

An xG of Their Own: Using Expected Goals to Explore the Analytical Shortcomings of Misapplied Gender Schemas in Football

Sachin Narayanan and N. David Pifer

Florida State University, Tallahassee, FL, USA

Although professional women's football has benefitted from recent surges in popularity, challenges to progress and distinguish the sport persist. The gender-schema theory explains the tendency for individuals to hold female sports to male standards, a phenomenon that leads to negative outcomes in areas such as media representation and consumer perception. One area in which schemas have a more discreet effect is player and team performance, where the assumption that technical metrics developed in men's football are transferable to women's football remains unfounded. Using expected goals, a metric synonymous with the probability of a shot being scored, we highlight how variables important to shot quality and shot execution differ across gender, and how attempts to evaluate female footballers with models built on men's data increase estimation errors. These results have theoretical and practical implications for the role they play in reframing schemas and improving the methods used to evaluate performance in women's sports.

Keywords: performance analytics, machine learning, player evaluation, soccer

The development of women's professional football, known as "soccer," in some parts of the world, has been laden with challenges. With issues ranging from wage disparities at the international level (Bass, 2022) to stadium bans lasting almost half a century (The FA, 2022), the sport's efforts to grow have been persistently stunted by limited publicity, negative stereotypes, and various forms of gender discrimination. However, the sport that was once labeled as being "quite unsuitable for females" by a prominent European football association has seen its popularity rise in recent years as the number of competitions and consumers increased (The FA, 2022). In Europe, for example, the number of women's national teams and youth national teams rose by nearly 35% in a 4-year span, jumping from a total of 173 in 2013 to 233 in 2017 (UEFA, 2019). Around this time, the number of training academies for women doubled (Van Lange et al., 2018), and recent estimates suggest close to 40 million women and girls participate in organized football around the globe (Pedersen et al., 2019). Further growth was evident at the 2019 FIFA Women's World Cup in France, where an average live television and streaming audience of 17.27 million marked a 106% increase in viewership over the 2015 FIFA Women's World Cup in Canada. The final match between the United States and Netherlands garnered 82.8 million views (FIFA, 2019), eclipsing viewership of the 2015 final by 56% and setting a record as the most-watched women's football match in broadcast history.


Accordingly, an increased number of job opportunities have coincided with the continued progress of professional women's football. Many of the same roles traditionally available to sport managers in the men's settings have become available to aspiring managers in women's football. The managers (coaches), scouts, and data analysts composing a club's technical staff, for instance, now regularly appear in both settings as they play important roles in shaping the product on the pitch. As the individuals responsible for evaluating the opposition,

developing effective tactical strategies, and improving the physical and technical qualities of the players, their tasks are becoming increasingly important to professional football clubs. However, while much has been observed and recorded as it relates to the performances of men's footballers and teams (Anzer & Bauer, 2021; Lucey et al., 2015; Pollard & Reep, 1997), relatively little is known about the technical skills and tactical styles unique to the women's game.

To this end, there is an inclination to transfer knowledge and techniques across genders with limited validation of their effectiveness. The gender-schema theory, which suggests people are inclined to process information based on sex-linked associations and gender-based constructs (Bem, 1981), presents a rationale for the tendency to frame male athletes, actions, and competitions as the ideal standard (Clément-Guillotin & Fontayne, 2011). Performance analysts, even those operating in successful women's football organizations, have admitted to haphazardly applying training methods and analytical tools (e.g., expected goals [xG] models) developed in men's football to the women's game (Mitchell et al., 2022); furthermore, the number of male coaches present in high-level women's football (McLoughlin, 2021), as well as the number of male coaches using women's jobs as springboards for managerial careers in men's football (Rampling, 2020), likely magnifies the propensity to implement a "one-gender-fits all" approach to player and team performance. Even if football is unique among most professional sports in that the men and women adhere to the same rules and regulations (Ashworth, 2020), research noting clear physiological differences and potential variations in technical performance suggests further investigation is needed. Although physical differences between male and female athletes have been clearly identified in prior literature (Bradley et al., 2014; de Araújo et al., 2020; Pedersen, 1997; Perroni et al., 2018), studies highlighting differences in technical performance and skill-based measures remain in relatively short supply (Bransen & Davis, 2021; Pappalardo et al., 2021).

In football, xG is a metric used to quantify the probability of a shot resulting in a goal based on a variety of factors. Because the

Pifer  <https://orcid.org/0000-0001-9523-9417>

Narayanan (sachin10dev@gmail.com) is corresponding author,  <https://orcid.org/0000-0002-6404-2424>

primary objective in football is to score goals, many other analytical metrics have been built backward from xG, allowing it to become a key measure of performance in the sport (Goodman, 2018). Reliable statistical models can help identify variables that are critical to shot quality (xG) and shot execution (post-shot expected goals [PSxG]) in different football contexts. Given the increased availability and quality of data in women's football, there now exists an opportunity to estimate these statistics in their proper context. Further analysis is warranted due to the potential for work in this area to reframe stereotypes, help managers prescribe appropriate training regimens, and improve the precision of analysts' evaluations. Therefore, the purpose of this study was to explore whether analytical frameworks established in one (gender) schema are directly transferable to another. More specifically, we trained and tested gradient-boosted women's and men's xG and PSxG models on 28,942 shots provided by StatsBomb, an industry-leading data provider, to answer the following research questions (RQs):

RQ1. Which performance variables are most important to shot quality and shot execution in professional women's football, and how do they differ from those in professional men's football?

RQ2. Are xG and PSxG estimates susceptible to increased error when women's (men's) football data are supplied to models built from men's (women's) data, and are there any notable trends in these deviations?

From a theoretical perspective, the answers to these questions highlight the potential shortcomings of misapplied gender schemas, prompting individuals to recall schemas more appropriate to a specific context. In practice, training sessions and player evaluations may benefit from insights that are more accurate and tailored to the unique traits of female players. If estimates of xG and PSxG vary across gender, sport managers working as analysts, scouts, and coaches would be wise to consider the resulting implications.

Literature Review

Expected Goals

Such as many professional sports in the 21st Century, football has embraced the use of data analytics for such purposes as player scouting and recruitment, performance evaluation, and tactical development, leading to an increase in the number of technical staff members being employed by clubs and teams around the world (Anderson & Sally, 2013; Pifer et al. 2018). Spurred by innovations in player tracking technology and motivations to remain competitive in settings where success and financial clout are highly correlated (Hall et al., 2002), the modern football organization collects and analyzes big data with the goal of creating and sustaining competitive advantages. The technical staff tasked with overseeing these efforts therefore plays an increasingly important role in developing metrics that shape team strategy and complement the decisions of upper management. In this regard, xG, which is defined as the probability of a shot resulting in a goal, has become one of football's primary measures of player and team performance.

Though many variations exist, each containing subtle differences in the types of data, predictor variables, or classification methods being used, xG models attempt to quantify a shot's likelihood of ending up in the back of the net based on factors observed at the time of the shot. Because goals are rare events in football, even a solitary strike can have a big impact on the outcome of a match (Anzer & Bauer, 2021). Pollard and Reep (1997) were two of the first to codify xG in scientific

research, using a hand-charted collection of 489 shots from the 1986 FIFA (Men's) World Cup to quantify the probability of a shot being scored based on variables related to shot location (distance and angle to the goal), whether the shot was taken on the first touch, the proximity of the nearest defender, and whether the possession preceding the shot originated from open play or a set piece (freekick). Using logistic regression, the pair was able to estimate the significance and size of each effect on the outcome, finding that chances taken more centrally and closer to the goal had a higher probability of being scored.

More recently, xG models have been constructed using larger, more detailed data and applied to scenarios and strategies of increasing complexity. Lucey et al. (2015), for example, combined tracking data related to nearly 10,000 shots with a conditional random field model to analyze how actions occurring during the 10-s window preceding a shot impacted goal probability. Results showed that variables related to the game phase (e.g., corner kick, freekick, counterattack); defender proximity; the interactions of surrounding players' speed of play; and shot location were all important determinants of shot outcomes in men's football. Anzer and Bauer (2021) extended work in this area by using nearly 106,000 shots from multiple seasons of the German Bundesliga to train an extreme gradient boosting (XGBoost) algorithm that—by their measure—was more accurate than the xG models published in prior literature. The private tracking data they acquired allowed them to incorporate such variables as the distance and angle of the shot, the speed of the player taking the shot, the number of defenders in the line of the shot, the positioning of the goalkeeper and pressure on the shooter at the time of the shot, the part of the body used to take the shot, how the shooter gained control of the ball before shooting, and whether the shot was the result of a freekick. In line with earlier analyses, they found distance from the goal to be the most important determinant of shot outcomes, with defender proximity and the speed of the shooter also playing prominent roles. Such as Lucey et al. (2015), this study highlighted the practical and predictive advantages of supplying granular tracking data to xG models.

It is not surprising, then, that the data providers—companies such as Chyronhego, Stats Perform (formerly known as Opta), and StatsBomb—have also contributed to the steady stream of xG-related research. As those tasked with developing and selling proprietary versions of xG models to their clients (e.g., professional clubs and university teams), analysts at these companies have helped refine the metric and entrench its use in the practical sector. One advancement in this arena has been the development of post-shot xG (PSxG). Whereas traditional xG incorporates all shot outcomes and constituent factors at the moment of contact between the ball and the part of the body striking it, PSxG condenses the sample to on-target shots (e.g., saves and goals) and considers what happens to the ball after the shot is taken. By including variables related to the trajectories and velocities of on-target shots, PSxG allows for more intuitive quantifications of finishing skill and save difficulty for efforts placed on goal (Goodman, 2018; Vatvani, 2022). This metric, given its novelty and general privatization, remains largely unexplored in scientific literature.

Technical and Physical Differences

For all its uses in quantifying chance quality, finishing skill, and shot-stopping ability, xG largely remains a by-product of men's football data. Each study in the reviewed literature, for example, was applied in the context of men's football, resigning applications of xG and PSxG in the women's game to dedicated bloggers and the proprietary work of companies such as StatsBomb. Among academic research, only a limited number of studies have focused

extensively on technical differences in men's and women's football.

One such examination was a white paper presented at a conference workshop by Bransen and Davis (2021). Using private event data provided through a connection with SciSports, the researchers were able to train a series of xG models (generalized additive models and XGBoost algorithms) on shot data from 2,100 professional women's football matches in the United States and Europe and 9,076 men's matches in European club football. The models incorporated predictors related to shot location (distance and angle to goal), body part used for the shot, assist type, and game state (goal difference and time in current half), and were used to examine how variable importance differed across the models for each gender and whether xG models trained on data from one gender were transferable to the other.

Descriptive findings revealed that female footballers had an overall higher shot conversion rate, tended to shoot from locations closer to and with a smaller angle to the goal, headed the ball from positions closer to the goal, and scored more frequently from headers than men. As for the xG-based findings, crossing out-of-sample data to models built for the opposite gender did not noticeably dampen the models' broad-based measures of performance; therefore, given the general transferability, greater focus was placed on specific shot types where xG was markedly different. To this end, the models trained on women's shot data assigned lower xG values to shots taken while cutting inward at the top of the 18-yard box, higher xG values to close range efforts near the goal line, and lower xG values to headers assisted by a set piece. Nonetheless, these results should be taken with caution. White papers are not held to the same research standards as peer-reviewed literature, and in addition to a relatively narrow scope and small number of predictors, the study's replicability is limited by its use of private data.

In an analysis designed to identify more general differences between the observable skills of men's and women's footballers, Pappalardo et al. (2021) examined whether a series of machine learning algorithms, including random forest and boosted models, could properly distinguish between male and female teams on the basis of their technical performances. Incorporating data that quantified the volume of basic football stats (e.g., fouls, passes, shots, and offsides) being observed, the proportion of accurate passes; the speed of the game; the qualities of individual performers; and the collective behavior of teams, they searched for variables that differentiated men's and women's matches at the 2018 FIFA (Men's) World Cup and 2019 FIFA Women's World Cup. Pass accuracy, which was much higher in men's matches, turned out to be the most important differentiator, with recovery time, average time between restarts, pass velocity, and pass length serving as the other relevant variables in models that correctly assigned 93% of out-of-sample matches to the appropriate gender. On average, the women's matches featured significantly fewer fouls, fewer passes, shorter and slower passes, lower pass accuracies, more shots from closer to the goal, and quicker recovery times after the ball was lost. Women were also significantly faster in resuming play when it had to be restarted from a freekick, goal kick, or throw in. A study by Garnica-Capparrós and Memmert (2021) identified similar trends, with a series of classification models showing female players in international matches to perform more—but less accurate—passes, fewer ground duels, and more clearances compared to males.

In summarizing their findings, Pappalardo et al. (2021) labeled the women's game as displaying greater "loyalty" in the sense that

female footballers committed fewer fouls and took less time to resume a match following stoppages. However, they also characterized women's games as being more "fragmented" due to the finding that women's teams exchanged possession of the ball more frequently throughout their matches and displayed less accuracy in their passing. The probable cause for this occurrence was attributed to differences in skill between men's and women's World Cup players, with the former containing more true professionals being paid to specialize in football at the club level. The findings are also linked to research showing that women's teams score more goals and win games by larger margins than men's teams in international tournaments (Sakellaris, 2017). The uneven score lines and higher rates of conversion could indicate that talent in international women's competitions is more concentrated within a subset of elite teams. Pappalardo et al. (2021) were also quick to attribute some of the results to physiological differences that have been widely observed in exercise science literature. The shorter passes and shots taken by the women, for example, were said to be indicative of them having less physical strength in their legs, a finding consistent with the conclusions drawn by research highlighting vivid differences in male and female athletes across the categories of endurance, acceleration, muscular strength, vertical jumping, and stride length (Castagna & Castellini, 2013; Bradley et al., 2014; Bartolomei et al., 2021). Specific to football, Pedersen et al. (2019) found that female footballers have lower muscle mass and must exert more energy in movement, while female goalkeepers are, on average, shorter in height than their male counterparts. Bradley et al. (2014) and Perroni et al. (2018) further found that women typically dribble the ball at a slower pace than men but tend to cover more distance with the ball at these lower speeds.

The Gender-Schema Theory

The notion that women's footballers and many other female athletes suffer from a comparative lack of skill and athleticism has further been at the crux of the messages communicated both subtly and directly by prominent media figures and other members of the professional sport hierarchy (Ross & Shinew, 2008; Lebel & Danylchuk, 2009). Combined with the relative lack of quality and quantity devoted to coverage of women's sports in the media (Bissell & Duke, 2019; Coche & Tuggle, 2016; Lumpkin & Williams, 1991), it is perhaps unsurprising that some of the more observable, male-oriented traits have come to typify the schemas developed by various onlookers and participants in these settings. A schema can be defined as a cognitive structure built through people's experiences and interactions that helps them interpret and understand what is occurring in a particular environment (Baran et al., 2012). Once developed, individuals can recall a particular schema to help process information and navigate issues that later arise from experiences in identical or similar settings; however, despite the efficiency that schemas can lend to estimation and decision making in such environments, they can also contribute to the formation of biases, unrealistic expectations, and stereotypes among their users (Augoustinos et al., 2014).

People can ultimately form schemas around a variety of constructs, including gender, where the gender-schema theory highlights the tendency to process information through sex-linked associations (Bem, 1981). Historically, sports have been heavily dominated by the male demographic in terms of viewership, participation, and management (Adams & Tuggle, 2004). Consequently, many of the decisions and actions that unfold in women's sports are based on mental constructs with heavy links to the male

versions of their games. For example, the assumption that traits associated with men's sports (e.g., pace, power, and precision) are more representative of true athleticism and skill persists among consumers, prominent stakeholders, and even female athletes (Lebel & Danylchuk, 2009; Lobpries et al., 2018). Society, driven in large part by the media, has accordingly classified certain sports as masculine (e.g., football, American football, basketball, hockey, and baseball) or feminine (e.g., individual sports such as gymnastics and figure skating) based on traditionally held views of males and females, forming schemas and stereotypes that are hard to break (Koivula, 2001).

Because many sports have origins in the men's version of the game, and men have historically had more opportunities to specialize in and profit from sport (Toffoletti, 2017; The FA, 2022), male-oriented features tend to dominate the schemas applied to sports such as football. Prior studies have shown that competitive sport contexts activate the masculine dimension of the gender schema (Clément-Guillotin & Fontayne, 2011) and that athletic identity is positively correlated to masculinity (Lantz & Schroeder, 1999). As Birrell (1983) noted, "sport remains highly associated with the so-called 'masculine' elements of our culture, and the female in sport is still considered a woman in man's territory" (p. 49). When male-oriented schemas are reinforced over time, men's sports get highlighted as the primary offering while women's sports are relegated to secondary roles and perceived in a matter-of-fact manner. This tendency, described by some as gender-bland sexism (Musto et al., 2017), positions traits inherent to male athletes and competitions as the standard for comparison.

An illustration of this phenomenon playing out in football took place among a panel of prominent managers at the 2020 Laureus World Sports Awards in Berlin. Posed with a question on what needed to happen for women's football to sustain the momentum of a successful 2019 World Cup, Fabio Capello, former manager of the England men's football team, suggested the women's game change its goal and pitch dimensions to accommodate physical differences. "I think the goal is too big for women and that the pitch is too wide," said Capello. "When they [women] play basketball and volleyball they lower the net because they are not tall such as men. I think the size of the goals makes it really difficult for the keeper, because in football, you have to jump" (Ashworth, 2020, para. 2–3). Capello's assertion was not universally shared by his fellow panelists. "You've got to understand that the men's game has been around for over 100 years," noted former U.S. women's national team manager Jill Ellis, who was also on the panel. "We're seeing taller athletes go into women's soccer. The purity of the game is the purity of the game, and we all love it" (Ashworth, 2020, para. 5).

In line with the tendencies of many other prominent figures in male and female sport (Walker et al., 2022), Capello had employed a schema established in men's football to articulate his views on the development of the women's game. The perception of female athletes (e.g., goalkeepers) as less-athletic and less-skillful became the focus, and the assumption that women's football needed to alter its rules and regulations to meet the same standards of entertainment and performance as the men's game was reemphasized. When occurring in a forum consisting of former coaches, the merits of such assertions are simply debated and discussed, yet when employed among active managers or other individuals in positions of prominence, schemas harness the power to directly influence the operations of a club.

Technical staff, due in part to a historical lack of data and resources in women's football, may be particularly susceptible to

biases arising from misapplied gender schemas. Given the comparative abundance of men's data and the tendencies of prior research and analysis to focus on events taking place in men's football, technical analysts at women's teams—even successful ones—often use xG models trained on men's data to estimate the xG of their female athletes (Mitchell et al., 2022). Such an inclination is likely reinforced in settings where men's coaches manage a disproportionately high number of female teams and use high-profile women's jobs as springboards to similar roles in the men's game. One industry report in professional football showed that, during the 2020–2021 season, 72% of managers across the National Women's Soccer League (United States); Women's Super League (England); FIFA's top-25 ranked women's national teams; and the professional women's club leagues in Germany, France, and Spain were men (McLoughlin, 2021). The opportunistic moves of managers such as John Herdman (Canadian women's national team to Canadian men's national team) and Phil Neville (England women's national team to Major League Soccer's Inter Miami CF) further underscore the propensity to build a resume in women's football while retaining a focus on men's management (Rampling, 2020). These tendencies, which are by no means specific to the sport of professional football (Walker & Bopp, 2011; Ladda, 2015), likely trickle down to analysts and other staff.

Relying on a schema that is heavily biased toward the men's game means these personnel risk overlooking or miscalculating the effects that known physical differences or perceived variations in skill have on key performance indicators such as xG. Because the football industry's current understandings of xG and PSxG are based on male-oriented features, various biases, and questions persist. Are the same variables that impact shot quality and shot execution in the men's game important in the women's game, and do they have similar effects? Is it safe to assume that the same conclusions can be drawn when supplying female performance data to models built from men's data? The answers to these questions are important for the implications they hold locally, for football managers, and globally in relation to the management of female athletes and sports; however, they remain largely unexplored in the extant body of peer-reviewed research.

Methodology

To better isolate the distinguishing variables of performance in women's football (RQ1) and identify the potential shortcomings of employing a men's (women's) schema in the technical analysis of women's (men's) performance data (RQ2), we needed metrics that were relevant to overall performance and heavily utilized among industry professionals. xG and PSxG, as functions of the contextual variables surrounding shots and on-target shots, satisfied these criteria. By following the process of (a) developing xG and PSxG models for each gender, (b) extracting interpretable importance scores for relevant variables, and (c) cross validating the models on differing samples of test data, we could compare the relative importance of key variables toward scoring and whether the estimated scoring probabilities were robust to cross-gender applications. Accordingly, this design improves on the foundational work of Bransen and Davis (2021) by (a) using public event data and transparent statistical methods to enable the important element of replicability (Szymanski, 2020), (b) analyzing a wider variety of predictor variables to allow for more precise estimates and interpretations across a range of scenarios, and (c) including PSxG as an additional outcome variable to present a more holistic view of certain variables' effects on goal-scoring across gender.

Data Source and Data Collection

A large quantity of quality event data related to shots and shot-preceding-actions were needed to create reliable xG and PSxG models and validate their use across different contexts. To this end, StatsBomb, one of the industry leaders in the collection and analysis of football performance data, served as our primary source. The company combines state-of-the-art camera and pitch detection technologies with the technical expertise of its employees to record an average of 3,400 events per match (StatsBomb, 2022). Following a quality assurance process that validates the collected data for accuracy and consistency, StatsBomb disseminates the resulting information and insights to its subscribed clientele. Fortunately, the company also makes sizable portions of its data freely available to the public as JSON (JavaScript Object Notation) files through a GitHub repository. These samples are useful for research because (a) they cover a wide range of competitive matches in professional men's and women's football, (b) they are collected by the same operator, thereby limiting the measurement error that can arise when merging data from separate sources (Anzer & Bauer, 2021; Garnica-Caparrós & Memmert, 2021), (c) they undergo a quality control process and adhere to the expectations of clients who are leaders in the football industry, and (d) they are publicly available, meaning analyses stemming from these data can be replicated and used to build on the existing body of knowledge (Szymanski, 2020). StatsBomb also has packages in Python and R (e.g., StatsBombPy and StatsBombR) that make the steps of importing the files and converting and cleaning the data more manageable.

Using StatsBombR in R (version 3.3.0), we pulled 84,538 rows of event data from StatsBomb's online repository. These data represented 1,242 matches in nine competitions across professional women's (four) and men's (five) football. Of these events, 28,942 were shots (10,923 from women and 18,019 from men), and 10,420 (3,796 for women and 6,624 for men) were on-target shots. Table 1 lists all events and seasons included in StatsBomb's free data at the time of collection. In its records of shooting events, StatsBomb uses a method known as "Freeze Frames" to capture detailed information on the shooter (e.g., distance to the goal, footedness, and angle to the goal), the surrounding environment at the time of the shot (e.g., distance of the nearest defender and position of the goal-keeper), and the outcome of the shot (e.g., goal, miss, save, or block). Moreover, the company charts events outside of shots, allowing researchers who chronologically arrange the data to observe and record the type of action (e.g., through ball, cross, or regular pass) preceding each shot. This attaches granular variables supported in prior xG research to each shot observation (Anzer & Bauer, 2021; Lucey et al., 2015), yielding an enriched data set spanning a diversity of matches and competitions.

Table 1 Football Competitions and Seasons Represented in StatsBomb's Event Data

Women's Football	Men's Football
FA Women's Super League (2018–2021)	UEFA Champions League (1999/2000, 2003–2019)
National Women's Soccer League (2018)	English Premier League (Arsenal, 2003/2004)
FIFA Women's World Cup (2019)	FIFA Men's World Cup (2018)
UEFA Women's Euros (2022)	La Liga (Barcelona, 2004–2021)
	UEFA Men's Euros (2020)

Model Variables

Following data collection, each shot was labeled as a goal ("1") or nongoal ("0"), resulting in a binary outcome variable to which relevant variables could be fit in the subsequent classification models. The data were filtered to only include open-play shots that occurred during normal phases of play rather than shots taken directly from set pieces or penalty kicks. The predictive variables used in the models were largely derived from the reviewed literature and industry best practices. They are listed and appropriately defined in Table 2. While most of these variables were directly measured or labeled by StatsBomb, we calculated a few more (e.g., *Ball Receipt Speed* and *Length of Prior Event*) to proxy for potential factors not recorded in the original data. The final three variables listed in Table 2—*Average Velocity*, *Shot End Location Y* (lateral, goal width); and *Shot End Location Z* (vertical, goal height)—are exclusive to the PSxG models that only include samples of on-target shots (i.e., goals and saves by the keeper). Finally, to meet the data formatting requirements of the chosen method, the one-hot encoding process was used to convert the levels of each categorical variable into binary values that indicated the presence ("1") or absence ("0") of that level from the shot observation.

Model Specifications

The shot observations from each gender were split into training (80%) and testing sets (20%) using a balancing procedure that ensured the natural distribution of goal outcomes remained consistent in each set. Then, the training data were used to build and parameterize a series of extreme gradient boosting (XGBoost) classification models for xG and PSxG in men's and women's football. XGBoost is a supervised machine learning algorithm involving decision trees and ensemble learning. Tree-based methods are prevalent among tasks involving the classification of an outcome where the training data provided to the model are gradually split into numerous branches at decision points called nodes. These trees grow until every observation in the data set can be grouped with the most commonly occurring observations from that class, ultimately ending in terminal nodes that do not split further (James et al., 2021). The nature of ensemble learning is that multiple tree models are being combined to improve the overall predictive power of a single model. XGBoost, as a type of ensemble learner, utilizes "boosting" to iteratively grow new trees using information from the previously grown trees (James et al., 2021). Therefore, in addition to the general benefits of tree-based models handling variable nonlinearities and interactions more efficiently, XGBoost models learn from the errors of the prior models and use the boosting process to subsequently generate more accurate estimates. This element helps XGBoost, and other boosted models outperform many of the more traditional machine learning classifiers such as CART (classification and regression tree) and random forest models (James et al., 2021). Accordingly, many of the xG models recently developed by industry leaders and academic experts have relied on XGBoost or similar methods (Anzer & Bauer, 2021; Bransen & Davis, 2021; Vatvani, 2022).

Like other tree-based algorithms, XGBoost has hyperparameters that can be tuned to avoid overfitting to the training data and ensure optimal performance on out-of-sample data. As values that help configure machine learning models' abilities to learn, hyperparameters must be set before a model is built and adjusted to improve subsequent performance. According to Yu and Zhu (2020), three of the most important hyperparameters controlling

XGBoost model performance are the learning rate (eta), maximum tree depth (max depth), and overall model complexity (gamma). The eta parameter controls the rate at which trees are added to the model by ensuring the improvements offered by new trees meet a certain threshold. The max depth, or interaction depth, controls how deep each individual tree grows in proportion to the number of variables and observations in the data. Gamma is similar to max depth except it seeks to reduce overfitting across the entire ensemble by instructing trees to only split and grow branches when estimation is improved by some amount (Chen & Guestrin, 2016).

To tune the models, we subjected the training data to a 10-fold (20% of training data in the folds) cross-validation procedure that sought to minimize log loss (the negative average of the log of corrected, predicted probabilities that indicates the proximity of predicted probabilities to their corresponding true values) across a

grid containing different combinations of eta, max depth, and gamma. We used log loss because it is slightly more robust to imbalanced classes than traditional measures of classification performance (e.g., accuracy) and because it was a more direct measure of the xG and PSxG probabilities we were estimating. In the samples used to build the xG models for women (men), just 10.5% (12.1%) of shots resulted in a goal, justifying our use of log loss as the primary performance measure. In addition, to account for the imbalance in the outcome classes and generate truer probabilities, an isotonic probability calibration was performed on the test data predictions for all models using the Generalized Pool-Adjacent-Violators Algorithm. Table 3 shows the optimal hyperparameters selected for each model and their respective log loss values following the tunes. Comparisons to models with default parameters, and additional performance metrics for each model (e.g., accuracy and area under the ROC curve [AUC]), are also reported.

Table 2 List and Descriptions of Predictive Variables Used in the xG and PSxG Models

Feature	Definition	Type
Distance to Goal	Distance, in yards, between shot location and goal center.	Numeric
Angle to Goal	Angle, in degrees, between shot location, goal center, and right sideline.	Numeric
Distance to Keeper	Distance, in yards, between the goalkeeper and the goal center.	Numeric
Angle to Keeper	Angle, in degrees, made by goalkeeper, goal center, and right sideline.	Numeric
Defenders In Cone	Number of defenders present in the conical area between the shot and goal posts.	Integer
Density In Cone	Aggregate inverse distance for each defender behind the ball in the shot cone.	Numeric
Distance to D1 (D2)	Distance, in yards, between shooter and nearest (second nearest) defender.	Numeric
Shot First Time	Indicates whether the shot was taken on the shooter's first touch.	Factor
Receipt to Shot Time	Time taken by shooting player from ball receipt to shot.	Integer
Distance Traveled	Distance traveled by the shooting player from ball reception to shot execution.	Numeric
Ball Receipt Type	Type of prior event leading to the shot (normal pass, through ball, cross, and cut back).	Factor
Length of Prior Event	Distance covered by the ball from the prior event to reception by the shot taker.	Numeric
Ball Receipt Speed	The speed the ball traveled to the shot taker when released by the prior event.	Numeric
Dominant Foot	Indicates whether the shot was taken with the player's dominant foot.	Binary
Open Goal	StatsBomb measure indicating whether the shot was in front of an open goal.	Binary
Under Pressure	StatsBomb measure indicating whether a defender pressured the shot taker.	Binary
Shot Technique Type	Shot technique used (half volley, volley, overhead, lob, backheel, header, and normal).	Factor
Average Velocity	The average velocity of the shot in yards per second.	Numeric
Shot End Location Y	Measure of horizontal shot location for on-target shots.	Integer
Shot End Location Z	Measure of vertical shot height for on-target shots.	Integer

Note. xG = expected goals; PSxG = Post-shot expected goals.

Table 3 Performance Metrics of Models With Tuning Parameters (Optimized on Log Loss)

Model	Type	Log loss	Accuracy	AUC	Brier score	ETA	Gamma	Max depth
W xG	Optimal	0.2793	0.8965	0.7914	0.0791	.01	5	6
	Default	0.5731	0.8828	0.7148	0.1016	.3	0	6
M xG	Optimal	0.2935	0.8903	0.8102	0.0836	.01	5	6
	Default	0.4871	0.8734	0.7739	0.1044	.3	0	6
W PSxG	Optimal	0.3318	0.8498	0.9176	0.0986	.3	2	10
	Default	0.5276	0.8524	0.9044	0.1199	.3	0	6
M PSxG	Optimal	0.3023	0.8625	0.9356	0.0908	.01	2	10
	Default	0.4631	0.8595	0.9323	0.1137	.3	0	6

Note. xG = expected goals; PSxG = post-shot expected goals; W = women; M = men.

Variable Importance

Following model tuning, RQ1 was addressed by calculating and comparing variable importance scores in the optimal xG and PSxG models for both genders. Important scores help make machine learning algorithms more interpretable and less of a “black box” method by providing information on which variables are being used to split observations into the appropriate classes. Though there are different variations of importance scores, each intending to quantify the relative usefulness of certain variables in generating a model’s predictions, we used *SHapley Additive exPlanations* (SHAP values) to quantify and rank our variables’ contributions toward overall model performance. At the global level, SHAP values are representative of the mean absolute magnitude of each variable’s contribution toward predicting the outcome class (e.g., goal or no goal); therefore, higher mean absolute SHAP values are associated with more influential variables (Yang, 2017).

Because variables’ SHAP values can be calculated globally for their effects on overall model performance and locally to quantify their importance in making predictions on specific instances in a data set, they allow users to observe the magnitude and direction (positive or negative) of variables’ effects across the range of values (e.g., low to high) corresponding to those variables (Yang, 2017). When plotted accordingly, SHAP values provide an intuitive means of analyzing and comparing the effects of the different variables as they relate to estimates of xG and PSxG in men’s and women’s football. To supplement the information conveyed in the SHAP plots, we also calculated the partial effects for each variable and visualized them in a series of partial dependence plots (PDPs). PDPs display the marginal effects of a variable on a model’s outcome across the range of values for that variable, holding all other predictors constant at their average (Greenwell, 2017).

Cross-Validations of Model Performance

After calculating the variable importance scores, the optimal XGBoost models for women’s and men’s xG, and women’s and men’s PSxG, were cross validated on the samples of test data that had been withheld prior to model construction. To fully address RQ2, these validations were performed using the out-of-sample test data from both the matching gender and the differing gender. By analyzing log loss and other measures of model performance, we could see if the estimated probabilities were notably impacted by cross-gender applications. A larger, “gender-blind” model combining all shots from the women’s and men’s training sets was also created and validated on the separate sets and a combined set of test data. This provided an additional view of the scenario and helped control for the potential influences of a larger sample. Finally, following the estimation of model performance differences at a more global level, localized investigations were conducted to identify specific instances where model estimates differed across genders; that is, we analyzed whether models built using men’s data were consistently overestimating or underestimating xG and PSxG for certain types of shots (i.e., shots containing specific values for certain variables) compared to the women’s models.

Results

Variable Importance for the xG Models

The SHAP values ranking variable importance and displaying variable effects for the standard xG models are presented as SHAP plots in Figure 1 (women) and Figure 2 (men). These plots offer a

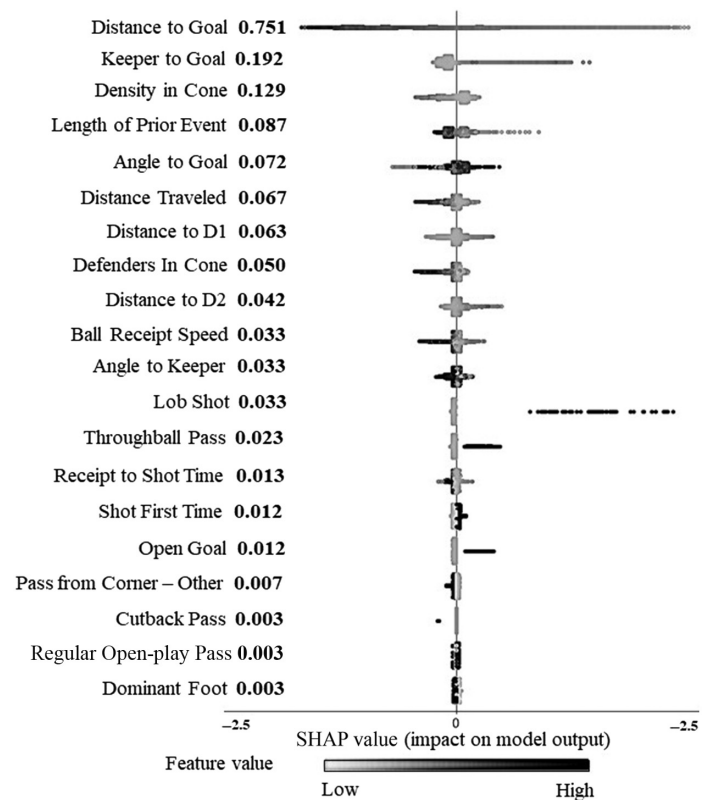


Figure 1 — Women’s xG SHAP plot for 20 most important variables. xG = expected goals; SHAP = *SHapley Additive exPlanations*.

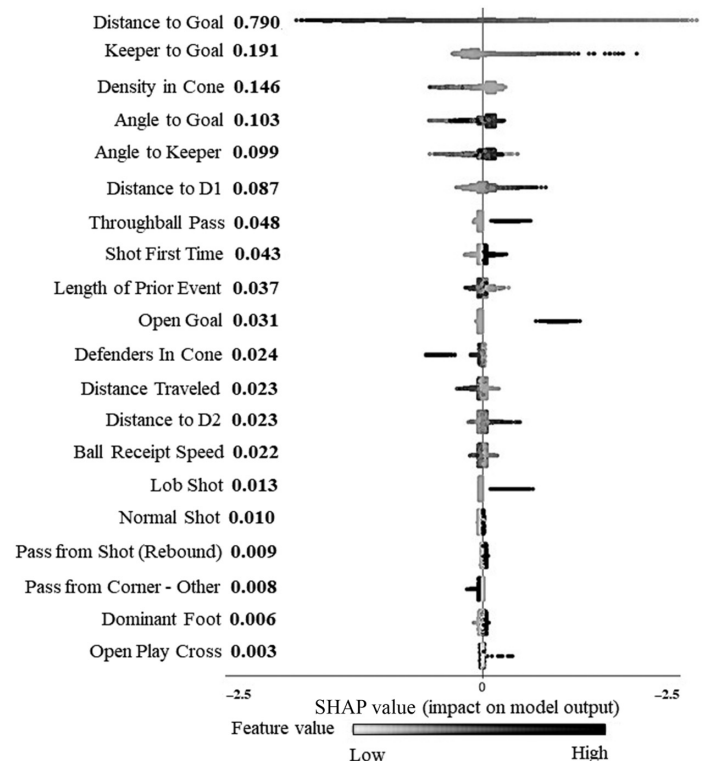


Figure 2 — Men’s xG SHAP plot for 20 most important variables. xG = expected goals; SHAP = *SHapley Additive exPlanations*.

multidimensional view of variable influence by plotting variables' negative or positive impacts on model output along the x -axis and ranking them by their absolute average (global) importance scores on the y -axis. Lower (brighter) and higher (darker) values for a variable are shown in the ranged scale, and the localized estimates of each shot prediction are visibly distributed across the effect continuum (x -axis).

For both genders, *Distance to Goal*, *Keeper to Goal*, and *Density In Cone* were the most important variables for their respective impacts on the xG models' predictions. In addition to identical rankings and similar SHAP values, their expected effects were comparable for different values of those variables. In both men's and women's football, for instance, taking a shot further from goal had a negative impact on its probability of being scored relative to shots taken from closer distances. Likewise, lower quality chances were associated with shots taken from more congested (dense) areas in front of goal. The PDP plots displayed in Figure 3 visually support these relationships, showing how the effects of longer distances and increased density consistently lead to lower xG, all else equal, in both settings.

Even so, some of these primary variables were subject to more discreet variations across gender. The average distance between the goalkeeper and the goal was slightly lower in women's football, and the average women's shot was taken with levels of pressure and congestion that were marginally higher than those observed among men's shots. Table 4 displays descriptive statistics supporting these findings. The PDP for *Keeper to Goal* in Figure 3 also shows that xG is more positively affected by female goalkeepers

coming off their lines over the short to mid ranges than when male goalkeepers assume similar positions. Even so, differences across this element of goalkeeper positioning were marginal compared to those observed in *Angle to Keeper*, a variable that was markedly lower in importance for women's shot outcomes ($\text{SHAP}_{\text{women}} = 0.033$, $\text{SHAP}_{\text{men}} = 0.099$). Its respective PDP in Figure 3 implies that, outside of shots taken from rare, extremely narrow angles to the keeper, the isolated effect of this variable is larger in men's football at most values.

The *Length of Prior Event* ($\text{SHAP}_{\text{women}} = 0.087$, $\text{SHAP}_{\text{men}} = 0.037$) and *Ball Receipt Speed* ($\text{SHAP}_{\text{women}} = 0.033$, $\text{SHAP}_{\text{men}} = 0.022$) variables also displayed notably different importance scores and respective impacts at high and low values. All else equal, women's shots had lower xG values if they received the ball at higher speeds, and they seemed less likely to score following longer passes. The PDP for *Ball Receipt Speed* reiterates that men's xG is more positively impacted at higher ball speeds than women's xG. Further distinctions were observed among variables related to the positions of surrounding defenders at the time of a shot. The number of defenders in the conical area between the shot and the goalposts (*Defenders In Cone*) and the distance to the second nearest defender (*Distance to D2*) both ranked higher in importance in the women's xG model. Combined with the means presented in Table 4, these results suggest female footballers took shots in more congested spaces, all else equal. Data for the shot locations provide some evidence for why, showing that women in the sample took a higher percentage of

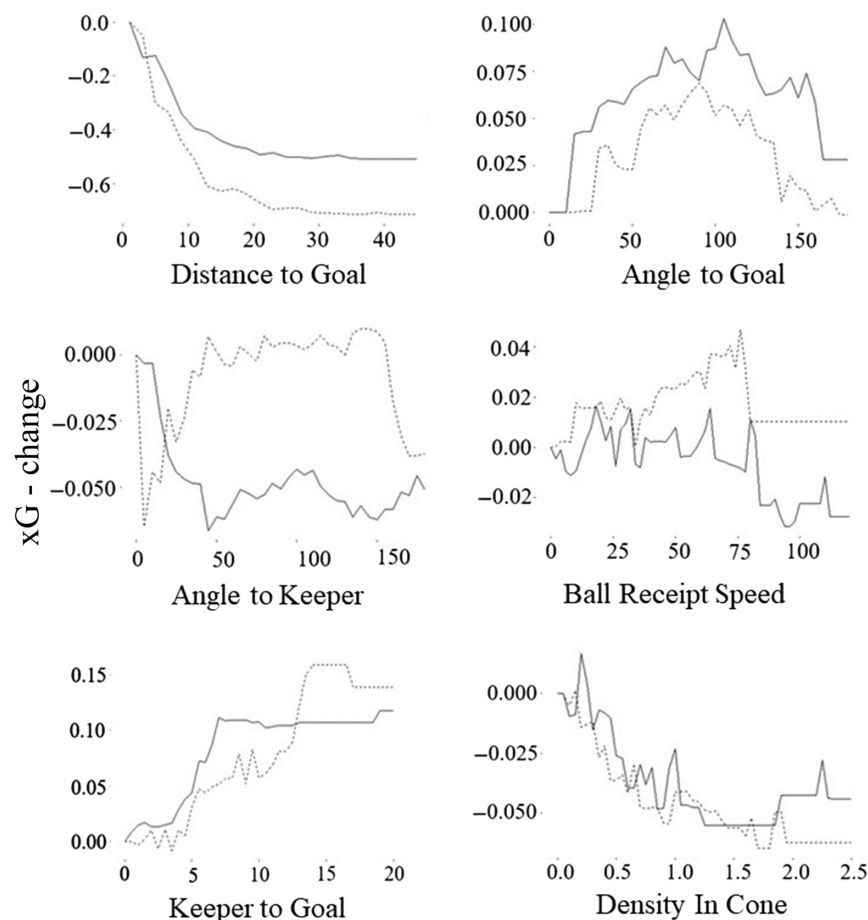


Figure 3 — Partial dependence plots for important xG variables. *Note.* Women's line is solid; men's line is dashed. xG = expected goals.

Table 4 Descriptive Statistics for Select xG and PSxG Variables

Parameter	Mean		SD		Median		First quartile		Third quartile	
	Women	Men	Women	Men	Women	Men	Women	Men	Women	Men
Distance to Goal	18.9	18.59	8.786	8.522	17.2	17.62	11.14	11.99	24.35	24.33
Angle to Goal	90.47	88.32	33.65	35.12	90.00	88.315	64.65	60.33	116.3	115.429
Keeper to Goal	2.966	3.424	2.697	2.662	2.302	2.832	1.526	1.884	3.468	4.101
Angle to Keeper	95.60	89.91	47.29	46.34	91.01	90.00	55.59	48.58	135.00	130.43
Defenders In Cone	1.117	0.8602	1.241	0.981	1.00	1.000	0.000	0.000	2.000	1.000
Density In Cone	0.3297	0.2737	0.514	0.445	0.1549	0.1162	0.000	0.000	0.4469	0.3661
Distance to D1	2.691	2.781	1.958	1.878	2.236	2.267	1.414	1.414	3.493	3.606
Distance to D2	5.369	5.537	3.142	3.133	4.827	4.825	3.162	3.220	6.943	7.201
Time from Receipt to Shot	4.03	3.556	8.745	7.743	1.000	1.000	1.000	1.000	3.000	3.000
Distance Traveled	13.724	12.511	17.352	15.07	6.891	6.576	2.484	2.640	18.237	16.621
Length of Prior Event	34.37	31.67	23.31	21.77	29.43	25.50	14.38	14.07	54.01	45.51
Ball Receipt Speed	19.626	19.294	16.455	15.63	15.418	15.403	7.642	7.846	27.214	27.007
Shot Velocity	17.756	20.56	10.385	11.06	16.865	19.67	10.825	12.91	22.028	25.96
Shot End Location Y	40.05	39.91	6.561	7.07	40.00	39.90	36.40	36.15	43.70	43.70
Shot End Location Z	1.228	1.272	1.586	1.76	0.50	0.40	0.00	0.00	1.80	1.80

Note. xG = expected goals; PSxG = Post-shot expected goals.

shots from close range areas more central to the goal while men took more shots from the edge of the 18-yard box and from positions wide of the 6-yard box.

In terms of shooting techniques, the SHAP plots showed that lobbed shots (*Lob Shot*) had a larger effect on scoring in the women's game and that their effect on xG was overwhelmingly positive compared to the men ($SHAP_{women} = 0.033$, $SHAP_{men} = 0.013$). The importance of first-time shots was similar in both ($SHAP_{women} = 0.012$, $SHAP_{men} = 0.043$) and generally led to positive outcomes, while shooting with the dominant foot was slightly more important and more frequently associated with positive outcomes in men's football ($SHAP_{women} = 0.003$, $SHAP_{men} = 0.006$). Shots labeled as being taken on an open goal also appeared to be less impactful in women's football ($SHAP_{women} = 0.012$, $SHAP_{men} = 0.031$). For the different types of actions preceding a shot, through balls had a slightly larger impact on men's xG but were associated with higher goal probabilities and ranked as the most important prior event in both settings ($SHAP_{women} = 0.023$, $SHAP_{men} = 0.048$). For women's xG, open-play crosses were the second most important preceding event, but regular, open-play passes were the second most important event in men's xG. All else equal, crosses have a faintly more negative effect on xG in women's football. The series of PDPs contained in Figure 4 depict these differences and lend further support to the effects witnessed in the SHAP plots.

Variable Importance for the Post-Shot xG Models

The SHAP values associated with variables in the PSxG models trained and tuned by on-target shots are displayed in Figure 5 (women) and Figure 6 (men). Here, differences in variable importance were particularly evident among the predictors related to shot speed and shot trajectory that are exclusive to PSxG. In women's football, the lateral end location (*Shot End Location Y*) of an on-target shot appeared to be less important ($SHAP_{women} = 0.769$, $SHAP_{men} = 1.196$), but the vertical end location (height) of an on-target shot (*Shot End Location Z*) was slightly more important ($SHAP_{women} = 0.197$, $SHAP_{men} = 0.155$).

This coincided with the greater importance of lobbed shots in women's football, a variable that retained its positive impact in the women's PSxG model. The *Average Velocity* of a shot also boasted a higher SHAP value while ranking higher in women's football ($SHAP_{women} = 0.300$, $SHAP_{men} = 0.233$). The mean values in Table 4 show the average velocity of female shots to be nearly three yards per second slower than men. This gap, which increased to almost 4 yards per second in the upper quartiles, indicates that female footballers tend to shoot the ball with less intensity and speed, all else equal. Figure 7 contains the PDPs for these three variables.

Among the remaining variables, *Distance to Goal*, *Keeper to Goal*, *Density In Cone*, and *Angle to Goal* displayed similarly important effects across gender as those seen in the regular xG models. *Keeper to Goal* ($SHAP_{women} = 0.359$, $SHAP_{men} = 0.510$) and *Angle to Keeper* ($SHAP_{women} = 0.195$, $SHAP_{men} = 0.253$) also exhibited the same differences that had been observed in the prior models, retaining lower importance scores in women's PSxG. In terms of overall importance, the goalkeeper positioning variables had an expectedly higher impact on PSxG than they did on xG. Conversely, the *Density In Cone*, *Defenders In Cone*, and *Distance to D2* variables dropped in importance in the PSxG models, likely due to a higher proportion of shots taken in congested areas not ending up on target. This might also explain the lesser importance of the *Ball Receipt Speed* and *Length of Prior Event* variables in the PSxG models when compared to the xG models. Essentially, these variables impacted whether a shot would be on target in the first place, but once on target, they did not matter as much.

However, while many circumstantial factors, shot techniques, and preceding actions mattered far less to the outcomes of shots, and particularly men's shots, once the shots were known to be on target, there were a few exceptions. First-time shots became more important when placed on target in both men's and women's football, likely signaling this type of shot's ability to catch goalkeepers flat-footed when accurately executed. Similarly, the importance of the *Open Goal* variable—much like the goalkeeper positioning variables—logically rose in importance from xG to PSxG, and the gender differences observed for this variable in xG were less apparent.

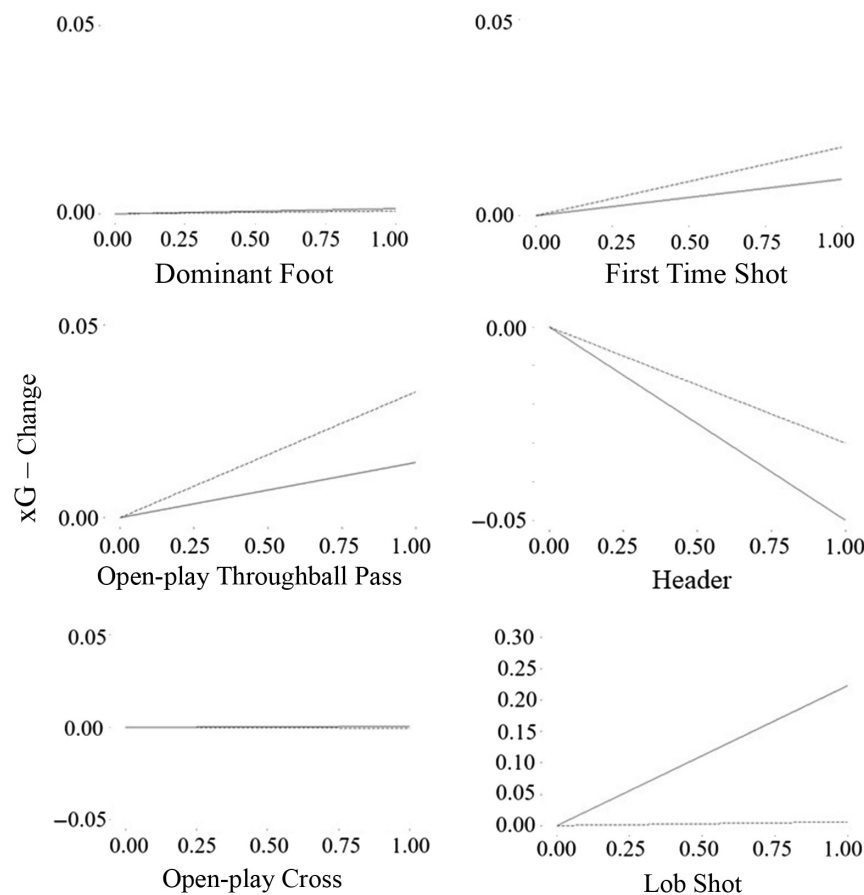


Figure 4 — Partial dependence plots for xG variables related to preceding events and shot types. *Note.* Women's line is solid; men's line is dashed. xG = expected goals.

Validating Model Performance

Table 5 shows performance metrics for the xG and PSxG models cross validated on the out-of-sample test data from both genders. It also displays performance metrics for the gender-blind models trained and tested on the gender-specific samples and combined samples of shot data from both settings. In terms of the xG models' global performances on the full sets of test data, the resulting log loss values (where lower values are associated with predicted probabilities that are closer to the actual outcomes) indicated the strongest performing model on any set of test data was the women's model validated on the women's data (0.2793). While fitting given the motivations of the study, it is worth noting this log loss was only marginally improved over the log loss that resulted from the predictions of the men's model on that same set of women's test data (0.2802); similarly, when the out-of-sample men's data were supplied to the men's (0.2935) and women's models (0.2973), the resulting log loss values were nearly identical. The lack of discernment between the models is further emphasized by the additional performance measures, with each combination of model and data achieving top scores in either the Brier Score, accuracy, or AUC category. In addition, using an xG model created from and validated on the combined data offered no major improvements in predictability over the gender-specific models. This combined xG model was also independently tested on separate men's and women's out-of-sample data to see whether it offered improved performance over the gender-specific models. However, the

combined model had higher log losses for the women's (0.2942) and men's (0.2939) data than their respective, gender-specific models.

For PSxG, the men's model validated on the men's data produced a better (lower) log loss (0.3023) than the women's model validated on women's data (0.3318) and the combined model (0.3085). Compared to the xG models, performance differences when crossing genders between the PSxG models were more evident, particularly when the men's test data were supplied to the men's model (0.3023) over the women's model (0.3667). A lower Brier Score, and higher AUC reaffirmed the strength of the men's PSxG model at estimating the probabilities of on-target men's shots. However, PSxG predictions for the out-of-sample women's data did not appear to be globally impacted by the choice of model, with the resulting log loss values only displaying minor improvements when predicted by the women's model (0.3318) versus the men's (0.3368). Additional performance metrics such as Brier Score, accuracy, and AUC were in similar ranges across each combination of model and test data and offered no further indications of key differences in predictability. The combined PSxG model was also tested independently with out-of-sample men's and women's data, but their respective log losses were only marginally lower than the gender-specific models. All told, there was little separation among the models, particularly the xG models, when it came to making valid, global predictions on differing sets of test data. Rather, it appeared the models were capable of being broadly applied to shot data from the opposite gender, with the various

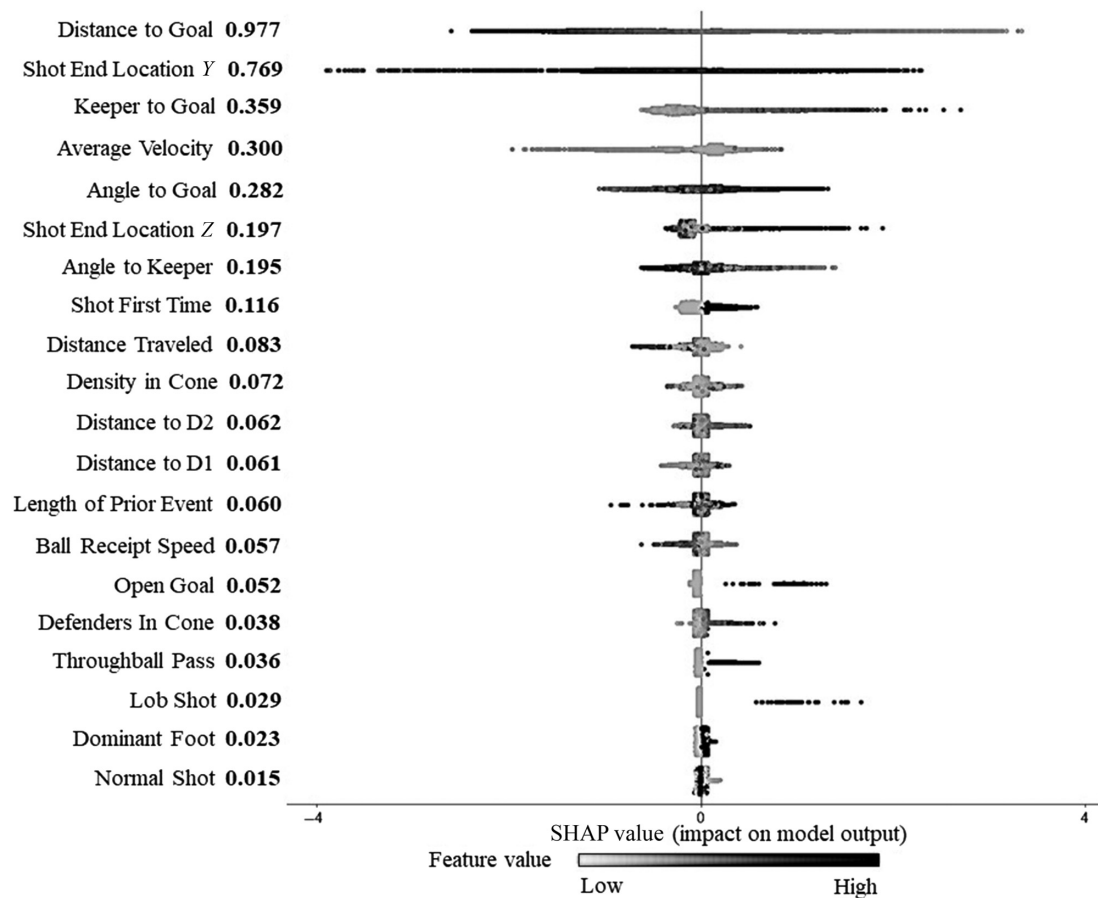


Figure 5 — Women's PSxG SHAP plot for 20 most important variables. PSxG = post-shot expected goals; SHAP = *SHapley Additive exPlanations*.

predictive measures appearing largely unphased by cross-gender applications.

Estimation Biases in Cross-Gender Applications

Minimal differences in global model performance on test data from the opposite gender prompted further exploration of specific, localized contexts in which model predictions might diverge. We started by analyzing the variables that were highly important to xG predictions for both genders (*Distance to Goal*, *Angle to Goal*, *Density In Cone*, and *Keeper to Goal*) and generated local predictions in the test data for each value of these variables. Figure 8 displays the results of the women's test data supplied to the men's xG model, with the upper half of each figure showing the extent to which the women's model tended to predict higher (positive bars) or lower (negative bars) xG values compared to the estimates of the men's model across each value of a variable. The lower half of each figure shows the distribution of values for that variable in the women's test sample. All else equal, higher (lower) bars indicate that if the women's (men's) model was used, it predicted a higher (lower) probability of the women's shot becoming a goal compared to that same datapoint being predicted by the men's (women's) model.

Starting with *Distance to Goal*, results suggest that the men's model estimates higher xG values for women's shots taken closer to the goal compared to the women's model, with an average

difference of approximately 8%. However, right around the 6-yard box, this relationship was briefly reversed as the women's model predicted a higher xG. Further increases in distance to the goal yielded milder fluctuations in the xG differences, but at distances beyond 25 yards, the women's model predicted consistently higher xG values. Next, *Angle to Goal* shows how, at extremely narrow angles close to 0° and 180°, the men's model predicted a higher xG than the women's model for the same shot data. In addition, the women's model predicted a higher xG on the left side of the pitch as indicated by the positive spikes between 10° and 25°; nonetheless, shot angle produced a similar xG in the most commonly occurring (central) areas of the distribution. Continuing, *Density In Cone* did not show much variation across the men's and women's models at its most frequent values, but clearer trends were observed in higher values of *Keeper to Goal*. When the goalkeeper was much closer to the goal, both models had similar xG values, but in situations where the keepers were further off their line (likely rushing out to meet a shot-taker), the women's model predicted a higher likelihood of goal-scoring with xG values that were typically 5%–10% higher than those predicted by the men's model.

Similar analyses of key PSxG variables such as *Distance to Goal*, *Shot End Location Y*, *Shot End Location Z*, and *Average Shot Velocity* are shown in Figure 9. At very close distances (five yards or less), the men's model predicted higher PSxG values than the women's model for the women's test data. Between six and 12 yards, the trend reversed, and the women's model predicted higher PSxG. The men's model again predicted higher PSxG nearing the

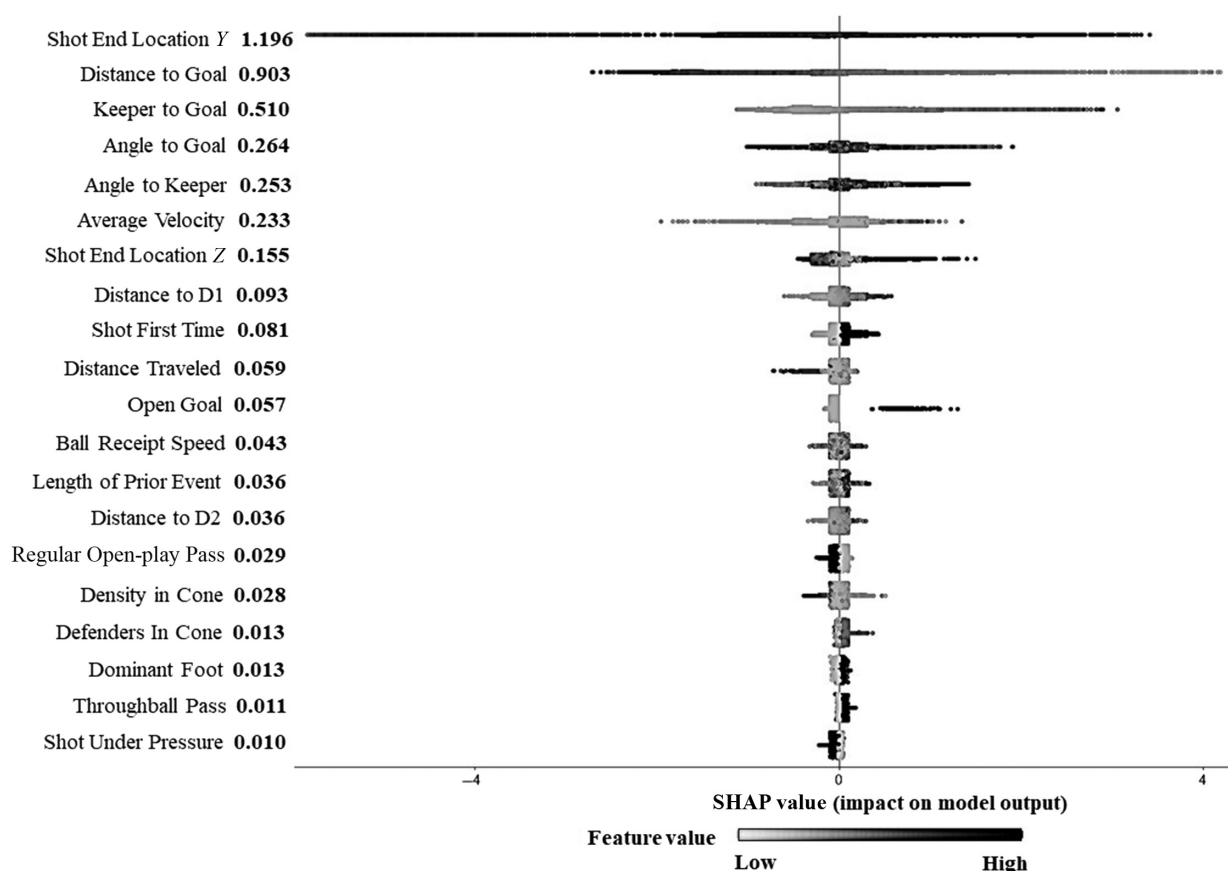


Figure 6 — Men's PSxG SHAP plot for 20 most important variables. PSxG = post-shot expected goals; SHAP = *SHapley Additive exPlanations*.



Figure 7 — Partial dependence plots for PSxG variables. *Note.* Women's line is solid; men's line is dashed. PSxG = post-shot expected goals.

18-yard box and beyond, though PSxG spiked in the women's model for a small sample of shots taken between 30 and 35 yards. Shifting to the trajectories of on-target women's shots, the women's model tended to produce higher PSxG estimates for shots along the ground, shots placed higher on goal (*Shot End Location Z*), and those placed in specific lateral areas of the goal (*Shot End Location Y*). For example, within 1.2 yards of the keeper's right post, the women's model produced a PSxG estimate that was approximately 5% higher than the value estimated by the men's model. For shots placed closer to the left post, the men's model tended to predict higher PSxG values, all else equal. Finishing with the *Average Velocity* of the shots, the PSxG estimates from the women's model were, on

average, 8%–10% higher than the men's model for shots frequently taken at lower velocities. However, once the recorded velocity moved beyond the average velocity for a women's shot (17.756), the men's model predicted higher PSxG values.

Discussion

"The biggest problem resulting from following a venerated tradition and hardened dogma is that they are rarely questioned. Knowledge remains static while the game itself and the world around it changes" (Anderson & Sally, 2013, p. 2). This quote,

Table 5 Results for xG and PSxG Models Validated on Women's (W), Men's (M), and Combined Data

Model/data	Log loss	Brier score	Accuracy	AUC
W xG/W	0.2793	0.0791	0.8965	0.7914
W xG/M	0.2973	0.0783	0.8900	0.8032
M xG/M	0.2935	0.0836	0.8903	0.8102
M xG/W	0.2802	0.0846	0.9011	0.7871
xG/Combined	0.2949	0.1001	0.8931	0.7806
xG Combined/W	0.2942	0.0858	0.8902	0.7738
xG Combined/M	0.2939	0.0838	0.8944	0.7951
W PSxG/W	0.3318	0.0986	0.8498	0.9176
W PSxG/M	0.3667	0.0946	0.8248	0.9117
M PSxG/M	0.3023	0.0908	0.8625	0.9356
M PSxG/W	0.3368	0.1083	0.8668	0.9210
PSxG/Combined	0.3085	0.1158	0.8594	0.9329
PSxG Combined/W	0.3281	0.0939	0.8399	0.9025
PSxG Combined/M	0.2986	0.0902	0.8632	0.9316

Note. PSxG = post-shot expected goals; xG = expected goals; AUC = area under the ROC curve.

which was used to frame the general need for data analytics in football, neatly captures the purpose of our study and the impact of our results. Because male-oriented frameworks traditionally serve as the schemas in many sports, including football, women's sports risk being overlooked or misinterpreted in certain capacities. The results of our investigations highlight the presence of this risk in technical analyses of player and team performance by showing how knowledge produced in one setting may not directly translate to another. In exploring RQ2, global estimates of xG and PSxG trained on data from one gender were found to vary little when tested on out-of-sample data from the other. On the surface, this gives the illusion that models can be applied across gender without any noticeable dips in predictive performance. However, in our more localized investigations of RQ1 and RQ2, clear differences were observed among specific variables that impact goal probability. This highlights the hidden dangers of drawing inferences for women's football from estimates based in the men's game and reiterates the need for schemas to be appropriately developed and interpreted. Although global estimates appear robust, localized estimates reveal unique trends across women's and men's football. These findings have notable implications for football managers, scouts, data analysts, and other members of the technical staff.

While variables innate to every open-play shot can lead to relatively consistent estimates of xG and PSxG across gender,

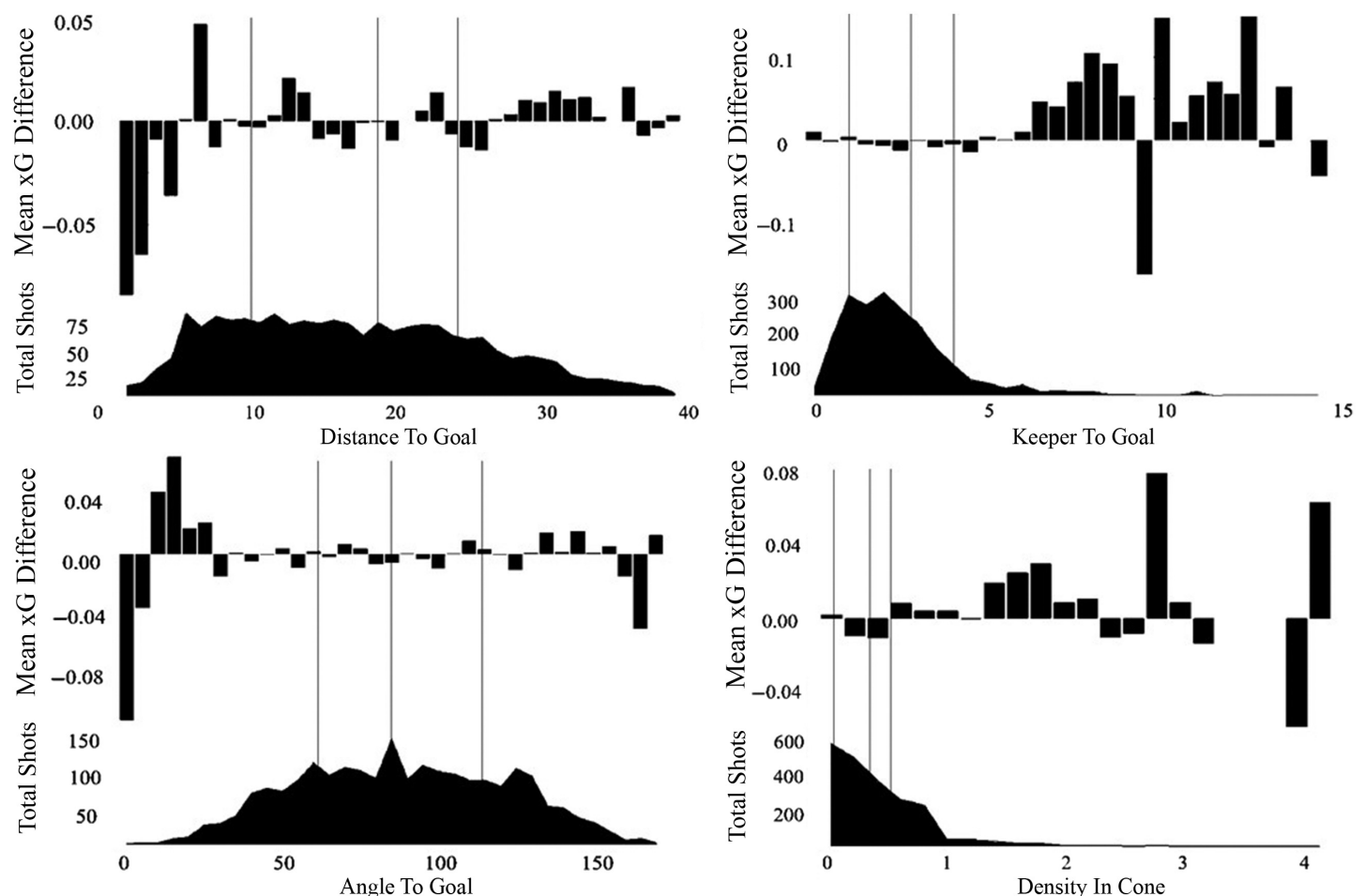


Figure 8 — Mean xG differences in women's data tested on women's and men's models for variation in key model variables. Note. Positive (negative) bars show women's (men's) model predicted higher xG; vertical lines show interquartile range and mean. xG = expected goals.

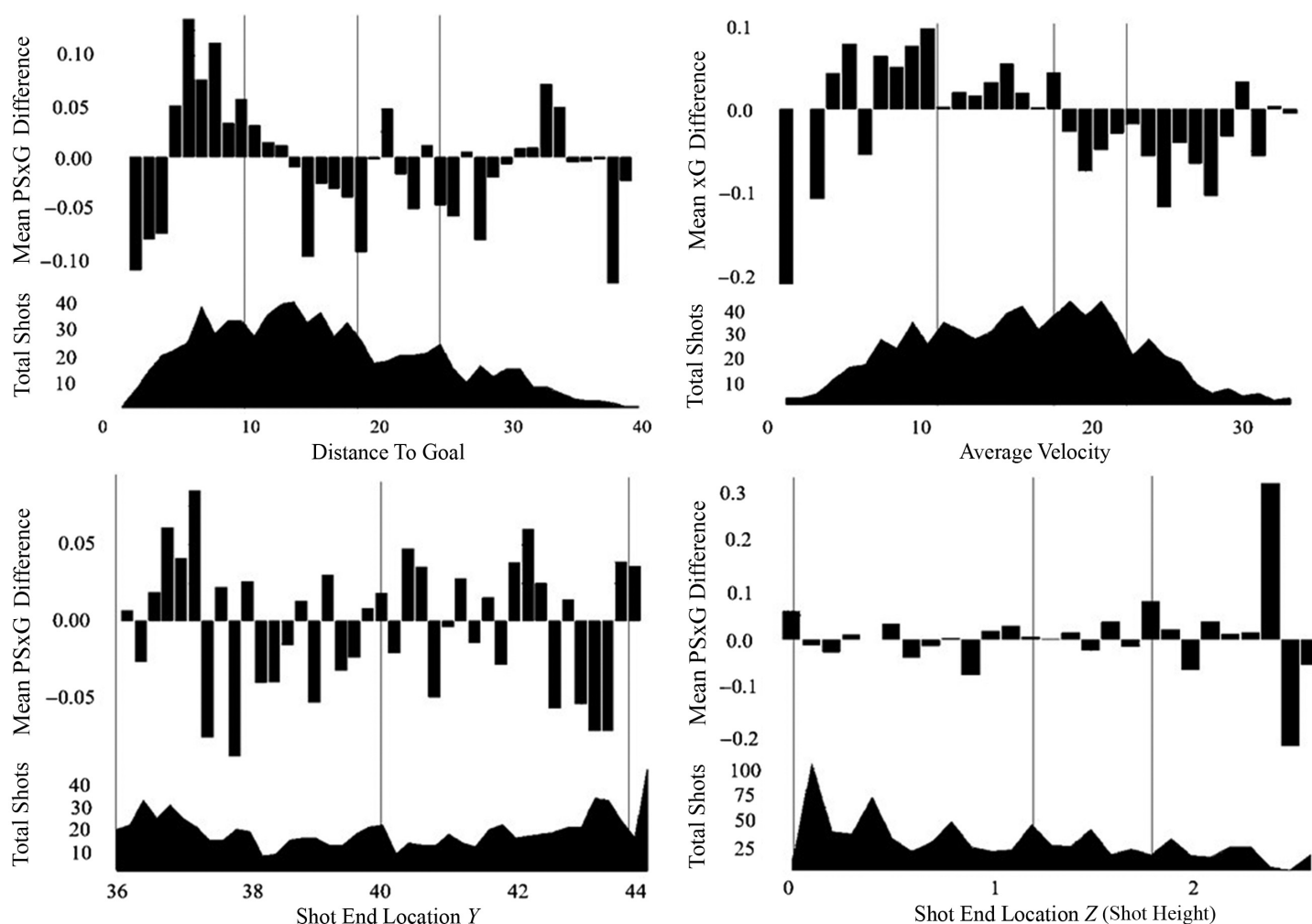


Figure 9 — PSxG differences in women's data tested on women's and men's models for variation in key model variables. *Note.* Positive (negative) bars show women's (men's) model predicted higher PSxG; vertical lines show interquartile range and mean. PSxG = post-shot expected goals.

variables present in more specific contexts affect the predicted xG and PSxG values to varying extents. This is particularly evident when certain types of passes are received, or certain shot techniques are chosen. For example, the speed at which a player received the ball via pass or other preceding event, and the distance the ball traveled from the prior event to the shot location, had more of an impact on xG in women's football. In both variables, higher values (i.e., faster speeds and longer distances) appeared to affect xG more negatively in the women. Further differences were evident when lobbed shots were taken (more positively related to xG in women's football) and in the spaces where females chose to take shots (the women generally took shots in more congestion and under more pressure).

The impact of goalkeeping was also found to vary. Closer shots, higher velocity shots, and shots placed nearer to the posts were all necessary to generate greater, positive changes in men's PSxG. Conversely, women's PSxG was more positively affected as the height of an on-target shot increased, with the models also suggesting that ground shots typically produce higher PSxG estimates. Combined with the greater effectiveness of the lob shot and a higher risk of conceding when far off their line, it seems that female goalkeepers exhibit a different style of goalkeeping in

comparison to their male counterparts, a style which may be further identified through attributes (e.g., ball distribution) not analyzed in this study (Riley, 2023).

Practical Implications

Because this study marked one of the first attempts in scholarly research to build xG models from publicly available women's shot data, and one of the first to explore and validate PSxG in either gender, our findings are relevant to football industry personnel looking to employ xG-based models in their technical operations. Through the use of accessible, high-quality data and appropriate statistical methods, models that accurately estimate the probabilities of shots and on-target shots resulting in goals can be produced and utilized. The resulting xG and PSxG estimates, when aggregated to individual players or entire teams, provide reliable approximations of shot quality and shot execution that can be used to examine a player or team's propensity to take or prevent shots from opportunistic areas of the pitch. PSxG is further useful for analyzing goalkeeping performances because it considers the pace and placement of the ball during on-target shots. By comparing goals allowed to the PSxG sums they faced, goalkeepers' shot-stopping

abilities can be quantified more appropriately. Offensively, finishing ability can be better distinguished from shot quality by comparing goals scored to sums of xG and PSxG.

In addition to their predicted probabilities, the models also highlight the key indicators of goal scoring. Knowing the precise location of a shot, for example, is vital to determining the overall quality of the chance. A shot's distance from the goal and its angle to the goal are highly important variables in any context, and these markers regularly have the largest effects on shot outcomes. Information pertaining to the location of the goalkeeper is also important. A quick reproduction of our models without variables related to goalkeeper positioning revealed observable dips in predictive performance when validated on the test data. Within both the women's and men's xG models, log loss increased by 0.04 and 0.03 and accuracy fell by over 2% and 3%, respectively, when goalkeeping variables were excluded.

Shifting from the more universal implications to the gender-based differences explored in this study, the findings should help persuade women's coaches, analysts, and other constituents to derive their xG and PSxG estimates from models tailored to the women's game. For example, when a women's PSxG model is used to assign probabilities to women's shot data, we see that increasingly higher shot velocities and shots placed closer to the posts do not affect PSxG at the same rate as the men's models. While the patterns are similar, the resulting effects are not as pronounced, suggesting that, all else equal, increased pace and the wider placement of shots have less discernible effects on women's shot outcomes compared to men. This is conversely echoed by the finding that shots placed high on target tend to produce higher PSxG values in the women's models. In terms of regular xG, higher ball receipt speeds and the further positioning of keepers from the goal present higher rates of change in the women's predictions relative to the men, in the negative and positive directions, respectively. All else equal, increases in the distance of a shot from the goal also have a weaker, less-adverse effect on women's xG.

These findings therefore hold implications for managers who could use training sessions to recreate scenarios and rehearse techniques that produce higher probability chances. For example, the added importance of lobbed shots and efforts placed higher on target could prompt women's football coaches to implement drills that reinforce these tendencies in shot takers. The positive effects of first-time shots and finishing from through balls could also be reinforced through related practice regimens. By comparison, the negligible impact of crosses on resulting shot outcomes suggests that women might create better chances playing through the middle and attempting shots with their feet, emphasizing a more centralized attack. Having players make runs that pull defenders away from the goal mouth and encouraging women's players to shoot quicker could also aid in the quality of opportunities created and their ability to convert shots more effectively.

On the opposite side of the pitch, defensive and goalkeeping coaches could use xG and PSxG-related information to design defensive drills that focus on nullifying the attacking strengths of female footballers. Aggregations of PSxG, for example, offer more reliable estimates of goals that keepers theoretically should have allowed or prevented, providing more appropriate quantifications of shot-stopping ability (Vatvani, 2022). Given the model's suggestions that on-rushing keepers and higher areas of the goal present a higher PSxG for women, coaches could focus on goalkeeper positioning and vertical leaping ability to improve their defensive prowess in these situations. This way, they could cover

the angles and heights more effectively when dealing with on-target shots.

Ultimately, while there is no shortage of training-related applications for xG and PSxG in football, these findings suggest that female footballers and teams should not always be evaluated by the same standards as the men. The specific elements that lend uniqueness to women's football should be considered when technical analyses are conducted, a feature with critical implications for those involved in player scouting and recruitment. While the primary objectives of any football team are still a combination of profit and win maximization (Garcia-del-Barrio & Szymanski, 2006), various clubs at the highest level now focus on need-specific talent recruitment emphasizing multiple dimensions of a player (Gavião et al., 2020). Recruiting the right players is crucial to winning matches, which is in turn related to a team's revenue and financial performance (Barajas et al., 2005; Hall et al., 2002). Furthermore, fans and sport consumers watch a sporting event to see teams perform to their liking, and their demand for an event is strongly linked to a team's actual performance (Walker et al., 2022). Investors, in turn, should have a stronger motivation to commit to a sport that attracts higher consumer demand through increased performance. Because even a moderately incorrect or misguided evaluation of a player can have negative effects on team performance, especially when based on inaccurately trained statistical models or faulty metrics, managers and analysts would be wise to consider the contextual factors surrounding their development and use.

Theoretical Implications

The gender-schema theory suggests sex-linked associations are the filters through which incoming stimuli are processed (Bem, 1981), and competitive sports are known to activate the male dimension of a gender schema (Clément-Guillotin & Fontayne, 2011). In football, industry professionals have admittedly taken a more aschematic approach by applying models grounded in men's data to the women's game (Mitchell et al., 2022), and media members and managers have established the men's version of the sport as the standard for comparison (Ashworth, 2020). This occurs despite prior literature indicating clear physical (Bradley et al., 2014; Perroni et al., 2018) and technical differences (Pappalardo et al., 2021) in male and female athletes. Prior studies on gender schemas in sport have primarily focused on the schema's malleability and applicability across various contexts. Given that prior research has tested and found a "strong and recurrent association between competitive sport and masculine attributes" (Clément-Guillotin & Fontayne, 2011, p. 428), our study went in the direction of examining the consequences of a misapplied gender schema. That is, if male-oriented frameworks are applied to a female sport context, can the same conclusions be drawn?

Our findings suggest that technical personnel and other sport managers involved in player and team performance should be cognizant of gender schemas associated with the sport and acknowledge the inherent differences between women's and men's football. While the effects of certain variables and broader estimates of goal probabilities appear robust to cross-validations on data from the opposite sex, local effects of individual variables and scenarios reveal unique outcomes in men's and women's football that could be overestimated or underestimated if the wrong model is used. Some differences, such as reaction times and shooting strength, are likely due to well-established physiological differences while others are indicative of differences in decision making

and skill execution (Bradley et al., 2014; de Araújo et al., 2020; Pappalardo et al., 2021; Pedersen, 1997; Perroni et al., 2018). The appropriate response among multiple parties, then, should involve a reorientation to the feminine dimension of the gender schema. Rather than embracing gender-aschematic views or forms of gender-bland sexism that ignore or downplay obvious feminine elements in favor of those prevalent in men (Cooky et al., 2021), media members, consumers, and sport managers should note the inherent variations and frame the respective athletes and competitions accordingly. While those taking an aschematic approach might justify their decisions in the name of equality, convenience, or a lack of data availability, further investigation suggests they should not remain ignorant of discreet, yet observable, differences in the performances of male and female footballers. To this end, improvements in the availability, quantity, and quality of women's football data would lend further validity to our findings and allow women's football to continue advancing in the associated areas.

Limitations and Recommendations for Future Research

Moving forward, it is important for researchers to know the limitations of our study and some avenues through which future research might be able to advance on the foundations that were provided. First, while the public event data used in our analyses aid in variable consistency and reproducibility, they still imposed several constraints on the study. To start, even though the number of shots in the data represented a diversity of matches and competitions, it was relatively small compared to the private data collections analyzed in prior research. Although the tradeoff between public and private data was necessary for purposes of replication and transparency—and many of our results were consistent with previous findings—it still highlights the need for additional, high-quality data to be made available to the public. This is particularly true for the women's data, as the number of datapoints available from their samples was considerably smaller than the number available in the men's data. The diversity of the data also presents a few issues of its own. International matches are generally considered to be of lower quality than professional club matches (Pappalardo et al., 2021), and one of the women's seasons was played during a shortened season in stadiums that were empty because of COVID-19 restrictions; conversely, portions of the men's data are from high-performing clubs such as Lionel Messi's Barcelona and Arsenal's undefeated "Invincibles" squad from the 2003 to 2004 English Premier League season. This might make some of the men's results less generalizable to all types of clubs and players while positioning their traits in a more positive light compared to the women. Furthermore, the amount of data available from StatsBomb is not conducive to analyses of performance changes over time. While the factors impacting xG and PSxG might be different now than they were as far back as 2003–2004, there were not enough shot observations within each specific timeframe to test for these differences. As such, some of the findings reported as gender differences in this study could represent general shortcomings of football analytics, and particularly xG models, being applied across varying contexts. Predictive models created and applied in future studies could further test whether the observed differences remain consistent across timeframes, leagues, levels of competition, and other factors.

Continuing, we only focused on xG and PSxG as viable, important measures of football performance. While this study marked a seminal effort to develop a women's xG model and

PSxG models for either gender from public data, shot quality and shot execution metrics represent just a sample of the statistics available to modern football analysts. Newer metrics such as expected threat and expected possession value would allow future research to look beyond shots at the events that help create and prevent them (Fernández et al., 2021). In addition, this study did not analyze differences in the effectiveness of certain playing styles (e.g., formations, defensive organization, pressing schemes, and possession tendencies) or set-piece strategies across women's and men's football. Therefore, future studies could develop a more comprehensive women's football schema that extends beyond shooting and scoring by examining other football details.

Finally, despite women's football representing one of the more popular and prominent women's sports, it is just one competition among a myriad of games played by both men and women. Each sport has its own key performance indicators and measures of skill, and certain ones may not translate as cleanly across gender as xG. Nonetheless, the primary problem outlined by this study (i.e., that men's schemas are more frequently applied to women's sports) and its general method (i.e., using data to identify specific differences between the two versions of a sport) are repeatable in other sports where athlete and team performance are evaluated across similar metrics and rulesets. The negative consequences of the gender-schema theory and gender-bland sexism are by no means unique to football or data analytics; as such, future studies have the opportunity to contribute to the progress of women's sports by further identifying the variables that distinguish them.

Conclusion

This study proceeded from the assumption that technical personnel in women's football tend to employ male schemas in their evaluations of player and team performance, which was grounded in prior literature and industry examples positioning men's sports as the filter through which managers, analysts, media members, and fans process the performances of female players. Using XGBoost machine learning algorithms and reliable, public event data from StatsBomb, we built a series of xG and PSxG models to explore the limitations of employing men's schemas in women's football. More specifically, we identified whether estimates of shot quality (xG) and shot execution (PSxG) were differently influenced by certain variables, and whether these models were robust to cross-gender applications. In addition to serving as seminal attempts to develop replicable xG models in women's football, our findings revealed observable differences in variable importance and highlighted the subliminal dangers of accepting model estimates at face value. When validated on data from the opposite gender, frequently observed variables innate to common shots blurred the lines of model performance. However, variables related to goal-keeper positioning, shot placement, preceding actions, and shot type had varying effects on xG and PSxG in more specific situations, suggesting that women's football clubs and players need to be evaluated according to their unique characteristics and skillsets.

References

- Adams, T., & Tuggle, C.A. (2004). ESPN's SportsCenter and coverage of women's athletics: It's a boys' club. *Mass Communication & Society*, 7(2), 237–248. https://doi.org/10.1207/s15327825mcs0702_6

- Anderson, C., & Sally, D. (2013). *The numbers game: Why everything you know about soccer is wrong*. Penguin.
- Anzer, G., & Bauer, P. (2021). A goal scoring probability model for shots based on synchronized positional and event data in football (soccer). *Frontiers in Sports and Active Living*, 3, Article 624475. <https://doi.org/10.3389/fspor.2021.624475>
- Ashworth, C. (2020). Women's sport: Jill Ellis "respectfully" disagrees with Fabio Capello's suggestion women's pitches and goals should be made smaller. *GiveMeSport*. <https://www.givemesport.com/1547965-womens-sport-jill-ellis-respectfully-disagrees-with-fabio-capellos-suggestion-womens-pitches-and-goals-should-be-made-smaller>
- Augoustinos, M., Walker, I., & Donaghue, N. (2014). *Social cognition: An integrated introduction*. Sage.
- Barajas, A., Fernández-Jardón, C.M., & Crolley, L. (2005). Does sports performance influence revenues and economic results in Spanish football? <https://doi.org/10.2139/ssrn.986365>
- Baran, S.J., Davis, D.K., & Striby, K. (2012). *Mass communication theory: Foundations, ferment, and future*. Cengage Learning.
- Bartolomei, S., Grillone, G., Di Michele, R., & Cortesi, M. (2021). A comparison between male and female athletes in relative strength and power performances. *Journal of Functional Morphology and Kinesiology*, 6(1), Article 17. <https://doi.org/10.3390/jfmk6010017>
- Bass, A. (2022). Opinion: It's time for all professional sports to pay attention to what us soccer just did. *CNN*. Retrieved December 31, 2022, from <https://www.cnn.com/2022/05/18/opinions/us-soccer-agreement-equal-pay-bass/index.html>
- Bem, S.L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, 88, 354–364. <https://doi.org/10.1037/0033-295X.88.4.354>
- Birrell, S. (1983). The psychological dimensions of female athletic participation. In M. Boutilier & L. SanGiovanni (Eds.), *The sporting woman* (pp. 49–91). Human Kinetics.
- Bissell, K.L., & Duke, A.M. (2019). Bump, set, spike: An analysis of commentary and camera angles of women's beach volleyball during the 2004 summer Olympics. In T. Reichert (Ed.), *Investigating the use of sex in media promotion and advertising* (pp. 35–53). Routledge. https://doi.org/10.1300/J057v13n01_04
- Bradley, P.S., Dellal, A., Mohr, M., Castellano, J., & Wilkie, A. (2014). Gender differences in match performance characteristics of soccer players competing in the UEFA Champions League. *Human Movement Science*, 33, 159–171. <https://doi.org/10.1016/j.humov.2013.07.024>
- Bransen, L., & Davis, J. (2021, August 19–26). *Women's football analyzed: Interpretable expected goals models for women* [Paper presentation]. AI for Sports Analytics (AISA) Workshop at IJCAI 2021.
- Castagna, C., & Castellini, E. (2013). Vertical jump performance in Italian male and female national team soccer players. *The Journal of Strength & Conditioning Research*, 27(4), 1156–1161. <https://doi.org/10.1519/JSC.0b013e3182610999>
- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system* [Conference session]. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). New York, NY. <https://doi.org/10.1145/2939672.2939785>
- Clément-Guillotin, C., & Fontayne, P. (2011). Situational malleability of gender schema: The case of the competitive sport context. *Sex Roles*, 64(5), 426–439. <https://doi.org/10.1007/s11199-010-9912-1>
- Coche, R., & Tuggle, C.A. (2016). The women's Olympics? A gender analysis of NBC's coverage of the 2012 London Summer Games. *Electronic News*, 10(2), 121–138. <https://doi.org/10.1177/1931243116647770>
- Cooky, C., Council, L.D., Mears, M.A., & Messner, M.A. (2021). One and done: The long eclipse of women's televised sports, 1989–2019. *Communication & Sport*, 9(3), 347–371.
- de Araújo, M.C., Baumgart, C., Jansen, C.T., Freiwald, J., & Hoppe, M.W. (2020). Sex differences in physical capacities of German Bundesliga soccer players. *The Journal of Strength & Conditioning Research*, 34(8), 2329–2337. <https://doi.org/10.1519/JSC.0000000000002662>
- Fernández, J., Bornn, L., & Cervone, D. (2021). A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. *Machine Learning*, 110(6), 1389–1427.
- FIFA. (2019). 2019: A breakthrough year for women's football. <https://www.fifa.com/tournaments/womens/womensworldcup/france2019/news/2019-a-breakthrough-year-for-women-s-football>
- García-del-Barrio, P., & Szymanski, S. (2006). *Goal! Profit maximization and win maximization in football leagues* (No. 0621), International Association of Sports Economists; North American Association of Sports Economists.
- Garnica-Caparrós, M., & Memmert, D. (2021). Understanding gender differences in professional European football through machine learning interpretability and match actions data. *Scientific Reports*, 11(1), Article 264. <https://doi.org/10.1038/s41598-021-90264-w>
- Gavião, L.O., Sant'Anna, A.P., Alves Lima, G.B., & de Almada Garcia, P.A. (2020). Evaluation of soccer players under the Moneyball concept. *Journal of Sports Sciences*, 38(11–12), 1221–1247. <https://doi.org/10.1080/02640414.2019.1702280>
- Goodman, M. (2018). A new way to measure keepers' shot stopping: Post-shot expected goals. *StatsBomb*. <https://statsbomb.com/articles/soccer/a-new-way-to-measure-keepers-shot-stopping-post-shot-expected-goals/>
- Greenwell, B. (2017). PDP: An R package for constructing partial dependence plots. *The R Journal*, 9(1), 421–436. <https://doi.org/10.32614/RJ-2017-016>
- Hall, S., Szymanski, S., & Zimbalist, A.S. (2002). Testing causality between team performance and payroll: The cases of Major League Baseball and English soccer. *Journal of Sports Economics*, 3(2), 149–168. <https://doi.org/10.1177/152700250200300204>
- James, G., Witten, D., Hastie, T., & Tibshirani, T. (2021). *An introduction to statistical learning: With applications in R*. Springer.
- Koivula, N. (2001). Perceives characteristics of sports categorized as gender-neutral, feminine and masculine. *Journal of Sport Behavior*, 24(4), 377–393.
- Ladda, S. (2015). Where are the female coaches? *Journal of Physical Education, Recreation and Dance*, 86(4), 3–4. <https://doi.org/10.1080/07303084.2015.1009804>
- Lantz, C.D., & Schroeder, P.J. (1999). Endorsement of masculine and feminine gender roles: Differences between participation in and identification with the athletic role. *Journal of Sport Behavior*, 22(4), 545–557.
- Lebel, K., & Danylchuk, K. (2009). Generation Y's perceptions of women's sport in the media. *International Journal of Sport Communication*, 2(2), Article 146. <https://doi.org/10.1123/ijsc.2.2.146>
- Lobpries, J., Bennett, G., & Brison, N. (2018). How I perform is not enough: Exploring branding barriers faced by elite female athletes. *Sport Marketing Quarterly*, 27(1), 5–17. <https://doi.org/10.32731/SMQ.271.032018.01>
- Lucey, P., Bialkowski, A., Monfort, M., Carr, P., & Matthews, I. (2015). *Quality vs quantity: Improved shot prediction in soccer using strategic Variables from spatiotemporal data* [Conference session]. Proceedings of the 8th Annual MIT Sloan Sports Analytics Conference, Pittsburgh, PA.
- Lumpkin, A., & Williams, L.D. (1991). An analysis of sports illustrated ft articles, 1954–1987. *Sociology of Sport Journal*, 8(1), 16–32. <https://doi.org/10.1123/ssj.8.1.16>

- McLoughlin, D. (2021). More female football coaches than ever before, but men still dominate. *Female Coaching Network*. <https://femalecoachingnetwork.com/2021/09/07/more-female-football-coaches-than-ever-before-but-men-still-dominate/>
- Mitchell, J., Brown, A., & Minion, N. (2022). *FSU Soccer Panel*. Florida State University Sports Analytics Summit.
- Musto, M., Cooky, C., & Messner, M.A. (2017). "From fizz to sizzle!" Televised sports news and the production of gender-bland sexism. *Gender & Society*, 31(5), 573–596. <https://doi.org/10.1177/08912432177260>
- Pappalardo, L., Rossi, A., Natilli, M., & Cintia, P. (2021). Explaining the difference between men's and women's football. *PLoS One*, 16(8), Article 255407. <https://doi.org/10.1371/journal.pone.0255407>
- Pedersen, A.V., Akssdal, I.M., & Stalsberg, R. (2019). Scaling demands of soccer according to anthropometric and physiological sex differences: A fairer comparison of men's and women's soccer. *Frontiers in Psychology*, 10, Article 762. <https://doi.org/10.3389/fpsyg.2019.00762>
- Pedersen, D.M. (1997). Perceptions of high-risk sports. *Perceptual and Motor Skills*, 85(2), 756–758. <https://doi.org/10.2466/pms.1997.85.2.756>
- Perroni, F., Pintus, A., Frandino, M., Guidetti, L., & Baldari, C. (2018). Relationship among repeated sprint ability, chronological age, and puberty in young soccer players. *The Journal of Strength & Conditioning Research*, 32(2), 364–371. <https://doi.org/10.1519/JSC>
- Pifer, N.D., Wang, Y., Scremin, G., Pitts, B.G., & Zhang, J.J. (2018). Contemporary global football industry: An introduction. In B.G. Pitts & J.J. Zhang (Eds.), *The global football industry* (pp. 3–35). Routledge. <https://doi.org/10.4324/9781351117982-1>
- Pollard, R., & Reep, C. (1997). Measuring the effectiveness of playing strategies at soccer. *Journal of the Royal Statistical Society: Series D*, 46(4), 541–550. <https://doi.org/10.1111/1467-9884.00108>
- Ramplig, A. (2020). The England women's job deserves to be more than the launch pad Phil Neville used it for. *90 Min*. <https://www.90min.com/posts/the-england-women-s-job-deserves-to-be-more-than-the-launch-pad-phil-neville-used-it-for-01e9dy8j2n8b>
- Riley, P. (2023). The unique (and not so unique) challenges of goalkeeping in women's soccer. *StatsBomb*. Retrieved April 8, 2023, from <https://statsbomb.com/articles/soccer/the-unique-and-not-so-unique-challenges-of-goalkeeping-in-womens-soccer/>
- Ross, S.R., & Shinew, K.J. (2008). Perspectives of women college athletes on sport and gender. *Sex Roles*, 58(1), 40–57. <https://doi.org/10.1007/s11199-007-9275-4>
- Sakellaris, D. (2017). The in-game comparison between male and female footballers. *Statathon*. Retrieved December 15, 2022, from https://statathon.com/author/sakellaris_dim_c8tnj45k/page/3/
- StatsBomb. (2022). *Free data: What we do*. Retrieved December 16, 2022, from <https://statsbomb.com/what-we-do/hub/free-data/>
- Szymanski, S. (2020). Sport analytics: Science or alchemy? *Kinesiology Review*, 9(1), 57–63.
- The FA. (2022). *The history of women's football in England*. <https://www.thefa.com/womens-girls-football/history>
- Toffoletti, K. (2017). *Women sport fans: Identification, participation, representation*. Routledge.
- UEFA. (2019). *Women's football*. <https://www.uefa.com/insideuefa/football-development/womens-football/>
- Van Lange, P.A., Manesi, Z., Meershoek, R.W., Yuan, M., Dong, M., & Van Doesum, N.J. (2018). Do male and female soccer players differ in helping? A study on prosocial behavior among young players. *PLoS One*, 13(12), Article 209168. <https://doi.org/10.1371/journal.pone.0209168>
- Vatvani, D. (2022). Upgrading expected goals. *StatsBomb*. <https://statsbomb.com/articles/soccer/upgrading-expected-goals/>
- Walker, N., Allred, T., & Berri, D. (2022). Could more dunking really help the WNBA? *International Journal of Sport Finance*, 17(4), 187–200. <https://doi.org/10.32731/IJSF/174.112022.01>
- Walker, N.A., & Bopp, T. (2011). The underrepresentation of women in the male-dominated sport workplace: Perspectives of female coaches. *Journal of Workplace Rights*, 15(1), 47–64. <https://doi.org/10.2190/WR.15.1.d>
- Yang, L. (2017). SHAP for XGBoost in R: SHAPforxgboost. *Text on Image*. Retrieved January 1, 2023, from <https://liuyanguu.github.io/post/2019/07/18/visualization-of-shap-for-xgboost/>
- Yu, T., & Zhu, H. (2020). *Hyper-parameter optimization: A review of algorithms and applications*. <https://doi.org/10.48550/arXiv.2003.05689>