A Statistical Analysis of

Landmark Conference Women's Soccer

By Nathan T. Rauscher

Submitted in Partial Fulfillment of the Requirements for the

Degree of Master of Education

May 2014

Graduate Programs in Education

Goucher College

Table of Contents

List of Tables

Abstract

The purpose of this study was to determine if a statistically significant relationship existed between box score statistics and success in winning Landmark Conference women's soccer games. The researcher used measures in the form of box score statistics from previously contested Landmark Conference women's soccer games. An independent samples median test was used for each of the following factors: number of shots taken, shots on goal, shots on goal percentage, corner kicks, and the number of fouls committed. A stepwise regression analysis, in particular a discriminant analysis, was used to determine whether variables entered after the best predictor contributed anything additional to the predictive power of the model. Initially, both shots taken and shots on goal were identified as significant predictors. However, once the best predictor, shots taken, was taken into account, none of the others added any additional information to improve the prediction. Shots taken and shots on goal are correlated and predictive of winning versus losing. The other two actions—fouls and percent shots on goal—are not. Therefore, the null hypothesis was retained for fouls and percent shots on goal and rejected for the other two actions. Despite the demonstrated importance of taking more shots, a team may not put all of its effort into shooting, as many other aspects of a match will affect shooting rates. This points to the need for future research on factors such as passing accuracy and team possession that affect the volume of shooting opportunities generated within a match.

CHAPTER I

INTRODUCTION

Overview

Over approximately the past 30 years, women's soccer has grown from a sport which once provided no opportunities for collegiate competition to one in which hundreds of teams compete across several divisions and conferences. As gaining entry into the NCAA tournament becomes increasingly prestigious, competition, especially between teams within conferences, has grown. Competition in the NCAA tournament demonstrates athletic success and helps attract attention for potential student enrollment, which has also become increasingly competitive.

While video scouting has become extremely common in preparing teams for important matches, teams continue to search for advantages. Especially within conference competition, which is the major factor determining NCAA tournament participation, teams can potentially gain an advantage by preparing for games with statistical analyses in mind. Previous studies in soccer and other athletic competitions have demonstrated various links between in-game playing factors, non-playing factors, and the likelihood of success for a team. Given unlimited resources, teams could analyze the importance of such statistics as dribbling success, passing, defensive tackling, fouling, offsides, and player positioning. With the limited resources of college soccer programs, however, it is important to examine those statistics that are available in order to find significant statistics that can contribute to success.

Statement of the Problem

The problem this study addresses is to determine how available box score statistics, including shots, corner kicks, and fouls, contribute to winning Landmark Conference women's soccer games.

Hypothesis

There will be no statistically significant relationship between wins and any of the measured statistics on shots taken, shots on goal, shots on goal percent, fouls, corner kicks, or game location (home versus away) in games played in Landmark Conference women's soccer.

Operational Definitions

The predictor variables are the number of shots taken, percent of shots on goal, number of fouls committed, and number of corner kicks in Landmark Conference women's soccer games.

The dependent variable is the outcome of each game. The operational definition is the number of wins in a season/winning percentage for Landmark Conference women's soccer games.

CHAPTER II

REVIEW OF THE LITERATURE

This literature will review success on the soccer playing field. The first section will examine various models used for predicting success in athletic events, particularly in soccer. The second section will discuss playing factors that can affect the results of soccer matches specifically. Finally, the third section will analyze non-playing factors that can affect the results of a competition.

Models for Predicting Success in Soccer Matches

Models are used in professional sports to predict winners and the likelihood of different results, largely for gambling purposes. Models are more specifically used by coaches and trainers to understand the specific causes and predictors for situation-specific scoring that can affect match preparation.

In line with predicting winners for specific matches, da Silva, Vainstein, Lamb, and Prado (2013) used two models to predict winners through the course of a round-robin tournament based on previous results within the same tournament. Their model does not predict winners using game statistics. However, it makes important predictions for wins based on previous results, which parallels competition for the NCAA tournament and league play. These types of models can assist in developing tactics to match the desired role of playing as the favorite or underdog. While da Silva et al.'s study provides a large-scale view of an entire competition, studies that examine specific factors that affect the result of a single match are more applicable for coaches. Coaches typically consider fitness, skill, their team's formation, and their opponent's formation as strong factors in determining the result of a match.

It is important to understand the basic impact goals scored and conceded can have over the course of an entire season in order to develop a more attacking or defensive playing style best suited to a team's particular league or competition. Maher (1982) conducted a more specific examination of total goals scored and goals conceded over the course of a league season to predict match results. However, this did not control for the tendency of more goals to be scored and fewer conceded by home teams in any match. For this reason, results were predicted to be draws far more frequently than is actually the case.

There are clearly many factors that control whether or not a team is able to score goals and prevent goals from being scored. Min, Kim, Choe, Eom, and McKay (2008) used Bayesian inference, similar to da Silva et al.'s (2013) methods, to predict more specific results between teams that have previously competed against each other. Their model, however, considers multiple factors including current score, morale, fatigue, and skills. These factors can clearly be important in preparing a team. Morale, for example, can be an especially important factor in cultivating a women's soccer team. In quantifying such factors as morale, fatigue, and skills, the study maintains content validity with a 16-question survey given to ten soccer-knowledgeable people. Each question in this survey rates a strong, normal, or weak effect on different controls for each factor. The study further examines the way these factors affect each other in breaking down the game into smaller frames. Min et al. compared their predictor to two similar football results predictors using a one-tailed *T*-test because they hypothesized their predictor was more accurate. They concluded their predictor was more accurate based on their small *p*-values for error.

While da Silva et al. (2013) examine winners over the scale of a tournament or league competition, Maher (1982) examines the specifics of goals scored and conceded, and Min et al. (2008) focus on even more specific factors that can affect the outcome of a match. Meanwhile, Goddard's (2005) study compares these types of models. In particular, he compares models that focus on goals to those that focus on overall results over the course of a season. He concludes that while the goals-based models use far more data and should therefore be more accurate predictors, there is very little difference in the prediction accuracy of these models. Ultimately, using a hybrid model, in which results are used as the dependent variable, provides the greatest accuracy.

The previous studies examined factors that affect either individual results or results over an entire competition. However, Moschini (2004) examines a more specific factor that coaches can utilize in training. The success rate for shots based on the location of the shot in relation to goalie positioning provides important clues that can help in both scoring and preventing goals. Important factors considered in this model are the fact that near post shots are more accurate than far post shots based on the distance. In examining all shots taken over the course of the 2002–2003 Italian league season, Moschini's model considers only those that are off-center, thereby providing the option of either a near post or far post shot. The results show that only approximately one-third of goals are scored at the near post in these situations. This is important information for training players both in shooting and saving goals. The implication is that goalies may overprotect their near post and should shift their positioning. Similarly, this indicates a clearer choice for players to increase their success rate for these types of shots. Of course, other factors may make adjustments to conventional wisdom inappropriate.

While the aforementioned studies focus on more specific factors that can be addressed in training, Carmichael, Thomas, and Ward (2000) examined the effect of 21 factors on match success. These range from attacking statistics including four different categorizations of shots, passes in different areas of the field, and dribbling success, to defending statistics including fouls, yellow cards, red cards, and goalkeeper plays. Given the type of factors and the huge number of factors examined in this study, it is much more indicative of a teams overall style of play than individual skills that can be addressed in training. For example, defensive tackling, clearances, blocks and interceptions are shown to be a significant indicator for success. These categories are more demonstrative of organized team defending than a statistic such as shots on goal percentage. In terms of defensive aggressiveness and tactics, differences in free kicks given away for fouls and off-sides are shown to be significant. Detailed statistical analyses such as these can demonstrate how effectively teams in a given league or conference are able to take advantage of a teams inability to defend well.

All of the models in these studies explore the ability to predict match results based on previous results or factors within a competition. Specifically, the importance of examining these results may help reveal patterns that teams can use to direct their training and increase their likelihood of success in a specific match, or, more importantly, over the course of an extended season.

Playing Style and Mentality as it Affects Success in Soccer Matches

Mentality, including a player's belief that he or she can win, can be a factor in his or her effort. Motivation and effort are clear factors in determining the result of an athletic contest. Though these must be taken as only a piece of the likelihood of success,

they are important to consider, especially from a coaching standpoint. In the instances of very evenly matched competition, it may prove the difference maker. Blavatskyy (2010) examines a contest success function (CSF) that "translates an individual's effort into his or her probability of winning" (p. 267). Essentially, Blavatskyy concluded that a draw is more likely when the contestants exhibit reduced effort. While this isn't applicable in most soccer matches over a league season, this can be important in instances when there is no incentive to win. An example would be the final game of a round robin group, in which the first two teams advance to the next round of the competition. If the two top teams face each other in the final game having already guaranteed entry into the next round, there is no incentive to give substantial effort, and so the likelihood of a draw is substantially increased.

More specific to playing styles, Dobson and Goddard (2010) examine the effect of a team's formations and playing style throughout a match. Defensive versus attacking tactics and violent versus nonviolent playing style can change for teams over the course of a match. The application of these tactics and styles at different times can influence the match result and simulations can help determine more effective times to apply these. Dobson and Goddard used data from over 12,000 matches over a period of six seasons in English soccer leagues. One significant finding was that the likelihood of a player being red carded out of the match is lowest when the score is even and highest for the losing team when they are losing by one or two goals. Additionally, players are significantly more likely (18 times for home teams and 9 times for away teams) to be red carded during the last five minutes of a match than the first ten minutes of a match. Essentially,

teams are more likely to play with a more violent style toward the end of the game, especially if they are a better team playing at home.

With regard to defensive versus attacking tactics, the data from Dobson and Goddard's (2010) study reveals several important trends that can help teams decide when to push for goals or be especially careful about conceding a goal. First, teams are significantly more likely to score in the last five minutes of a match than the first ten minutes. Though this may be at least in part due to the physical exhaustion at the end of a game, it is still important in showing that a team must be fit enough to focus through the entire game and that a team should be pushing for goals in the latter time frame.

Rampinini, Impellizzeri, Castagna, Coutts, and Wisloff (2009) examined the effect of fatigue on match results specifically. This study compared the technical and physical performance of players on successful teams to those on unsuccessful teams. By comparing a player's success in the first half to his or her success in the second half, Rampinini et al. were able to conclude that successful teams demonstrated far better fitness in the second half of a game. Though the conclusion itself isn't surprising, the method used to demonstrate it was. It is reasonable to expect more successful teams would have more ball possession and would therefore be required to run less. However, players on successful teams covered greater overall distances and completed more high-intensity running and very high-intensity running than players on less successful teams. The importance of being able to run extremely hard late in a game may be an important factor to consider in preparation as well as playing style. Possessing the ball in order to rest is an important tactical approach that can clearly play a factor in a team's success.

Less importantly, though still worth considering especially immediately following a goal being scored, is that goals are least likely to be scored while the game is tied. Finally, these different trends can be used by players and coaches, not only in English soccer, but in other leagues, as well, to help determine a style of play that provides the best opportunity to win a competition.

In addition to the factors cited above, optimism is a major factor in individual and therefore team performance. This is important to consider in preparing a team for a match. Players will undoubtedly take the tone of a coach's preparation as positive or negative depending on the types of drills the team practices, the style with which the team plays (playing to win or playing for a tie), and how the coach presents the chances of winning. Gordon (2008) examined the individual performance of 20 male athletes and 18 female athletes. The study further surveyed these athletes in order to determine any correlation between the dependent variable, athletic performance, and the independent variables, attributional style and dispositional optimism. The study was reliable in that the Attributional Style Questionnaire (ASQ) was given before the season as well as one month into the season. This questionnaire essentially indicated whether the athletes perceived positive versus negative events as being caused by themselves or outside forces. Assessors measured the athletic performance of the soccer players by examining five statistics over the course of eight matches. Gordon pointed out that because there was a smaller data set for three of these statistics, "the ratio of passes completed over passes attempted was used as a general performance index" (p. 340). The study conducted on female athletes demonstrated stronger performance for pessimists. This

may reflect the athletic personality of athletes being hard on themselves, which drives them to work continually harder in search of greater and greater success.

Non-playing Factors that May Affect Competition Results

Hagemann, Strauss, and Leissing (2008) showed taekwondo bouts of one competitor wearing blue and the other wearing red to 42 referees. They then showed the same bouts after digitally swapping the red and blue. For bouts with a clear winner, the results did not vary significantly. However, in close bouts, the competitor wearing red was awarded 13 percent more points than the competitor wearing blue despite actually being the same matches. This demonstrates an important bias that may affect soccer games and identifying potentially better players for opponents. First, soccer referees may see a team that wears red as more aggressive, therefore awarding more fouls against that team. Additionally, players who stand out with red may be viewed as more aggressive and therefore more likely to receive fouls that can eventually send them out of a match.

Heuer and Rubner (2009) examined more than 12,000 match results from the German soccer league from 1965 to 2007. They found that while home teams scored 0.7 more goals per game than away teams, no team was actually better at home than they are away. Essentially if a team is more likely to win at home, then they are also more likely to win away from home. Heuer and Rubner, however, only examined whether a game was played at home or away from home. Oberhofer, Philippovich, and Winner (2010) demonstrated that it is important to minimize the effects of travelling for away games by examining the distance travelled for away games. They found "that distance increases a guest team's propensity to concede goals and exhibits a negative but insignificant impact on the ability to score goals" (p. 200). Specifically, Oberhofer et al. found that a team's

ability to prevent goals is worst when travelling up to approximately 450 km. However, beyond this distance, defensive performance improves slightly. They hypothesize that this is because games that require travelling a greater distance are better prepared and often involve arriving at the host city one or two days early to be better prepared.

Summary

Models that predict results based on previous data can provide clues about how to prepare for matches. Additionally, direct factors such as style of play and tactics can be used to increase the likelihood of success if one knows more about the relationships between box-score statistics and winning a contest. The fitness level of a team is another potentially important factor to success that a coach can control. Finally, factors that are outside of a coach's control on the training field that can positively and negatively affect a team's performance can be used and minimized respectively if they are addressed when necessary. Something as simple as uniform color can affect the judgment of the performance in close matches. This could be applied to identifying recruits and could be a factor in officiating matches. Additionally, the manner in which a team travels to away games can be an important factor in gaining an advantage for away games.

CHAPTER III

METHODS

The study examines and compares relationships among specific performance indicators and winning versus losing soccer matches in Landmark Conference women's Soccer. The indicators were selected from a broader group of game statistics collected by the NCAA based on their potential for affecting coaching strategies.

Design

For this study, a causal comparative design was selected. The independent-like variables are the game statistics, including the number of shots, shots on goal, shots on goal percentage, corner kicks, number of fouls committed, and game venue (home versus away). The dependent variable is the game result being a win or loss.

Participants

The analysis is based on regular season games played by seven Division III collegiate women's soccer teams during the 2011, 2012, and 2013 seasons. Only the 118 games with a winner and loser were included. Tied games were not included in the analysis. The number of games by each team ranged from 15 to 19, attributable to Landmark Conference playoff games added in some cases to regular season competition.

Instrument

The researcher used measures in the form of box score statistics from previously contested Landmark Conference Women's soccer games.

Procedure

For each Landmark Conference women's soccer game, the home team provides a game manager for the purpose of tallying statistics in the following categories: number of

shots, saves, corner kicks, and number of fouls committed. The researcher calculated shots on goal by adding the goals scored for a team to the number of saves by the opposition goalie(s).

An independent samples median test was used for each of the following factors: number of shots taken, shots on goal, shots on goal percentage, corner kicks, and the number of fouls committed.

Additionally, a stepwise regression analysis was used to determine whether variables entered after the best predictor contributed anything additional to the predictive power of the model.

CHAPTER IV

RESULTS

Four key strategic actions performed by each team in each game were analyzed to show central tendency (mean and median), dispersion (standard deviation), and range. Results of those analyses are shown in Table 1.

Table 1

Key Strategic Actions

Variable	Mean	SD	Median	Range
Shots taken	12.25	6.25	10.50	2-29
Shots on goal	6.00	3.69	6.00	0-15
Fouls Committed	6.84	4.44	7.00	0-21
Percent shots on goal	48.67	20.84	48.05	0-100

Next, independent samples median tests were performed to determine whether differences in performance medians related to winning versus losing are significant. Results are shown in Table 2.

Table 2

Hypothesis Test Summary

Null Hypothesis	Sig.	Decision
Medians of Shots Taken are the same across Winning-Losing Status	.000	Reject null hypothesis
Medians of Fouls Committed are the same	.582	Do not reject null hypothesis
Medians of Shots on Goal are the same	.000	Reject null hypothesis
Medians of Percent Shots on Goal are the same	.8555	Do not reject null hypothesis

As Table 2 indicates, only two of the four indicators examined, shots taken and number of shots on goal, are significantly related to winning versus losing a game.

Further analysis was performed to determine whether either of these indicators independently predicted winning versus losing. A discriminant analysis, which helps to

determine which variables discriminate between two or more groups, in this case winning versus losing a game, was used for that purpose.

Table 3

Discriminant Analysis Results

Step		F to Enter
0	Shots Taken	47.641
	Fouls	.262
	Shots on Goal	30.117
	Percent Shots on Goal	.048
1	Fouls	.039
	Shots on Goal	.042
	Percent Shots on Goal	.001

As Table 3 indicates, a stepwise discriminant analysis was performed to identify which actions best discriminate between winning versus losing a game. In this analysis, the variable that best predicts winning versus losing is examined first. Once that variable's contribution to the prediction is accounted for, then the next-best predictor is entered, etc. Initially, both shots taken and shots on goal were identified as significant predictors. However, once the best predictor, shots taken, was taken into account, none of the others added any additional information to improve the prediction.

Table 4 shows the relationships between the discriminant function, which consists of the single predictor, shots taken, and each of the other variables.

Table 4

Structure Matrix

	Function 1
Shots taken	1.000
Fouls	.108
Shots on goal	.773
Percent shots on goal	.026

Notice that shots taken and shots on goal are correlated and predictive of winning versus losing. The other two actions, fouls and percent shots on goal, are not. Therefore, the null hypothesis was retained for fouls and percent shots on goal and rejected for the other two actions.

CHAPTER V

DISCUSSION

The null hypothesis was retained for fouls and percent shots on goal and rejected for shots taken and shots on goal.

Implications of Results

The most important implication from the results is that taking more shots is predictive of winning versus losing. Effectively, attacking aggression with regard to taking shots is important even without accuracy. Additionally, physical aggression, in the form of aggressive play resulting in fouls is not a predictor of winning versus losing.

Theoretical Consequences

Despite the demonstrated importance of taking more shots, a team may not put all of its effort into shooting, as many other aspects of a match will affect shooting rates. For example, better passing is likely to result in a team possessing the ball near the opponent's goal to create a shooting opportunity. Also, defending is key to preventing shooting opportunities for the opposition.

Threats to the Validity

Because of the complexity of factors that may affect the result of a soccer match, internal threats to validity include the impact of factors that were not examined in the box score statistics. These include the effects of the quality and quantity of passing for each team, player movement without the ball, player positioning, team formation, fatigue, and morale to name a few. While these statistics are used for NCAA record keeping, the nature of the collection by a range of game-day managers from different schools make

the possibility of inaccuracies a concern, especially in interpreting if a shot is on goal or not on goal from the sideline view of the game.

As this study is generalized to Landmark Conference women's soccer and results were consistent for three seasons of competition, external threats to validity are not present.

Connections to Previous Studies/Existing Literature

Carmichael et al. (2000) examined statistics that are comparable to this study, in addition to many others, that may affect the outcome of a soccer competition. An interesting difference in the finding of the current study is the statistical significance of fouls given away as a determinant of match success. Carmichael et al.'s study analyzed these factors at the professional level in which players are much more capable of taking advantage of shooting/set-play opportunities close to the opponents' goals. It is not surprising Carmichael et al. had different findings compared to the current study's examination of Division III women's soccer.

Implications for Future Research

The importance of shooting quantity is an unsurprising result. However, the lack of importance of shooting accuracy, in terms of percent shots on goal, is surprising. This points to the need for future research on the factors that affect the volume of shooting opportunities generated within a match. For example, such factors that are within the realm of reasonable statistical analysis, given the resources typically afforded to Division III women's soccer teams, include passing accuracy and team possession. An additional factor that may be important to the performance of a single team with Landmark Conference women's soccer competition may be the timing of goals scored and

conceded. For example, is a team more likely to score at certain times of a game, either just after a team has scored, or in the last 10 minutes of a half? The answer to this type of question could point to the importance of greater focus during certain periods, or to the need for better fitness.

Conclusions

The importance of this study in demonstrating the statistical insignificance of most statistics compiled for NCAA records shows the need for a more careful examination of game analysis for women's soccer in the Landmark Conference. Teams cannot rely on current statistics to demonstrate successful and unsuccessful performances. As a result, teams need to use qualitative data in terms of team performance or develop alternative quantitative analyses of team performances in order to train for success.

References

Blavatskyy, P. R. (2010). Contest success function with the possibility of a draw: Axiomatization. *Journal of Mathematical Economics, 46*(2), 267–276.

Carmichael, F., Thomas, D., & Ward, R. (2000). Team performance: The case of English premiership football. *Managerial and Decision Economics*, *21*(1), 31–45.

da Silva, R., Vainstein, M. H., Lamb, L. C., & Prado, S. D. (2013). A simple non-Markovian computational model of the statistics of soccer leagues: Emergence and scaling effects. *Computer Physics Communications, 184*(3), 661–670.

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

Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting, 21*(2), 331–340.

Gordon, R. A. (2008). Attributional style and athletic performance: Strategic optimism and defensive pessimism. *Psychology of Sport and Exercise, 9*(3), 336–350.

Hagemann, N., Strauss, B., & Leissing, J. (2008). When the referee sees red. *Psychological Science*, *19*(8), 769–771.

Heuer, A., & Rubner, O., (2009). Fitness, chance, and myths: An objective view on soccer results. *The European Physical Journal*, *B67*, 445–477.

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

Min, B., Kim, J., Choe, C., Eom, H., & (Bob) McKay, R. I. (2008). A compound

framework for sports results prediction: A football case study. *Knowledge-Based Systems, 21*(7), 551–562.

Moschini, G. (2004). Nash equilibrium in strictly competitive games: Live play in soccer. *Economics Letters, 85*(3), 365–371.

Oberhofer, H., Philippovich, T., & Winner, H. (2010). Distance matters in away games: Evidence from the German football league. *Journal of Economic Psychology, 31*(2), 200–211.

Rampinini, E., Impellizzeri, F., Castagna, C., Coutts, A., & Wisloff, U. (2009). Technical performance during soccer matches of the Italian serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, *12*, 277–233.