Appalachian
STATE UNIVERSITY

Department of Economics Working Paper

Number 25-08| October 2025

Analyzing the Influence of Team Formation on Soccer Match Outcomes

O. Ashton Morgan
Appalachian State University

Jayjit Roy
Appalachian State University

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Analyzing the Influence of Team Formation on Soccer Match Outcomes

O. Ashton Morgan (corresponding author)
416 Howard Street
3094 Peacock Hall
Appalachian State University
Boone, NC 28608
Tel: (828) 262 2927
Email: morganoa@appstate.edu

Jayjit Roy
416 Howard Street
3108 Peacock Hall
Appalachian State University
Boone, NC 28608
Tel: (828) 262 2927
Email: royj@appstate.edu

Analyzing the Influence of Team Formation on Soccer Match Outcomes

Abstract

This paper investigates the strategic role of soccer formations on match outcomes. Using six seasons (2016–2022) of English Premier League data we estimate the impact of four dominant formations—4-3-3, 4-4-2, 4-5-1, and wingbacks—on expected and actual goals differential with a three-stage empirical strategy. Results from a fixed effects linear model show that playing a 4-3-3 at home yields the highest expected goal advantage across nearly all opponent formations. Predictive margins, conditional on opponent formation, further show that the wingback system also performs well at home. In contrast, the 4-4-2 formation is the least effective for away teams, likely due to being overloaded in midfield and limited press resistance. These findings align with tactical shifts over time away from a more rigid 4-4-2 formation toward more flexible systems that maximize spatial control and pressing efficiency.

Keywords: Strategic Behavior; Soccer Formations; Expected Goals; English Premier League

Acknowledgements: The paper was presented in an Appalachian State University Sports Economics Seminar in 2025. We thank participants in this session for helpful comments.

Introduction

In soccer, the selection of a team's formation or playing system can be conceptualized as a strategic decision game in a duopoly scenario, where there are two economic agents (teams) and each team's payoff (e.g., goal difference or difference in expected goals) depends not only on its own formation but also on the formation of its opponent. Just like two rival firms, either choosing a level of quantity or price, based on what they perceive the other firm to be doing, two soccer teams play a formation-selection game. This duopoly framework highlights that the formation decision is a dynamic (as opposed to static) and strategic interaction influenced by the opposition's expected set up. For example, if the home team knows that the away team will play a more conservative 4-5-1 system, then should the home team match that system or play a 4-3-3 or a system with wingbacks to maximize expected outcomes?

This paper analyzes six seasons of English Premier League data to investigate the influence of team formation on expected match outcomes. Team formations, or systems, have consistently evolved over time to reflect changes in tactical ideas. In the early years of the sport (19th Century), formations and tactics were basic. It wasn't until 1870 that a goalkeeper became a recognized position and further until 1912 that the keeper was prevented from handling the ball outside of the penalty area (Wilson 2008).[1] At this time, a formation typically consisted of two or three defenders and eight or nine attackers. When the English Football Association changed the offside rule to require two players (down from three) to be between the attacker and the opponent's goal at the time the ball is played, increasing the number of goals scored in a game,

[1] The rule changed in 1912 to stop the Sunderland AFC goalkeeper from carrying the ball to the halfway line (Wilson 2008).

the Arsenal manager at the time (Herbert Chapman) implemented the W-M (3-2-2-3) formation (Wilson 2008). This new system allowed for an improved defensive structure.

In the mid-20th Century, the global expansion the sport led to further developments in formation tactics. Brazil's 1958 World Cup winning team successfully employed a 4-2-4 structure, which later, in the 1960s and 1970s, morphed into the 4-4-2 formation. The 4-4-2 provided more defensive capabilities with its two lines of four (defense and midfield) but still had an attacking threat with two strikers. This formation was the primary system used particularly in English soccer, both at the domestic and international level (Wilson 2008) during this time.

At least part of the reasoning for a dedication to the 4-4-2 formation in English soccer at this time was based on the influential work of Charles Reep, an English trained accountant and the first known analyst in professional soccer. Reep is cited with collecting data on the first soccer match between Swindon Town and Bristol Rovers in 1950, noting pitch positions, passing sequences, and outcomes. In the years that followed, he continued to collect data by hand on soccer matches in England. During this time, his innovative research with soccer data meant that he was hired by Brentford Football Club (FC) in a part-time advisory role before taking up a similar position at Wolverhampton Wanderers FC and Sheffield Wednesday. Reep investigated the relationship between passing movements and goals. His most-cited academic work entitled "Skill and Chance in Association Football" was published in the Journal of the Royal Statistical Society. Having studied data from 578 games, his primary findings include that only 5% and 1% of all moves consist of four or more passes and 6 or more passes, respectively. Further, 50% of goals are scored with one pass or fewer; and 80% of goals are scored with 3 passes or fewer (Reep 1968).

While regarded as an astute mathematician and analyst, Reep is also controversially renowned for setting back soccer in England at the national level for many years. His analysis encouraged managers of several clubs in England to play the 4-4-2 system with a long-ball, or route one-style soccer (Wilson 2008). Perhaps the most famous of these managers was Graham Taylor. Taylor was a proponent of the long-ball game that Reep's analyses indicated to be the most effective style of play to score goals. His methods were successful, too. Taylor won four promotions in five years and led an unfashionable Watford team to second in the then first division (now called the Premier League). In 1990, Taylor was appointed as the manager of the England National Team. Using the same long-ball tactics though, Taylor's reign as England manager was unsuccessful with England failing to get out of the group stages of the 1992 European Championships and subsequently failing to qualify for the 1994 World Cup.

The 4-4-2 system, while once the dominant global soccer formation has declined in popularity due to the emergence of more possession-based systems. In the 1970s, the Ajax coach (Rinus Michels) adopted the 4-3-3 system as part of the "total football" revolution (Wilson 2008). Also, Johan Cruyff used a 3-4-3 system while coach at Barcelona with a focus on total football, high pressing, and short passes (Wilson 2008). Cruyff's philosophy has famously influenced more modern-day tactics used by coaches like Pep Guardiola (at Barcelona, Bayern Munich, and Manchester City). Overall, systems such as the 4-3-3 and 3-4-3 (wingbacks) focus on midfield overloads and spatial control that provided qualities lacking in the more rigid 4-4-2 formation (Tenga et. al 2020; Guilianotti and Robertson 2004). While still used today, especially for teams

that prefer a defensive set up with a focus on counterattacking, use of the 4-4-2 has decreased over the last three decades.

The tactical evolution over time has created a rich environment for analyzing the influence of formation on outcomes in soccer games. The purpose of this research is to – using recent data from the English Premier League – examine the influence of different soccer formations on match outcomes. The principal outcome metric used in the analysis is expected goals (xG). In soccer matches, the xG metric represents an aggregated probabilistic measure of the likelihood a goal is scored by a team in a match based on the quality of goal-scoring opportunities (shots). Each shot in a match is given a value between 0 and 1 with 0 representing no chance of a goal being scored from the shot and 1 indicating a guaranteed goal. xG estimates are derived from historical data and machine learning models to assign a value to each shot based on a number of factors, including shot location, distance from goal, angle of shot, body part used (foot, head, etc.), type of play (open play, corner, free kick, etc.), number of defenders between the shot and goal, position of goalkeeper, shot context (first time effort, ball under control, etc.). xG is a relatively new data point to present an outcome metric of a match, beginning as a formal metric for soccer games in the early 2010s. Its popularity grew as a metric by the mid 2010s when mainstream media outlets, such as Match of the Day and Sky Sports, began to provide it in match summaries. We use expected goals – as opposed to goals – as the metric for match outcomes as it provides a more long-term assessment of the influence of formation strategies on outcomes. While the actual scoreline of a match can mask the dominance of one team over another due to a host of factors (such as missed chances or goalkeeper heroics), expected goals provide a more thorough indication of the effectiveness of one team's formation over another by

looking at the quality of chances created. A second set of results are also presented that use actual goals scored as the dependent variable. This research is novel as, despite the recent research on the role of formations on player and team physical performance, to our knowledge, this is the first analysis to examine the effects of formation on potential match outcomes.

In recent years, there has been an increase in research investigating the relationship between formations and player physical performance. This group of studies typically use GPS or multi-camera tracking systems to observe physical performance (Bradley et al. 2011; Tierney et al. 2016; Aquino et al. 2019). For example, Aquino et al. (2019) investigate how team formation influences player running performance. They find greater running demands for players in a 4-4-2 formation compared to a 4-2-3-1 system. Modric et al. (2020) analyze differences in position-specific running performances as a function of defensive formations. They find that midfielders that play on teams that play three at the back, as opposed to a back four, have significantly more accelerations and decelerations. For center backs, high-intensity running was greater for those playing in a back three compared to a four. Vilamitjana et al. (2021) also examine differences between playing three and four at the back, more specifically comparing a 3-4-3 system to a 4-2-3-1. They find more high-intensity patterns in a 4-2-3-1 formation, especially in midfielders and attackers. Another group of research examines the role of player positions on physical performance across team formation. Bradley et al. (2011) examine three distinct positional groups (defenders, midfielders, and attackers) using data from English Premier League games. They find lower total distances and high-intensity distances for defenders on teams playing in a 4-4-2 formation, compared to defenders in a 4-3-3 or 4-5-1 formation. Aquino et al. (2021) find that teams playing with wingbacks increase the running demands of players in wide areas

compared to those teams playing four at the back. Castellano et al. (2013) use data from Spain's top-flight division (LaLiga) and look at the 4-2-3-1, 4-3-3, and 4-4-2 formations and find that formation influences running distances and high-speed running. Results indicate that center backs have increased running distances in wingback systems than in a 4-3-3 system. Finally, Morgans et al. (2024), using two seasons of data from the English Premier League, examine the relationship between possession, the 3-5-2 and 4-3-3 formations, and player positions. They find a significant interaction effect between formation and possession for high metabolic load efforts. For example, they find that center backs run less in a 4-3-3 when playing on teams with high possession compared to playing in a wingback system. They also find that players exert fewer high metabolic load efforts in a 4-3-3 formation compared to a 3-5-2 formation when both teams exhibit very high possession stats.

A more limited body of research examines the effect of formation on team performance metrics. For example, Bradley (2011) analyzes 20 English Premier League games and find that the fraction of successful passes is greater in a 4-4-2 than either a 4-3-3 or 4-5-1. In terms of preferred formations, González-Rodenas et al. (2023) analyze 3,420 matches (from 2012-2021) in Spain's top flight division (LaLiga) and provide a comparative analysis to examine the distribution in team formations over this period. The 4-2-3-1 and 4-4-2 formations were the most often used, while there was a reduction in formations with five players in midfield (3-5-2) and an increase in the use of two players up front (strikers) over this time period.

Using a three-step empirical framework, our results show that formations do influence the difference in xG and goals. We show that the 4-3-3 system for the home team has its distinct advantages, as does the use of wingbacks against specific away-team set ups. Results also show that the 4-4-2 is the least effective system for the away team. In the remainder of the paper, we describe the data used in the analysis, the methodological framework, regression and predicted outcomes results, followed by concluding remarks.

Data

The data were collected and merged from two sources for all English Premier League games taking place between the 2016-17 and 2021-22 seasons. First, all match scores, xG, and expected goals against (xGA) data were collected from the "fivethirtyeight" sport data and prediction website.[2] All data were downloaded into Stata for analysis. The remaining match-related data, such as possession stats, completed passes, and home and away team formations, were collected individually for each match. There are 380 matches played in each Premier league season, providing 2,280 observations across the six seasons of collected data. Table 1 describes these data.

The mean number of goals for the home team per game is 1.51 while the away team scores 1.24 goals per game. Pairwise The difference is statistically significant (p=0.00).[3] For the home team,

[2] The website offers metrics and forecasts for a variety of sports and is owned by ABC and ESPN is the U.S. < https://fivethirtyeight.com>

[3] Each observation in the dataset represents a single match that includes metrics (goals scored, possession, etc.) for both the home team and away team. These metrics constitute paired measurements from the same event. The paired

expected goals of 1.56 are statistically greater than the away team's expected goals of 1.29 (p=0.00). The possession for the home team is, on average, 51.2% with 48.8% for the away team. The home team, on average, makes significantly (p=0.00) more passes (466) than the away team (447). The home teams also has significantly more shots (13.6 versus 11.4) and significantly more shots on target (4.7 versus 4.0). In terms of disciplinary records, the away team receives significantly more yellow cards (p=0.00) than the home team but there is not any statistical difference in red cards received between the two teams (p=0.21).

This research analyzes 4 formation types. Specifically, 4-4-2, 4-3-3, 4-5-1, and wingbacks. All formations in the dataset are allocated to one of these four formation types. Some formations are therefore aggregated. For example, a 4-2-3-1 is coded in the analysis as a 4-3-3. Essentially, a 4-3-3 has a back four (defenders) and three forward players (a central striker, a left-sided forward and a right-sided forward). In the midfield, a 4-3-3 operates with a single central defensive midfielder (CDM) and operates with two Number Eights or central attacking midfielders (CAM). A 4-2-3-1 system has a very similar structure. The back four and forward three are the same. The only difference is that a 4-2-3-1 operates with two CDMs and one CAM. As such, both formations operate with a back four, three midfielders, and three attacking players. Therefore, we code all 4-3-3 and 4-2-3-1 formations as a 4-3-3. Aggregating like systems (such as a 4-3-3 and 4-2-3-1) has two main advantages. First, it provides more variation across systems for the analysis. Second, it also helps deal with the potential issue of within-game system changes. Formations are fluid so, for example, with the similarity of a 4-3-3 and 4-2-3-1, a team playing a

t-test is therefore the appropriate test of whether the metric means differ between home and away teams, as it accounts for the within-match dependence of these metrics rather than treating them as independent samples.

4-3-3 may drop one of the Number Eights back to play as a Number Six in the build-up phase, as such shifting the formation to a 4-2-3-1. Aggregating the two formations will help overcome the potential issue of within-game formation fluidity. The same logic applies to any teams operating with wingbacks. For example, a team that plays 3-at-the-back or 5-at-the-back are both using wingbacks. Both systems typically have three defending players when in possession of the ball (wingbacks pushing high and wide) and five defending players when not in possession of the ball (wingbacks dropping back to provide a back five). As such, all wingback systems in the data are coded as "wingbacks" regardless of whether they set up as a 3- or 5-at-the-back system.

Table 2 provides the breakdown of usage of the four systems. Both home and away, the 4-3-3 formation is the most popular system (47.7% at home and 44.3% away). The use of wingbacks, both home and away, is the next most-used system (23.9% at home and 26.7% away). Approximately one-fifth of teams use a 4-4-2 and less than 10% of teams use the 4-5-1 formation in matches.

With the main purpose of this research to examine the effect of formation on xG differences, Table 3 provides a cursory glance at xG differences across formation types. The table shows the average differences in expected goals calculated as:

$$xg\ diff = (xg\_home_f - xg\_away_f) \quad (1)$$

where *f* represents team formation. Just looking at the average differences in expected goals provides the first insight into the potential effects of formation differences on match outcomes. The largest average difference exists when the home team plays a 4-3-3 formation and the away side plays a 4-5-1. In fact, a home side that plays a 4-3-3 formation appears to have the greatest

impact on xG differences. If the home side plays a 4-3-3 and the away side plays either a 4-4-2 or with wingbacks, then the home team is expected to score 0.791 more goals per game. The weakest formation appears to be the 4-4-2. If the away side plays a 4-4-2 against a different home formation, then the magnitude of home expected goals over away expected goals ranges from 0.315 to 0.791. If the home team plays a 4-4-2 then the xG difference ranges from -0.391 to 0.355.

Home team advantage is visible when comparing the same formations for home and away sides. When using the same formation, the home side has a higher xG. The largest difference is when both teams play a 4-5-1, the home side is expected to score 0.491 more goals than the away team.

Methodology

To evaluate the tactical effectiveness of different formation matchups, we estimate the impact of home and away formations on differences in expected goals. This is a three-stage analysis. The first stage conducts a fixed effects linear regression using the following specification.

$$xg\_diff_{\mathrm{i}} = \sum_h \beta_{1h} \cdot HomeFormation_{\mathrm{i}h} + \sum_{\mathrm{a}} \beta_{2\mathrm{a}} \cdot AwayFormation_{\mathrm{i}\mathrm{a}} + \sum_h \sum_{\mathrm{a}} \beta_{3h\mathrm{a}} \cdot (HomeFormation_{\mathrm{i}h} \cdot AwayFormation_{\mathrm{i}\mathrm{a}}) + \gamma_1 \cdot ShotsTarget\_Diff_{\mathrm{i}} + \gamma_2 \cdot Possession\_Diff_{\mathrm{i}} + \gamma_3 \cdot Red\_Diff_{\mathrm{i}} + \alpha\_team(_{\mathrm{i}}) + \alpha\_season(_{\mathrm{i}}) + \varepsilon_{\mathrm{i}} \quad (2)$$

where $xg_{diff_{\mathrm{i}}}$ is the difference in expected goals in match *i* (home – away), β_{1h}, $\beta_{2\mathrm{a}}$, and $\beta_{3h\mathrm{a}}$ are the coefficients on dummy variables for home, away, and the home*away interaction. Their inclusion enables full factorial interaction between all levels of home formation, away formation,

and the interaction of home and away formations. Essentially, their inclusion lets the model estimate how each combination of home and away formations influences xg_{diff_i}. The γs represent coefficients on the included covariates of differences in shots on target (home – away) for each *i,* differences in percent possession (home – away), and differences in the number of red cards (home – away). The αs represent fixed effects at the team and season level, and ε_i is the error term. So, the model relies on within-team and within-season variation, removing potential bias from time-invariant team characteristics such as team personnel and strength, or seasonal trends such as viewership affecting all sides. Standard errors are computed using robust (Huber-White) estimators to account for heteroskedasticity.

The second stage in the analysis predicts the effects of away or home team formation on xG differences using the regression coefficient estimates, with an assumed formation for the other team fixed. So, for example, the predictive margins – average predicted values of the difference in xG – for each away formation are estimated, holding constant the assigned home team formation, controlling for the other covariates. This step is performed for each possible formation combination (e.g. 433, 442, etc.). So, this step estimates the expected differences in xG for each possible away formation given the home team's assumed formation and vice versa.[4]

In the final stage of the analysis, to really get at the research question of "what is the optimal formation for the home (away) team given the formation of the away (home)?", pairwise comparisons between the predicted margins (xG differences) are estimated for all the different

[4] The model is symmetric in how it treats home and away formations so the reverse calculations are not required. That is, the model includes a full factorial interaction of home*away, so the model estimates a separate effect for every unique home and away combination. As such, results can be presented in a matrix form.

possible home or away formations. For example, given a home team's formation, the predicted margin for an away team playing a 4-4-2 compared to a 4-3-3 is calculated as

$$margin(442\ v.\ 433) = predicted\ xg\_diff(away\ formation = 442) - predicted\ xg\_diff(away\ formation = 433) \quad (3)$$

Standard errors and 95% confidence intervals are also estimated using the delta method. From this, we can determine whether the pairwise comparisons in xG between home (away) team formations are statistically significant when the away (home) team plays a specific formation.

Results

The results from both the xG and goals models are presented, however, in line with the focus of the research, most of the discussion is reserved for the xG results. The results from the first-stage fixed effects linear regression models indicate both xG and goal differences (home-away) based on formation (see Table 4). The omitted formation in both models is a 4-3-3 by both teams. The results on home team formation provide the first insight into a potential advantage to playing a 4-3-3 system. For example, suppose the away team opts for a 4-3-3 formation. If the home team plays a 4-4-2, then the findings indicate that the difference in xG is 0.226 lower than if the home team played a formation of 4-3-3. This is statistically significant at the 0.01 level. Further, if the home team plays a 4-5-1 then it underperforms compared to a home 4-3-3 formation by 0.201 xG (significant at the 0.01 level). A home team playing with wingbacks has no statistical effect on xG compared to playing a 4-3-3. Next, suppose the home team chooses a 4-3-3 formation. With respect to the away teams, we don't see any statistical influence of team formation on xG differences compared to a 4-3-3. In terms of actual goal difference, the results are qualitatively

similar. The main finding is that a home 4-3-3 formation outperforms a 4-4-2 by 0.258 goals, on average, assuming that the away team plays 4-3-3.

When interacting other home and away formations, the interactions show some significant effects of formation on xG. These coefficients are relative to the base case of both teams playing a 4-3-3. So, for example, for the 4-4-2 v. 4-4-2 combination, i.e., if both teams play a 4-4-2, the difference in xG is 0.022 (i.e. -0.226 + 0.125 + 0.123) higher than if they both play a 4-3-3 (this result is statistically significant at the 0.01 level).[5] A similar combination of 4-4-2 by both teams significantly reduces actual goal difference by 0.265 (i.e., -0.257 + 0.121 - 0.129). We also see a statistically significant effect if, relative to both teams playing a 4-3-3, the home team plays with wingbacks and the away team plays a 4-4-2; the match xG difference is then expected to be 0.152 (i.e., 0.025 + 0.125 – 0.302) lower (significant at the 0.10 level). For the same combination, a statistically insignificant result holds for actual goals; the difference between home and away goals reduces by 0.218 (i.e., 0.041 + 0.121 - 0.380).

For the covariates, the results are intuitive. If the home team has one more shot on target than the away team, their xG difference is 0.203 higher. For actual goals, 0.276 more goals are scored for the home team for each additional shot on target. For possession, an increase in possession difference (home – away) of 1 percentage point is associated with an insignificant difference in xG. So, possession alone doesn't explain xG differences. This is not an unexpected result as teams, regardless of formation, can sometimes allow the other team to have more possession and try and play in transition on the counter attack. This is perhaps further supported by the negative and statistically significant possession coefficient in the goals model. Here, more possession

[5] Note, this entails a joint test of the three coefficients corresponding to the home and away side's 4-4-2 formation.

leads to fewer goals. The results do indicate xG effects with respect to red cards, with the direction of the coefficient as expected. One more red card for the home team than the away team is associated with a 0.256 decrease in the difference in xG, significant at the 0.01 level. A similar effect occurs for goals scored.

In the second stage of the analysis, results from the first-stage linear regression are used to estimate the adjusted predictions of differences in xG (and actual goals) for each home (away) formation, while holding the away (home) team formation fixed. Tables 5 and 6 shows all possible home and away team formation comparisons with respect to xG and goal differences. Again, with the focus on the xG model, the first main takeaway is that the 4-3-3 and wingback formations dominate for the home team. With respect to playing a 4-3-3 at home, all predicted values of xG differences are positive, indicating that a home-team 4-3-3 formation outperforms the away team in terms of expected goals irrespective of the away team's formation. Specifically, the home team's 4-3-3 outperforms an away team's 4-3-3, 4-4-2, 4-5-1, and wingback system by an expected goal difference of 0.317, 0.442, 0.317, and 0.276, respectively. All these results are statistically significant at the 0.05 level. The predicted goal margins (Table 6) also support the finding that the 4-3-3 home formation dominates with teams scoring, on average, 0.380, 0.501, and 0.240 more goals against a 4-3-3, 4-4-2, and wingback systems, respectively. The findings also show that playing wingbacks at home generates a positive and significant xG difference against each away formation except a 4-5-1.[6] Other significant comparisons are that a home 4-4-2 positively influences xG difference when playing an away team using wingbacks or 4-4-2, and a home team 4-5-1 outperforms an away team 4-4-2 formation.

[6] The positive effect of a home wingback system against an away 4-4-2 is significant at the 0.1 level.

The dominance of home teams playing the 4-3-3 and wingback formations may be suggestive of the importance of advanced width for the home team. Both formations typically use forward players high and wide to stretch the opposing team. Perhaps as the home team with an added incentive to attack, the use of such width can apply sustained pressure on the away team in a way that a 4-4-2 or 4-5-1 cannot match. Further, again as a home team, these two systems may help the home team be aggressive in their high pressing (and just as importantly, counter-pressing) structure to further place the away team under stress. Finally, it may also be that a 4-3-3 or use of wingbacks at home takes advantage of the fluidity of these systems in the build-up and creativity phases, relative to the perhaps more rigid 4-4-2 or 4-5-1 systems.

Another main finding from the predicted margins is that the 4-4-2 away team formation consistently underperforms against other home team formations with respect to xG differences. As mentioned, if the home team plays a 4-3-3, then the home team is expected to score 0.442 more goals than an away team with 4-4-2 formation. If the home team plays with a 4-5-1 or with wingbacks against an away 4-4-2, then the home is expected to score 0.406 and 0.165 more goals per game, respectively. If both teams set up with a 4-4-2, then the home team is expected to outperform the away team by 0.339 xG. The goals model lends support to this finding with an away 4-4-2 conceding, on average, 0.501 and 0.645 more goals to a home 4-3-3 and 4-5-1 formation, respectively. Combined, these findings provide support and understanding for the movement away from a traditional 4-4-2 in English soccer. The fact that a 4-4-2 underperforms as a formation away from home is likely because the other formations can overload the central midfield areas with three midfielders relative to two in a 4-4-2. This means that the 4-4-2 can be overrun in the midfield area, meaning they cede possession and control to the home team.

Finally, formation matchups involving a 4-5-1 and wingbacks are balanced. If either formation is home or away against the other, there is no statistical difference in xG differences.

In the third stage of the analysis, pairwise comparisons for home (away) team formations conditional on away (home) team formations are considered. The pairwise comparisons are used to calculate the predicted margins in terms of xG and goal differences. The results are shown in Tables 7 and 8. This essentially extends the second-stage results and provides direct formation-on-formation comparisons.

Again, focusing on the more of a long-run analysis using the xG model, results reinforce the finding that a home 4-3-3 or wingback system have their distinct advantages, and that an away team 4-4-2 formation is least effective. For example, given the home team plays a 4-4-2, then if the away team plays a 4-4-2, the xG difference is significantly higher (0.249 xG) for the home team compared to if the away team set up in 4-3-3 formation. In other words, if the home team is playing a 4-4-2, then the away team's xG performance is better if they play a 4-3-3 rather than a 4-4-2. The same story holds if the home team plays a 4-5-1. Again, the away team will underperform if it plays a 4-4-2 relative to a 4-3-3 in terms of xG difference (significant at the 0.1 level). Further, if the home team plays 4-5-1, the difference in xG is 0.331 lower when the away team plays a 4-4-2 instead of wingbacks (significant at the 0.1 level). So, when competing against a 4-5-1 formation at home, both the 4-3-3 and wingback setups by the away side outperform the 4-4-2. Reasons for this may include the ability of a 4-3-3 and a wingback system (relative to a 4-4-2) to compete in the midfield areas as the away side plays against a team that also has three midfielders. Further, with three forward players, a 4-3-3 or a wingback system may be better set up (compared to 4-4-2) to aggressively press a 4-5-1. As such, the home team

4-5-1 will likely have more time to build possession against a 4-4-2 compared to the away team employing 4-3-3 or wingbacks.

Now turning to the home formation comparisons, again the strength of the 4-3-3 is apparent. The results show that a home 4-4-2 and 4-5-1 formation underperform compared to a 4-3-3 when playing against an away 4-3-3. In other words, if the away team plays a 4-3-3, then the home team generates 0.226 and 0.201 greater xG difference if it also plays a 4-3-3 rather than a 4-4-2 or 4-5-1, respectively. Results also indicate that using wingbacks at home can perform favorably against an away 4-3-3 relative to playing either a 4-4-2 or 4-5-1. Specifically, if the away team sets up as a 4-3-3, then using wingbacks as the home team is expected to increase xG differences for the home team by 0.251 and 0.226 compared to using a 4-4-2 or 4-5-1, respectively. A potential reason why wingbacks perform better against a 4-3-3 than a 4-4-2 or 4-5-1 may be in the use of width in an attacking transition. The 4-3-3 away team may be compact without the ball and so high and wide wingbacks may be able to take advantage of that. Relative to that, a home 4-4-2 or 4-5-1 may not have the same width to exploit the compactness of the away team 4-3-3.

The goals model (Table 8) typically supports the findings that home 4-3-3 and wingbacks perform well, especially against a 4-4-2. For example, if the away team plays a 4-3-3 or 4-4-2 then a home 4-3-3 scores 0.258 and 0.387 more goals than a home 4-4-2, respectively. A home team playing with wingbacks also scores 0.299 more goals than a 4-4-2 if the away team plays 4-3-3, and an away team playing wingbacks scores, on average, more goals than if they play a 4-4-2, against a home team 4-3-3.

Conclusion

The purpose of this paper is to examine the influence of soccer formations on game outcomes. While other studies have considered the role of team formation on individual player output and performance, this is the first study to examine formation effects on match outcomes. Conceptually, this is a classic duopoly problem with two teams trying to make strategic decisions where the payoff to Team A depends on not only Team A's formation but also the formation set up of Team B. The decision on formation is not independent. It reflects how one team believes the other will set up. For example, if Team A believes (based on historical and current evidence) that Team B will play a 4-4-2 system, Team A will want to choose the system that they believe will optimize their payoff. The payoff to a team in this instance is reflected in increased expected goal or actual goal difference. So, just as firms in a duopoly competition iterate on strategy in anticipation of the other's move, soccer teams must anticipate and optimize formation choices dynamically. This behavior is representative of a strategic game with both sides trying to maximize payoffs.

Six seasons of match-level English Premier League data were collected and merged with match-level expected goals data. Four formation types are considered – 4-3-3, 4-4-2, 4-5-1, and wingbacks. The principal metric used for match outcome is expected goals. From a coaching perspective, we believe that analyzing the effect of formations on expected goals, rather than actual goals scored, represents a more informed examination of its effect. While goals scored cam sometimes mask the dominance of one team over another, expected goals provides a more realistic representation of the chances created in a match.

A three-step analysis was conducted. First, a fixed effects linear regression, controlling for the home team, away team, and home*away interactions, plus covariates, was estimated to examine the effect of team formation on expected goal differences. The coefficient estimates were then used to predict the expected goals differences for each formation type, holding constant the opposing team's structure, and then the predicted pairwise comparisons between formation types for home (away) teams, given the formation of the away (home) team.

The findings indicate that playing a 4-3-3 or wingback formation dominates for the home team. Playing a 4-3-3 at home has distinct performance advantages over all away formations. This result holds when controlling for in-game covariates and fixed effects at the team and season level. Its apparent success may help explain why almost half of teams in the data play a 4-3-3 home formation. Pairwise comparisons provide more detailed information regarding the success of playing a 4-3-3 home formation, indicating that a home 4-4-3 outperforms both a home 4-4-2 and 4-5-1 formation when playing against an away 4-3-3. Playing wingbacks at home generates a positive xG difference against each away formation except a 4-5-1. We reason that the success of the 4-3-3 and wingback systems for the home team may be due to their ability to press aggressively in the opposition's defensive third, establishing a territorial advantage, engaging the home support, and ultimately creating goal-scoring opportunities. Further both formations have a three-player midfield, so can gain control of the midfield and create player overloads in this important area of the pitch.

Findings also indicate that the more traditional 4-4-2 formation appears to be the least effective system when used by away teams. When matched against a home 4-5-1, wingbacks, or 4-3-3 formation, an away 4-4-2 set up results in significantly lower xG differences. This finding is

consistent with the the tactical shift away from this formation in the modern game. The 4-4-2 can be overloaded in the midfield area when competing against a 4-4-3 or a team with wingbacks so it can be difficult for the away 4-4-2 team to maintain possession and control the game in the midfield area, perhaps placing more physical demands on the players. This notion is supported by Aquino et al. (2019), who find greater running demands for players in a 4-4-2 formation compared to a 4-3-3 system.

This is the first research to investigate the strategic impact of soccer formations on match outcomes. There are limitations to the work. Principally, teams may shift formation during matches. While this work cannot account for the total fluidity of the modern game, we try to mitigate this effect by aggregating formations. For example, a team that sets up as a 4-2-3-1 may push one of their Number Sixes further forward while in possession to create a 4-3-3. By aggregating the two formations as we do, then results are less sensitives to these changes. Future research is challenged to find and utilize more granular real-time player-tracking data to try and control for within-match formation shifts.

References

Aquino, R., Carling, C., Palucci Vieira, L. H., Martins, G., Jabor, G., Machado, J., Santiago, P., Garganta, J. and E. Puggina. 2020. Influence of Situational Variables, Team Formation, and Playing Position on Match Running Performance and Social Network Analysis in Brazilian Professional Soccer Players. *Journal of Strength and Conditioning Research* 34(3): 808–817. https://doi.org/10.1519/JSC.0000000000002725

Bradley, P. S., Carling, C., Archer, D., Roberts, J., Dodds, A., Di Mascio, M., Paul, D., Diaz, A.G., Peart, D. and P. Krustrup. 2011. The Effect of Playing Formation on High-Intensity Running and Technical Profiles in English FA Premier League Soccer Matches. *Journal of Sports Sciences* 29(8): 821–830. https://doi.org/10.1080/02640414.2011.561868

Castellano J., Álvarez D., Figueira B., Coutinho D. and J. Sampaio. 2013. Identifying the Effects from the Quality of Opposition in a football Team Positioning Strategy. *International Journal of Performance Analysis in Sport* 13(3): 822–832. https://doi.org/10.1080/24748668.2013.11868691

González-Rodenas J, Moreno-Pérez V, Campo RL, Resta R, Coso JD. Evolution of Tactics in Professional Soccer: An Analysis of Team Formations from 2012 to 2021 in the Spanish LaLiga. J Hum Kinet. 2023. 15(87): 207–216. doi: 10.5114/jhk/167468. PMID: 37559775; PMCID: PMC10407311.

Guilianotti, R., and R. Robertson. 2004. The Globalization of Football: a Study in the Glocalization of the 'Serious Life'. *The British Journal of Sociology* 55(4): 545–568.

Modrić, T., Versić, S. and D. Sekulić. 2020. Position Specific Running Performances in Professional Football (Soccer): Influence of Different Tactical Formations. *Sports* 8(12): 161. https://doi.org/10.3390/sports8120161

Morgans, R., Radnor, J., Fonseca, J., Haslam, C., King, M., Rhodes, D., Żmijewski, P. and R. Oliveira. 2024. Match Running Performance is Influenced by Possession and Team Formation in an English Premier League Team. *Biology of Sport* 41(3): 275–286. https://doi.org/10.5114/biolsport.2024.135414

Reep, C., and B. Benjamin. 1968. Skill and Chance in Association Football. *Journal of the Royal Statistical Society, Series A (General)* 131(4): 581–585

Tenga, A., Holme, I., Ronglan, L. T., and R. Bahr. 2010. Effect of Playing Tactics on Achieving Score-Box Possessions in Football. *Journal of Sports Sciences* 28(9): 967–973.

Tierney, P. J., Young, N., Clarke, A., and M. Duncan. 2016. Match Play Demands of 11 Versus 11 Professional Football Using Global Positioning System Tracking: Variations Across Common Playing Formations. *Human Movement Science 49*: 1–8. https://doi.org/10.1016/j.humov.2016.05.007

Vilamitjana, J. J., Capdevila, L., Torregrossa, M., Buscà, B. and B. Fèrre. 2021. High-Intensity Activity According to Playing Position with Different Team Formations in Soccer: 1-3-4-3 vs. 1-4-2-1-3. *Acta Gymnica* 51, Article e2021.003. https://doi.org/10.5507/ag.2021.003

Wilson, J. 2008. *Inverting the Pyramid: The History of Football Tactics*. Orion Publishing Group.

Table 1. Data Summary

Variable	Description	Mean	Standard Deviation	Min	Max
Hgoals	Home goals	1.51	1.31	0	9
Agoals	Away goals	1.24	1.22	0	9
xG	Expected home goals	1.56	0.91	0.0	7.1
AxG	Expected away goals	1.29	0.82	0.0	5.1
Hpossession	Home team possession (%)	51.15	13.38	17	83
Apossession	Away team possession (%)	48.86	13.39	17	83
Hpasses	Home team number of passes	465.96	139.75	176	1015
Apasses	Away team number of passes	446.58	135.68	60	976
Hshots	Home team number of shots	13.63	5.71	0	38
Ashots	Away team number of shots	11.36	5.01	0	31
Hshotstarget	Home team number of shots on target	4.68	2.67	0	19
Ashotstarget	Away team number of shots on target	3.96	2.39	0	15
Hred	Home team number of red cards	0.05	0.23	0	2
Ared	Away team number of red cards	0.06	0.25	0	2
Hyellow	Home team number of yellow cards	1.56	1.25	0	11
Ayellow	Away team number of yellow cards	1.67	1.26	0	9

Table 2. Team Formations

Formation	Home Percent	Away Percent
4-3-3	47.7%	44.3%
4-4-2	21.1%	20.9%
4-5-1	7.3%	8.1%
Wingbacks	23.9%	26.7%

Table 3. Differences in Expected Goals, by Formation

Home Formation	Away Formation	xG Difference
433	433	0.236
433	442	0.792
433	451	1.115
433	Wingbacks	0.718
442	433	-0.391
442	442	0.355
442	451	0.280
442	Wingbacks	0.076
451	433	-0.488
451	442	0.699
451	451	0.491
451	Wingbacks	-0.179
Wingbacks	433	-0.175
Wingbacks	442	0.315
Wingbacks	451	0.035
Wingbacks	Wingbacks	0.385

Table 4. Fixed Effects Linear Regression Results

	Difference in Expected Goals		Difference in Goals	
Variable	Coefficient	P-Value	Coefficient	P-Value
Home Formation				
442	-0.226	0.006	-0.258	0.029
451	-0.201	0.056	-0.128	0.383
Wingbacks	0.025	0.734	0.041	0.735
Away Formation				
442	0.125	0.121	0.121	0.312
451	0.001	0.994	-0.214	0.185
Wingbacks	-0.040	0.602	-0.141	0.218
Home Away Interaction				
442 442	0.123	0.355	-0.129	0.509
442 451	0.016	0.932	0.156	0.569
442 Wingbacks	0.155	0.183	0.166	0.351
451 442	0.165	0.347	0.271	0.259
451 451	0.029	0.894	-0.392	0.532
451 Wingbacks	-0.000	0.999	0.095	0.707
Wingbacks 442	-0.302	0.021	-0.380	0.069
Wingbacks 451	-0.270	0.193	-0.191	0.624
Wingbacks Wingbacks	-0.006	0.956	-0.200	0.235
Covariates				
Shots on Target (home-away)	0.203	0.000	0.276	0.000
Possession (home-away)	0.000	0.784	-0.026	0.000
Red Cards (home-away)	-0.256	0.000	-0.577	0.000
Constant	0.168	0.000	0.238	0.000
Obs	2280		2280	
Adjusted R-Squared	0.614		0.517	

Table 5. Predicted xG Differences for Away and Home Team Formations

Away Formation Predicted Values	Home Team Formation							
	433		442		451		Wingbacks	
	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.
433	**0.317**	**0.232; 0.401**	0.090	-0.048; 0.229	0.116	-0.075; 0.306	**0.342**	**0.222; 0.461**
442	**0.442**	**0.312; 0.572**	**0.339**	**0.148; 0.530**	**0.406**	**0.143; 0.669**	*0.165*	*-0.029; 0.359*
451	**0.317**	**0.153; 0.482**	0.107	-0.191; 0.404	0.145	-0.201; 0.491	0.072	-0.281; 0.426
Wingbacks	**0.276**	**0.150; 0.403**	**0.205**	**0.066; 0.344**	0.075	-0.191; 0.341	**0.295**	**0.177; 0.414**

Bold coefficients indicate significance at the 0.05 level

Italicized coefficients indicate significance at the 0.1 level

Table 6. Predicted Goal Differences for Away and Home Team Formations

Away Formation Predicted Values	Home Team Formation							
	433		442		451		Wingbacks	
	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.
433	**0.380**	**0.249; 0.513**	0.123	-0.066; 0.312	*0.252*	*-0.010; 0.515*	**0.422**	**0.221; 0.623**
442	**0.501**	**0.315; 0.689**	0.115	-0.167; 0.397	**0.645**	**0.290; 1.000**	0.163	-0.153; 0.480
451	0.167	-0.119; 0.453	0.066	-0.369; 0.500	-0.354	-1.525; 0.818	0.017	-0.677; 0.711
Wingbacks	**0.240**	**0.060; 0.420**	0.148	-0.099; 0.397	0.207	-0.194; 0.608	0.080	-0.128; 0.288

Bold coefficients indicate significance at the 0.05 level

Italicized coefficients indicate significance at the 0.1 level

Table 7. Predicted xG Margins for Away and Home Team Formations

Away Formation Pairwise Comparisons	Home Team Formation							
	433		442		451		Wingbacks	
	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.
442 v. 433	0.125	-0.033; 0.284	**0.249**	**0.020; 0.477**	*0.290*	*-0.032; 0.613*	-0.176	-0.404; 0.052
451 v. 433	0.001	-0.185; 0.186	0.016	-0.301; 0.334	0.029	-0.356; 0.415	-0.269	-0.639; 0.100
Wingbacks v. 433	-0.040	-0.192; 0.111	0.115	-0.072; 0.301	-0.040	-0.364; 0.283	-0.046	-0.210; 0.117
451 v. 442	-0.125	-0.332; 0.083	-0.232	-0.575; 0.110	-0.261	-0.685; 0.164	-0.093	-0.494; 0.308
Wingbacks v. 442	*-0.166*	*-0.350; 0.018*	-0.134	-0.363; 0.096	*-0.331*	*-0.702; 0.040*	0.130	-0.096; 0.357
Wingbacks v. 451	-0.041	-0.246; 0.164	0.099	-0.214; 0.411	-0.070	-0.505; 0.365	0.223	-0.147; 0.594
Home Formation Pairwise Comparisons	Away Team Formation							
	433		442		451		Wingbacks	
	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.	xG Diff	95% Conf. Int.
442 v. 433	**-0.226**	**-0.389; -0.064**	-0.103	-0.326; 0.119	-0.211	-0.550; 0.128	-0.071	-0.259; 0.116
451 v. 433	*-0.201*	*-0.407; 0.005*	-0.036	-0.322; 0.249	-0.172	-0.554; 0.210	-0.201	-0.489; 0.086
Wingbacks v. 433	0.025	-0.119; 0.168	**-0.277**	**-0.497; -0.057**	-0.245	-0.635; 0.144	0.019	-0.151; 0.189
451 v. 442	0.025	-0.206; 0.256	0.067	-0.245; 0.379	0.039	-0.419; 0.496	-0.130	-0.427; 0.167
Wingbacks v. 442	**0.251**	**0.070; 0.432**	-0.174	-0.436; 0.088	-0.034	-0.496; 0.427	0.090	-0.093; 0.273
Wingbacks v. 451	**0.226**	**0.006; 0.446**	-0.241	-0.555; 0.073	-0.073	-0.565; 0.419	0.220	-0.070; 0.511

Bold coefficients indicate significance at the 0.05 level

Italicized coefficients indicate significance at the 0.1 level

Table 8. Predicted Goal Margins for Away and Home Team Formations

Away Formation Pairwise Comparisons	Home Team Formation							
	433		442		451		Wingbacks	
	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.
442 v. 433	0.121	-0.113; 0.356	-0.008	-0.335; 0.320	*0.393*	*-0.048; 0.833*	-0.259	-0.626; 0.108
451 v. 433	-0.214	-0.529; 0.102	-0.057	-0.511; 0.396	-0.606	-1.806; 0.594	-0.405	-1.116; 0.305
Wingbacks v. 433	-0.141	-0.365; 0.083	0.025	-0.265; 0.316	-0.046	-0.514; 0.423	**-0.342**	**-0.615; -0.069**
451 v. 442	*-0.335*	*-0.670; 0.000*	-0.049	-0.559; 0.459	-0.999	-2.207; 0.210	-0.146	-0.896; 0.604
Wingbacks v. 442	**-0.262**	**-0.521; -0.003**	0.033	-0.333; 0.399	-0.438	-0.970; 0.094	-0.083	-0.455; 0.289
Wingbacks v. 451	0.073	-0.258; 0.404	0.083	-0.395; 0.560	0.560	-0.679; 1.800	0.063	-0.650; 0.776
Home Formation Pairwise Comparisons	Away Team Formation							
	433		442		451		Wingbacks	
	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.	G Diff	95% Conf. Int.
442 v. 433	**-0.258**	**-0.489; -0.027**	**-0.387**	**-0.714; -0.059**	-0.102	-0.618; 0.415	-0.092	-0.397; 0.214
451 v. 433	-0.128	-0.416; 0.160	0.143	-0.248; 0.534	-0.521	-1.723; 0.682	-0.033	-0.460; 0.393
Wingbacks v. 433	0.041	-0.197; 0.280	*-0.339*	*-0.691; 0.014*	-0.150	-0.896; 0.596	-0.160	-0.426; 0.107
451 v. 442	0.130	-0.189; 0.448	**0.530**	**0.094; 0.966**	-0.419	-1.664; 0.826	0.059	-0.404; 0.521
Wingbacks v. 442	**0.299**	**0.017; 0.581**	0.048	-0.364; 0.460	-0.049	-0.874; 0.777	-0.068	-0.394; 0.258
Wingbacks v. 451	0.170	-0.156; 0.495	**-0.482**	**-0.946; -0.017**	0.370	-0.997; 1.738	-0.127	-0.573; 0.320

Bold coefficients indicate significance at the 0.05 level

Italicized coefficients indicate significance at the 0.1 level