less favourable when losing at the moment of the substitution. In summary, we answer the following research questions.

- Research question 3a (R3a). Is the change of the goal-scoring probability after a substitution heterogeneous by the timing of the substitution?
- Research question 3b (R3b). Is the change of the goal-scoring probability after a substitution heterogeneous by the tactical direction of the substitution?
- Research question 3c (R3c). Is the change of the goal-scoring probability after a substitution heterogeneous by the match status at the moment of the substitution?

Next to the aforementioned methodological contribution of the current study, this study is—to the best of our knowledge—the first in the psychological literature to investigate the impact of substitutions on team performance in professional sports. Thus, by investigating the consequences of coaches' substitutions decisions, it contributes to the extensive literature on judgement and decision making in (professional) sports (Raab, Bar-Eli, Plessner, & Araújo, 2019). Moreover, we contribute to the broad and interdisciplinary literature on determinants of game outcomes in professional soccer such as (i) shirt colour (Attril, Gresty, Hill, & Barton, 2008; Krenn, 2014), (ii) experiencing a player dismissal (Bar-Eli, Tenenbaum, & Geister, 2006; Mechtel, Bäker, Brändle, & Vetter, 2011), (iii) playing at the home venue (Carron, Loughead, & Bray, 2005; Courneya & Carron, 1992; Van Damme & Baert, 2019), and (iv) referee bias (Dohmen, 2008; Nevill, Balmer, & Williams, 2002).

2 Data

To answer our research questions, we constructed a dataset including all 2,025 games played in the UEFA Champions League between 2008 and 2016 and the UEFA Europa League between 2011 and 2016.[3] Both competitions are organised by the UEFA to let the best European soccer clubs compete. Based on the past performance of their teams in these two competitions, every national football association receives a number of entry tickets for the tournaments. The highest ranked team(s) receive an entry ticket for the (preliminaries of the) UEFA Champions League. Next, the entry tickets for the (preliminaries of the) UEFA Europa League are received by the teams ranked next at the end of the previous season.

Each season, 32 teams enter the first round of the UEFA Champions League, also known as the group stage. All teams are divided into eight groups of four teams who meet their opponents twice, once at home and once on the opponent's field. After these group games, the first and second ranked[4] team of each group proceed to the first knock-out phase where group winners of the first round meet a runner-up of another group. This knock-out phase consists of both a home and an away game. Only the aggregated winner[5] proceeds to the next round. This procedure is repeated until two teams remain. Next, those two teams meet each other for one final game, which is played on a neutral[6] field. As a result, 125 games are played each season in the UEFA Champions League.

The UEFA Europa League is organised similarly but allows for more participating teams. The group stage is entered by 48 teams, which are divided into 12 different groups of four teams. In analogy with the UEFA Champions League, the first and second ranked team in the group stage proceed to the first knock-out round. These 24 qualified teams are accompanied by the eight teams that were ranked third in their group in the UEFA Champions League group stage for the first knock-out phase of the tournament. Next, the knock-out procedure of the UEFA

[3] Before 2011, another competition format was used for the UEFA Europa League.

[4] Points are allocated to each team based on the result of every game. A team is granted three points for every win, one point for a draw, and no points for a loss.

[5] In case of an aggregated tie, the away goals rule is applied. This rule states that the team that scores more away goals qualifies for the next stage. If this procedure does not result in a winner, extra time is played and potentially penalty kicks may be decisive (UEFA, 2018a).

[6] Prior to the start of the tournament, the UEFA decides where the final game will be hosted. Thus, in principle, the final game is played on a neutral field, unless the team playing in the hosting city reaches the final round.

Champions League is followed. Due to the higher number of participating teams, each UEFA Europa League tournament contains 205 games.

In total, our dataset consists of 2,025 games, of which 1,000 in the UEFA Champions League and 1,025 in the UEFA Europa League. We deleted 293 group stage games which had no substantial competitive value because at least one of the teams was mathematically unable to change its qualification status for the next round (Baert & Amez, 2018).[7] Additionally, because in our estimations (*infra*) we want to control for crowd size effects, we deleted 19 games for which such data was unavailable. Therefore, our final sample size consists of 1,713 games. Multiple variables (*infra*) were collected for these games. The main sources were the game reports available at the official website of UEFA (http://www.uefa.com). Summary statistics for the variables used in our analyses are presented in Table 1. Panel A presents the dependent variables. Next, variables with respect to the substitutions are presented in Panel B. Finally, Panel C presents control variables for both within-game and between-game dynamics.

< Table 1 about here >

We used a pair of dependent variables, i.e. a goal scored in a certain minute by the home and away teams, respectively. Both binary variables take the value 1 if a goal was scored by the observed team in the specific minute. As can be seen in Panel A from Table 1, home teams scored a goal in 1.70% of the observed minutes, which corresponds to one goal every 59 minutes, on average. Away teams scored less often, namely once every 77 minutes.

Next, Panel B of Table 1 presents our independent variables of interest. Binary variables are used to introduce information with respect to (prior) substitutions. The binary variable takes the value 1 if the home (away) team made a substitution prior to the observed minute. For example, when the home team made its first substitution in the 60th minute, this variable takes the value 1 at minute 60 and all following minutes until the end of the game. In order to distinguish the order of substitutions, a binary variable was constructed for every possible substitution. Hence, six different binary variables with respect to the substitution scheme were derived from the game reports. As can be seen in Table 1, both home and away teams seem to allocate substitutions—on average—similarly with respect to timing. In 35.44%

[7] Analyses based on the full sample including games without competitive value yield very similar results which are available upon reasonable request.

(34.98%) of the observed minutes, a home (away) team already used its first substitution.

To answer R2, we constructed additional variables that take the value 1 when the first, second, or third substitution of the home (away) team was made in the three previous minutes. Next, a variable was constructed that takes the value of the minute in which the substitution was made. More concretely, when the home team made its first substitution in the 75th minute, this variable has the value 75. To answer R3b, the field positions of the participating players were derived from Transfermarkt (http://transfermarkt.de). A substitution was labelled as *offensive* when the player entering the field was playing a position 'higher' in the formation. Stated otherwise, entering an attacker for a defender was labelled offensive while substituting a midfielder by a defender was not. With respect to R3c, a binary variable indicated whether the substituting team was losing at the time of the replacement.

Further, we distinguish two groups of control variables in Panel C of Table 1. On the one hand, we collected data on variables to control for pre-game influences. First, we control for the stage of the tournament—i.e. first round, knock-out first leg, knock-out second leg, or final. First round games made up 69.06% of our dataset. The remaining games in our dataset were played in the knock-out stages, except for the 13 final games. Second, we constructed a binary variable that takes the value 1 if the game was part of the UEFA Europa League tournament and 0 if the game belonged to the UEFA Champions League. Third, because crowd size may influence the effort of both teams (Goumas, 2013; Van Damme & Baert, 2019), we controlled for the attendance by introducing the natural logarithm of the crowd size into our empirical model. As the UEFA game reports did not consistently provide the crowd size, we derived this information from Worldfootball (http://worldfootball.net). Fourth, we controlled for the relative strength of both teams prior to the game. Relative strength is—in line with Baert and Amez (2018)—defined as the home team's UEFA coefficient in the observed season plus 1 divided by its opponent's UEFA coefficient for the observed season plus 1.[8] Additionally, we take into account season fixed effects.

On the other hand, our empirical approach (*infra*) allows us to control for the influence of within-game dynamics. First, we include the variable 'minute' and its square in our analyses

[8] We add 1 to the UEFA coefficient in order to avoid division by 0. Teams may have an UEFA coefficient of 0 if they did not participate in UEFA Champions League or UEFA Europa League during the five previous seasons as the UEFA coefficient is based on a team's results in both tournaments of these five seasons.

as control variables as we expect—in analogy with Buraimo, Simmons, and Maciaszczyk (2012)—that the goal-scoring probability increases (Ridder, Cramer & Hopstaken, 1994) every minute but at a decreasing rate. Second, we add two dummy variables for the 45th minute and the 90th minute because due to added time granted by the referee at the end of the first half and at the end of the game, both these 'minutes' may cover more than just one minute. Third, because missing a clear scoring opportunity like a penalty may lead to inaction inertia (Tykocinski, Pittman & Tuttle, 1995) and thereby change the goal scoring probability, we control for missed penalties prior to the observed minute. Fourth, as Červený, van Ours, and van Tuijl (2018) argued that goal-scoring rates changes after receiving a red card, we also controlled for red cards received earlier in the game.

In contrast to previous studies (e.g. Myers, 2012; Silva & Swartz, 2016) on the association of substitutions and later match outcomes, we did not use the substitution as unit of observation (*supra*). In line with Buraimo et al. (2010), we estimate a minute-by-minute bivariate probit model. As a result, our dataset consists of 154,170 observations on the game-minute level. This approach allowed us to take within-game dynamics into account. For example, an improved goal difference after a substitutions may be caused by a penalty kick for the substituting team or by an exclusion of a player of the opposing team, instead of the impact of the substitution itself. Similarly, a deteriorated goal difference may be the result of the exclusion of a player of the substituting team instead of by the impact of the substitution itself. As pointed out by Hirotsu and Wright (2006), players' and coach's decisions results from the interaction between both teams, so we assumed that goal-scoring opportunities for the home team are associated to the goal-scoring chances of the away team. Therefore, we jointly estimated the goal-scoring opportunities for both teams. Standard errors were clustered on the game level in all models.

3 Results

3.1 Benchmark analysis

Table 2 presents the estimation results of our benchmark model. In this benchmark model, the goal-scoring probabilities of the home and away teams are regressed on six different variables with respect to substitutions as well as on all control variables described in Section 2. We find statistically significant associations between substituting and the goal-scoring probability. More concretely, for the home team, we find a positive association between both its first *(b = 0.047, p = 0.094)* and its second *(b = 0.108, p = 0.004)* substitution and the probability that the home team scores. Making a third substitution as the home team is not significantly associated with a home team's goal-scoring probability. As the average marginal effects[9] in Table 2 show, the home team's goal-scoring probability is 0.18 percentage points (0.41 percentage points) higher every minute following this team's first (second) substitution. In contrast, all substitutions of the away team are negatively associated with the home team's goal scoring probabilities. Following the first, second and third substitution of its opponent, the goal-scoring probability for the home team is 0.56 percentage points (hereafter: *pp*), 0.74 pp, and 0.84 pp lower, respectively.

< Table 2 about here>

Similarly, we find that the first *(b = 0.070, p = 0.026)* and second *(b = 0.079, p = 0.052)* substitution of the away team is positively associated with its own goal-scoring probabilities while all substitutions of its opponent are negatively associated with their goal-scoring probabilities. As for the home team, making a third substitution is not associated with a change in the away team's own goal-scoring probability. Concretely, the away team's first and second substitution are associated with a 0.21pp and 0.23pp higher probability of scoring a goal every minute following this substitution, respectively. The substitutions of the home team are—on average—associated with 0.36pp, 0.35pp, and 0.43pp lower goal-scoring probabilities for the away team after the first, second, and third substitution, respectively.

Before turning to the results of our alternative models, we discuss some secondary results from our benchmark analysis concerning the control variables. First, both home and away teams score less often during the first leg in the knock-out phase. This may intuitively be explained by a cautious approach of this kind of game as both teams know a decisive second match follows. Second, a bigger crowd size is associated with higher goal-scoring probabilities for both teams. Third, in line with our expectations (*supra*, section 2) more goals are scored during the 45th and the 90th minute. Fourth, a higher lead is associated with an increased probability of scoring a goal. This may be explained by relative differences in the strength or flow on the matchday (Heuer & Rubner, 2012). The higher goal-scoring probability of a

[9] Although, we do not claim a causal interpretation of our measures—we return to this issue in Section 4—we apply the label *average marginal effects* for conventional reasons.

leading home team decreases when the goal difference increases. Thus, it seems that leading teams decrease their offensive efforts once a reassuring lead is obtained. Finally, home teams score less often when their opponent missed a penalty kick earlier in the game.

3.2 Heterogeneity analysis

To answer R2, we add six interaction variables between a prior substitution and whether this substitution was made in the three minutes before the observed minute. Table 3 presents the main average marginal effects of this analysis where we investigate whether the change of goal-scoring probabilities applies immediately after a substitution is made. Recall that home teams experience an increase in the goal-scoring probability after their first substitution. However, this increase does not apply immediately after the substitution. The first three minutes after the first substitution of the home team, their probability to score is even lower *(0.24pp minus (−)0.34pp = −0.10pp)* than before the substitution. The second substitution is associated with an immediately increase of the goal-scoring probabilities although this increase is substantially smaller during the first three minutes following the substitution *(0.47pp minus 0.41pp = 0.06pp, p = 0.010)*. All three substitutions of their visiting opponent are associated with a decreased probability of scoring a goal for the home team but this change in goal-scoring probability is smaller immediately after the substitution as well.

< Table 3 about here>

Away teams experience similar changes in their probability of scoring after a substitution as their hosting colleagues. While the first substitution of the away team is associated with an increased *(0.29pp, p = 0.002)* goal-scoring probability, this increase is countered during the first three minutes following the substitution *(−0.44pp, p = 0.014)*. We find a similar result for the third substitution of the away team as well, while the second substitution is not associated with a change in the goal-scoring probability. The lower goal-scoring chances experienced by the away team after the home team's first and second substitutions occur immediately after the substitution. Immediately after the third substitution of the home team, the decrease in the probability of scoring is only limited while this decrease is bigger later on in the game.

< Table 4 about here>

Next, we test the robustness of these findings by looking how the effects discussed above change when we focus on alternative time intervals of one and five minutes, respectively. The main average marginal effects are presented in Table 4. Regardless of the concrete length of the time interval, we find that the change in goal-scoring probability following a substitution is smaller during the very first minute(s) the substitute player is on the field.

In summary, with respect to R2, we conclude that the change in the probability of scoring a goal does not take place immediately after the moment of the substitution. We see two possible explanations. On the one hand, this may indicate that a substitute player needs a short period of time to adapt to the team dynamics and the intensity on the field. On the other hand, it may be possible that the entire team has to adapt to a new tactical plan following the substitution.

To answer our last three research questions with respect to the heterogeneity of the change in the goal-scoring probability after a substitution, we add multiple interaction terms to our benchmark model. Table 5 shows the main estimated average marginal effects of our three models to investigate R3a, R3b, and R3c respectively.

< Table 5 about here>

First, we extend our benchmark model with an interaction variable capturing the minute of the substitution. Since the players physical fitness decreases during the game, the difference in physical energy between the substitute and the substituted player will be higher as the substitution is made later in the game. Therefore, we expect the change in the goal-scoring probability to be higher when the substitution takes place at the end of the game. As shown in Model (1) of Table 5, we only find limited evidence supporting this hypothesis. Only for the second substitution of both the home and away team, we find that the increase in goal-scoring probability after a substitution reduces over time. However, this reduction in the increase of the goal-scoring probability is negligible in magnitude.

In Model (2), we add an interaction term between our substitution variables and an indicator of whether the substitute player plays at a more offensive position than his substituted teammate. An offensive substitution may indicate a tactical change leading to a more offensive oriented formation.[10] Therefore, we expect that an offensive substitution may be

[10] The distinction between offensive and defensive substitutions does not capture actual changes in the team's tactical formation. Tracking team formations at every minute of the game is very difficult and might be considered subjective (Myers, 2012). However, this distinction clearly signals the tactical intensions of the substituting coach at the time of the substitution.

associated with an increased goal-scoring probability compared to a neutral or defensive substitution. In contrast, this more offensive orientation could lead to a higher chance of conceding a goal. Stated otherwise, we expect an increase in the goal-scoring probability of the opponent. However, none of the estimated average marginal effects presented in Table 5 are statistically significant, suggesting that the change in goal-scoring probability does not depend on the tactical orientation of the substitution.

Last, we investigate whether the change in the goal-scoring probability depends on the match status at the moment of the substitution. Therefore, we constructed two binary variables that take the value 1 if the home (away) team is behind at the moment of the substitution. Then, we enter the interaction of this dummy variables with the substitution variables into Model (3). We find strong statistical evidence that making a substitution while losing the game is associated with increased goal-scoring probabilities. On the one hand, this finding does not corroborate with the finding of Schneemann and Deutscher (2016) who find that substitute players reduce their effort when their team is losing. Moreover, our finding contradicts Silva and Swartz (2016) as they find that the goal-scoring probability of a losing team does not change after a substitution. This might be due to the setting of the games included in the sample. Silva and Swartz (2016) focus on soccer games played in robin-round national football competitions In contrast, the current study looks at games played in the context of prestigious international club tournaments. However, our findings are in line with Myers (2012) who analysed a very similar sample as Silva and Swartz (2016). Differences between the latter two studies might be explained by other statistical approaches (*supra*, section 1). On the other hand, the found heterogeneity of the association depending on the current match status is in line with Rey et al. (2015) who stated—based on games played in the UEFA Champions League—that substitutions are only associated with an improved match outcome when the substituting team is losing at the moment of the substitution.[11]

4 Conclusion

With the current study we substantially contributed in three ways to the empirical literature

[11] Rey et al. (2015) analysed the substitutions made in all games played in the UEFA Champions League during the 2013/2014 season. Since all these games are also included in the sample of the current study, the former sample might considered to be a subsample of the latter.

on the association between substitutions and team performance in soccer games. First, we were the first to take into account within-game dynamics by applying a minute-by-minute approach in line with Buraimo et al. (2010). Second, we explored whether the association between substitutions is heterogeneous by the motivation underlying the substitution, namely (i) the timing of the substitution, (ii) the tactical consequences of the substitution, and (iii) the current match status. Third, we test our hypotheses on a large sample covering combined 13 seasons of the two most prestigious international club tournaments in soccer, i.e. the UEFA Champions League and the UEFA Europa League.

In this study, we investigated the association between substitutions and the probability of scoring a goal. To this end, we applied a minute-by-minute bivariate probit approach on a rich dataset containing 2,025 games. We found that teams experience an increased probability of scoring a goal after their first and second substitution and a decreased goal-scoring probability following the three substitutions of their opponent. This changes in the goal-scoring probability are less pronounced during the three minutes following immediately after the substitution. This might be explained by a short period of adjustment needed by (i) the player to the game intensity or (ii) the team to tactical changes. Furthermore, we found evidence for heterogeneity of this change in the goal-scoring probability by the current match status. When the team is losing at the moment om the substitution, the increase in goal-scoring probability is bigger. As such our empirical results tend to the finding of Myers (2012) and Rey et al. (2015). In particular, these findings might be a good starting point for future research investigating the decision-making process of soccer coaches with respect to substitutions in the course of the game. Concretely, linking coaches' underlying intentions at the moment of substitutions with its consequences might substantially improve our understanding of the decision-making process in highly competitive circumstances.

We end this article by acknowledging its main limitation. Although our minute-by-minute approach allowed us to control for a substantial amount of pre- and within-game dynamics, it is possible that the association between substitutions and goal-scoring probabilities is confounded by unobserved factors. Therefore, our empirical results do not allow us to claim that the change in goal-scoring probability is caused by the substitutions. For example, our data does not capture whether the main motivation of the substitution was to replace an injured or warned player. Future research could collect and use information on the coaches' motivation for the substitution in a smaller sample and investigate whether the association between substitutions and goal-scoring probability is heterogeneous by this motivation. Moreover, more advanced statistics that were not available in the current study might capture currently unobserved information. Scholars might aim to enhance the understanding of the effect of substitution on later performance by introducing these statistics into the innovative minute-by-minute approach.

References

Albanese, A., Baert, S., Verstraeten, O. (forthcoming2020). Twelve eyes see more than eight. Referee bias and the introduction of additional assistant referees in soccer. *PLos One, 15*(2), e0227758.

Ascari, G., Gagnepain, P. (2006). Spanish Football. *Journal of Sports Economics*, *7*(1), 76–89.

Attril, M.J., Gresty, K.A., Hill, R.A., Barton, R.A. (2008). Red shirt colour is associated with long-term team success in English football. *Journal of Sports Sciences, 26*(6), 577–582.

Baert, S., Amez, S. (2018). No better moment to score a goal than just before half time? A soccer myth statistically tested. *PLoS ONE, 13*(3), e0194255.

Bar-Eli, M., Tenenbaum, G., 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.

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.

Buraimo, B., Forrest, D., Simmons, R. (2010). The 12th man?: Refereeing bias in English and German soccer. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 173*(2), 431–449.

Buraimo, B., Simmons, R., Maciaszczyk, M. (2012). Favoritism and referee bias in European soccer: Evidence from the Spanish League and the UEFA Champions League. *Contemporary Economic Policy, 30*(3), 329–343.

Carron, A.V., Loughhead, T.M., Bray, S.R. (2005). The home advantage in sport competitions: Courneya and Carron's (1992) conceptual framework a decade later. *Journal of Sports Sciences, 23*(4), 395–407.

Červený, J., van Ours, J.C., van Tuijl, M.A. (2018). Effects of a red card on goal-scoring in World Cup football matches. *Empirical Economics, 55*(2), 883–903.

Courneya, K.S., Carron, A.V. (1992). The home advantage in sport competitions: A literature review. *Journal of Sport & Exercise Psychology, 14*, 13–27.

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.

Dohmen, T.J. (2008). The influence of social forces: Evidence from the behaviour of football referees. *Economic Inquiry, 46*(3), 411–424.

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), 553–568.

Goumas, C. (2013). Home advantage and crowd size in soccer: A worldwide study. *Journal of Sport Behavior, 36*(4), 387–399.

Heuer, A., Rubner, O. (2012). How does the past of a soccer match influence its future? Concepts and statistical analysis. *PLoS ONE, 7*(11), e47678.

Hirotsu, N., Wright, M.B. (2006). Modeling tactical changes of formation in association football as a zero-sum game. *Journal of Quantitative Analysis in Sports, 2*(2), article 4.

Hirotsu, N., Wright, M.B.(2002). Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. *Journal of the Operational Research Society, 53*(1), 88–96.

IFAB (2019). Laws of the Game 2019/20. Zurich: The International Football Association Board.

Krenn, B. (2014). The impact of uniform color on judging tackles in association football. *Psychology of Sport and Exercise, 15*(2), 222–225.

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

Myers, B.R. (2012). A proposed decision rule for the timing of soccer substitutions. *Journal of Quantitative Analysis in Sports, 8*(1), Article 9.

Nevill, A.M., Balmer, N.J., Williams, A.M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise, 3*(4), 261–272.

Raab, M., Bar-Eli, M., Plessner, H., Araújo, D. (2019). The past, present and future of research on judgment and decision making in sport. *Psychology of Sport and Exercise, 42*, 25–32.

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), 840–850.

Ridder, G., Cramer, J.S., Hopstaken, P. (1994). Down to ten: estimating the effect of a red card in soccer. *Journal of the American Statistical Association, 89*,1124–1127.

Schneemann, S., Deutscher, C. (2016). Intermediate information, loss aversion, and effort: Empirical evidence. *Economic Inquiry, 55*(4), 1759–1770.

Silva, R.M., Swartz, T. (2016). Analysis of substitution times in soccer. *Journal of Quantitative Analysis in Sports, 12*(3), 113–122.

Tykocinski, O.E., Pittman, T.S., Tuttle, E.E. (1995). Inaction inertia: Foregoing future benefits as a result of an initial failure to act. *Journal of Personality and Social Psychology, 68*(5), 793–803.

UEFA (2018a). Regulations of the UEFA Champions League 2018-21 Cycle. 2018/19 Season. Nyon: UEFA.

UEFA (2018b). UEFA Champions League – Real Madrid - Liverpool. Retrieved 22 January 2020 from https://www.uefa.com/uefachampionsleague/history/season=2018/matches/round=2000885/match=2021711/index.html?iv=true

Van Damme, N., Baert, S. (2019). Home advantage in European international soccer: which dimension of distance matters? *Economics: The Open-Access, Open-Assessment E-Journal, 13*(2019–50), 1–17.

Table 1. Summary statistics

	Mean	SD
A. Dependent variables		
Goal home	0.017	-
Goal away	0.013	-
B. Independent variables		
First substitution prior home	0.354	-
Second substitution prior home	0.202	-
Third substitution prior home	0.088	-
First substitution prior away	0.350	-
Second substitution prior away	0.201	-
Third substitution prior away	0.088	-
C. Control variables		
First round	0.691	-
Knock-out first leg	0.152	-
Knock-out second leg	0.151	-
Final	0.008	-
Europa League	0.491	-
Log attendance	10.156	0.777
Relative strength	4.549	14.112
Minute	45.500	25.979
Minute 45	0.011	-
Minute 90	0.011	-
Goal difference	0.206	1.199
Missed penalty prior home	0.000	-
Missed penalty prior away	0.000	-
Red card prior home	0.001	-
Red card prior away	0.001	0.038
N (minutes)	154,170	

Notes. A definition of these variables can be found in Section 2.1. No standard deviation (SD) is provided for binary variables.

Table 2. Association between substitutions and goal scoring chances: benchmark analysis

	Goal home team	Goal away team
A. Bivariate probit model estimates		
First substitution prior home	0.047* (0.028)	−0.122*** (0.032)
Second substitution prior home	0.108*** (0.038)	−0.119*** (0.041)
Third substitution prior home	−0.007 (0.044)	−0.145*** (0.048)
First substitution prior away	−0.146*** (0.029)	0.070** (0.032)
Second substitution prior away	−0.194*** (0.037)	0.079* (0.041)
Third substitution prior away	−0.220*** (0.041)	0.050 (0.049)
Knock-out first leg	−0.051** (0.020)	−0.052** (0.021)
Knock-out second leg	0.004 (0.019)	0.000 (0.023)
Final	−0.015 (0.055)	0.043 (0.068)
Europa League	0.030* (0.018)	0.024 (0.020)
Log attendance	0.032*** (0.012)	0.034*** (0.012)
Relative strength	−0.001 (0.001)	−0.002 (0.002)
Relative strength squared	0.000 (0.000)	0.000 (0.000)
Minute	−0.005*** (0.002)	0.001 (0.002)
Minute squared	0.000*** (0.000)	0.000 (0.000)
Minute 45	0.410*** (0.060)	0.529*** (0.061)
Minute 90	0.918*** (0.058)	0.881*** (0.060)
Goal difference	0.458*** (0.027)	−0.350*** (0.020)
Goal difference squared	−0.041*** (0.007)	−0.020*** (0.007)
Missed penalty prior home	0.375 (0289)	−0.080 (0.414)
Missed penalty prior away	−4.748*** (0.107)	0.101 (0.441)
Red card prior home	-0.268 (.0424)	−0.182 (0.292)
Red card prior away	0.005 (0.174)	−0.073 (0.209)
Intercept	−2.614*** (0.130)	−2.772*** (0.136)
Season fixed effects	Yes	Yes
B. Average marginal effects		
First substitution prior home	0.002* (0.001)	−0.004*** (0.001)
Second substitution prior home	0.004*** (0.001)	−0.004*** (0.001)
Third substitution prior home	−0.000 (0.002)	−0.004*** (0.001)
First substitution prior away	−0.006*** (0.001)	0.002** (0.001)
Second substitution prior away	−0.007*** (0.001)	0.002** (0.001)
Third substitution prior away	−0.008*** (0.002)	0.001 (0.001)
Log pseudo-likelihood	−21,145.823	
N (minutes)	154,170	

Notes. The presented statistics are bivariate probit model estimates (panel A) and average marginal effects (panel B). Standard errors, clustered at the game level, are presented in parentheses. A definition of the variables used can be found in Section 2. *** (**) ((*)) indicate significance at the 1% (5%) ((10%)) significance level.

Table 3. Association between substitution and goal scoring chances: short term versus longer term association

	Goal home team	Goal away team
First substitution prior home	0.002** (0.001)	−0.004*** (0.001)
Second substitution prior home	0.005*** (0.002)	−0.004*** (0.001)
Third substitution prior home	−0.000 (0.002)	−0.005*** (0.002)
First substitution prior away	−0.006*** (0.001)	0.003*** (0.001)
Second substitution prior away	−0.009*** (0.002)	0.002 (0.001)
Third substitution prior away	−0.010*** (0.002)	0.003* (0.002)
First substitution prior home × previous three minutes	−0.003* (0.002)	0.001 (0.002)
Second substitution prior home × previous three minutes	−0.004* (0.002)	0.003 (0.002)
Third substitution prior home × previous three minutes	−0.001 (0.002)	0.004** (0.002)
First substitution prior away × previous three minutes	0.000 (0.002)	−0.004** (0.002)
Second substitution prior away × previous three minutes	0.004** (0.002)	−0.001 (0.002)
Third substitution prior away × previous three minutes	0.008*** (0.002)	−0.007*** (0.002)
Control variables	All	All
Season fixed effects	Yes	Yes
Log pseudo-likelihood	−21,120.445	
N (minutes)	154,170	

Notes. The presented statistics are average marginal effects of our bivariate probit model estimates and standard errors, clustered at the game level, in parentheses. A definition of the variables used can be found in Section 2. *** (**) ((*)) indicate significance at the 1% (5%) ((10%)) significance level.

Table 4. Association between substitution and goal scoring chances: short term versus longer term association: Alternative time intervals

	(1) Goal home team	(1) Goal away team	(2) Goal home team	(2) Goal away team
First substitution prior home	0.002** (0.001)	−0.003*** (0.001)	0.002* (0.001)	−0.003*** (0.001)
Second substitution prior home	0.004*** (0.001)	−0.004*** (0.001)	0.004** (0.002)	−0.005*** (0.001)
Third substitution prior home	−0.000 (0.002)	−0.004*** (0.001)	0.001 (0.002)	−0.004** (0.002)
First substitution prior away	−0.006*** (0.001)	0.003*** (0.001)	−0.006*** (0.001)	0.003*** (0.001)
Second substitution prior away	−0.008*** (0.001)	0.002* (0.001)	−0.009*** (0.002)	0.003** (0.001)
Third substitution prior away	−0.009*** (0.002)	0.002 (0.001)	−0.010*** (0.002)	0.003 (0.002)
First substitution prior home × previous minute	−0.008** (0.004)	−0.001 (0.003)	-	-
Second substitution prior home × previous minute	−0.007* (0.004)	0.003 (0.003)	-	-
Third substitution prior home × previous minute	−0.004 (0.004)	−0.001 (0.003)	-	-
First substitution prior away × previous minute	−0.001 (0.003)	−0.008** (0.003)	-	-
Second substitution prior away × previous minute	0.001 (0.003)	−0.004 (0.003)	-	-
Third substitution prior away × previous minute	0.008** (0.004)	−0.007** (0.004)	-	-
First substitution prior home × previous five minutes	-	-	−0.001 (0.002)	−0.000 (0.001)
Second substitution prior home × previous five minutes	-	-	−0.001 (0.002)	0.004*** (0.002)
Third substitution prior home × previous five minutes	-	-	−0.002 (0.002)	0.002 (0.002)
First substitution prior away × previous five minutes	-	-	−0.000 (0.002)	−0.003** (0.001)
Second substitution prior away × previous five minutes	-	-	0.004** (0.002)	−0.003** (0.002)
Third substitution prior away × previous five minutes	-	-	0.007*** (0.002)	−0.006*** (0.002)
Control variables	All	All	All	All
Season fixed effects	Yes	Yes	Yes	Yes
Log pseudo-likelihood	−21,129.887		−21,119.290	
N (minutes)	154,170			

Notes. The presented statistics are average marginal effects of our bivariate probit model estimates and standard errors, clustered at the game level, in parentheses. *** (**) ((*)) indicate significance at the 1% (5%) ((10%)) significance level.

Table 5. Association between substitution and goal scoring chances: other dimensions of heterogeneity in the association

	(1) Goal home team	(1) Goal away team	(2) Goal home team	(2) Goal away team	(3) Goal home team	(3) Goal away team
First substitution prior home	0.000 (0.002)	−0.003 (0.002)	0.001 (0.001)	−0.004*** (0.001)	0.001 (0.001)	0.002* (0.001)
Second substitution prior home	0.030*** (0.008)	−0.024*** (0.007)	0.004** (0.002)	−0.004*** (0.001)	0.001 (0.002)	0.001 (0.001)
Third substitution prior home	0.013 (0.017)	−0.012 (0.015)	0.001 (0.002)	−0.007*** (0.002)	−0.009*** (0.002)	0.001 (0.002)
First substitution prior away	−0.004 (0.003)	0.005** (0.002)	−0.007*** (0.001)	0.001 (0.001)	0.005*** (0.001)	0.000 (0.001)
Second substitution prior away	−0.022*** (0.007)	0.017*** (0.006)	−0.007*** (0.002)	0.003** (0.001)	0.003 (0.002)	−0.002 (0.001)
Third substitution prior away	−0.058*** (0.016)	0.013 (0.018)	−0.012*** (0.002)	0.002 (0.002)	−0.001 (0.003)	−0.009*** (0.002)
First substitution prior home × minute substitution	0.000 (0.000)	−0.000 (0.000)	-	-	-	-
Second substitution prior home × minute substitution	−0.000*** (0.000)	0.000*** (0.000)	-	-	-	-
Third substitution prior home × minute substitution	−0.000 (0.000)	0.000 (0.000)	-	-	-	-
First substitution prior away × minute substitution	−0.000 (0.000)	−0.000** (0.000)	-	-	-	-
Second substitution prior away × minute substitution	0.000** (0.000)	−0.000** (0.000)	-	-	-	-
Third substitution prior away × minute substitution	0.001*** (0.000)	−0.000 (0.000)	-	-	-	-
First substitution prior home × offensive substitution	-	-	−0.000 (0.001)	−0.000 (0.001)	-	-
Second substitution prior home × offensive substitution	-	-	0.000 (0.001)	0.000 (0.001)	-	-
Third substitution prior home × offensive substitution	-	-	−0.000 (0.001)	0.001 (0.001)	-	-
First substitution prior away × offensive substitution	-	-	−0.000 (0.001)	0.001 (0.001)	-	-
Second substitution prior away × offensive substitution	-	-	−0.000 (0.001)	−0.000 (0.001)	-	-
Third substitution prior away × offensive substitution	-	-	0.001 (0.001)	0.001 (0.001)	-	-
First substitution prior home × home behind	-	-	-	-	0.022*** (0.002)	−0.017*** (0.001)
Second substitution prior home × home behind	-	-	-	-	0.015*** (0.003)	−0.009*** (0.002)
Third substitution prior home × home behind	-	-	-	-	0.019*** (0.004)	−0.005** (0.002)
First substitution prior away × away behind	-	-	-	-	−0.023*** (0.001)	0.013*** (0.001)
Second substitution prior away × away behind	-	-	-	-	−0.011*** (0.002)	0.009*** (0.002)
Third substitution prior away × away behind	-	-	-	-	−0.009*** (0.003)	0.015*** (0.002)
Control variables	All	All	All	All	All	All
Season fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudo-likelihood	−15,031.881		−15,076.947		−14,181.745	
N (minutes)	105,390					

Notes. The presented statistics are average marginal effects of our bivariate probit model estimates and standard errors, clustered at the game level, in parentheses. A definition of the variables used can be found in Section 2. *** (**) ((*)) indicate significance at the 1% (5%) ((10%)) significance level.