

Research Article

Michael A. Lapré* and Elizabeth M. Palazzolo

Quantifying the impact of imbalanced groups in FIFA Women's World Cup tournaments 1991–2019

<https://doi.org/10.1515/jqas-2021-0052>

Received June 21, 2021; accepted September 30, 2022;
published online October 17, 2022

Abstract: The FIFA Women's World Cup tournament consists of a group stage and a knockout stage. We identify several issues that create competitive imbalance in the group stage. We use match data from all Women's World Cup tournaments from 1991 through 2019 to empirically assess competitive imbalance across groups in each World Cup. Using least squares, we determine ratings for all teams. For each team, we average the ratings of the opponents in the group to calculate group opponents rating. We find that the range in group opponents rating varies between 2.5 and 4.5 goals indicating substantial competitive imbalance. We use logistic regression to quantify the impact of imbalance on the probability of success in the Women's World Cup. Specifically, our estimates show that one goal less in group opponents rating can increase the probability of reaching the quarterfinal by 33%. We discuss several policy recommendations to reduce competitive imbalance at the Women's World Cup.

Keywords: balance; fairness; FIFA Women's World Cup; least squares; logistic regression; rating sports teams.

1 Introduction

Soccer is the most popular sport in the world. The sport's world governing body Fédération Internationale de Football Association (FIFA) has more members than the United Nations (Haan, Koning, and van Witteloostuijn 2007). Since 1991, FIFA organizes the Women's World Cup in every four years. The 2015 World Cup in Canada reached 764 million viewers worldwide (FIFA 2015b). More than 1 billion viewers watched the 2019 World Cup in France. In

2019, FIFA estimated the total number of females playing organized soccer at 13.36 million and the total number of registered female players at 4 million globally (FIFA 2019c). Despite the tremendous popularity of the Women's World Cup and continued growth in the total number of female soccer players, sports analytics papers studying the structure of the World Cup tend to focus on the Men's World Cup (see, e.g., Jones 1990, Rathgeber and Rathgeber 2007, Scarf and Yusof 2011, Groll, Schauburger, and Tutz 2015, Guyon 2015, Stone and Rod 2016, Laliena and López 2019, Cea et al. 2020, Guyon 2020, Stronka 2020, Chater et al. 2021, Csató 2022a).¹ Given the lack of research around the Women's World Cup, we analyze the structure of the Women's World Cup in this paper.

The World Cup, the pinnacle of women's soccer, consists of a qualification phase and a tournament phase. In the qualification phase, teams compete within their continental confederation to earn participation in the tournament phase. There are six continental confederations: North and Central America (CONCACAF), South America (CONMEBOL), Europe (UEFA), Africa (CAF), Asia (AFC), and Oceania (OFC). Prior to each qualification phase, FIFA determines the number of teams from each confederation that will advance to the tournament phase.²

The World Cup tournament phase consists of two stages – a group stage and a knockout stage. In the group stage, teams are allocated to groups of four teams each. In each group, the four teams play a round-robin tournament

*Corresponding author: Michael A. Lapré, Owen Graduate School of Management, Vanderbilt University, Nashville, TN, USA, E-mail: m.lapre@vanderbilt.edu. <https://orcid.org/0000-0003-2259-8739>

Elizabeth M. Palazzolo, Lazard, New York, NY, USA

¹ A notable exception is Groll et al. (2019). The authors use a hybrid machine learning approach to predict winning probabilities for all teams in the FIFA Women's World Cup 2019. In contrast, we are not concerned with prediction. Instead, we retrospectively assess competitive imbalance in Women's World Cup tournaments.

² Occasionally, FIFA has allocated 0.5 slots to two confederations each. For example, for the 2019 World Cup, CONCACAF received 3.5 slots and CONMEBOL received 2.5 slots. The fourth-placed team from CONCACAF and the third-placed team from CONMEBOL played an inter-continental play-off to determine which team advanced to the tournament. In the history of the Women's World Cup, inter-continental play-offs have determined only 5 out of 136 tournament participations.

that determines which teams will advance from the group stage to the single elimination knockout stage. Since this paper is concerned with the tournament phase, we will use “World Cup” and “tournament” interchangeably. The number of teams in the tournament has grown over time: 12 in 1991 and 1995; 16 from 1999 through 2011; 24 in 2015 and 2019. FIFA recently announced that 32 teams will participate in 2023. Figure 1 shows the evolution of the World Cup bracket in the knockout stage.

1.1 Competitive excitement

It is imperative in sports to produce “competitive excitement”, because competition is the product of sports industries and soccer leagues in particular (Haan, Koning, and van Witteloostuijn 2007). Uncertainty about a game result creates competitive excitement (Scarf and Yusof 2011) which in turn attracts customers, i.e., fans (Koning 2000). If the teams playing each other are balanced, meaning the teams are of equal strength, uncertainty about a game result is greater (Koning 2000). FIFA is aware of the importance of competitive excitement. After each World Cup, FIFA publishes a technical report compiled by FIFA's Technical Study Group (TSG) to track

progress in balancing team strength across all matches. The TSG argued in their technical report for the 2011 World Cup that “(t)he general improvement in the game was . . . reflected in the narrowing of the gap between countries, as only five of the 32 games ended with a goal difference of three or more” (FIFA 2011, p.10). The TSG implied that countries participating in the World Cup should be similar in ability to deliver high-quality play with competitive matches. A goal difference of 0 goals represents a draw which could be changed into a win/loss by a single goal in the last minute; a goal difference of 1 represents a game which could be tied by a single goal in the last minute; a goal difference of 2 is still considered close, as one goal could set up the possibility of a last-minute equalizer. So, we adopt the TSG's implied benchmark of goal differences of two goals or less to indicate competitive matches with excitement about the outcome of the match until the end of the match. The technical report for the 2015 World Cup went one step further stating: “It was also noticeable that the gap between teams is now smaller than ever before. Although it is fair to say that some teams are still more advanced than others, 72% of all matches were decided by a single goal” (FIFA 2015a, p.58). Yet, games such as Germany – Ivory Coast (10–0) and Switzerland – Ecuador (10–1) at the 2015 World Cup were sorely lacking in competitive excitement.

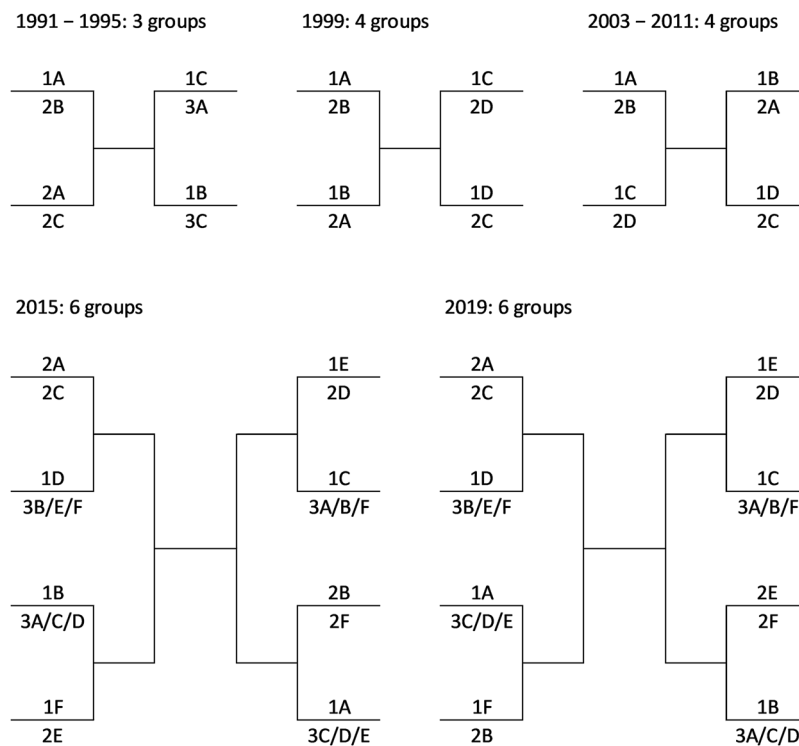


Figure 1: Evolution of the Women's World Cup bracket in the knockout stage. Each group consists of four teams. Based on the group standings, the brackets indicate which teams advance to the knockout stage as well as the matchups. For 1991 and 1995, FIFA had three brackets depending on which two third-placed finishers advanced from the group stage. The bracket shown is the bracket that was used both times. 1A represents the winner of group A. So, in 1991 the winner of group A (1A) played the runner-up of group B (2B) in the quarterfinals. The winner of 1A–2B played the winner of 2A–2C in the semifinals, etc. From 1991 through 2011, the knockout stage started with eight quarterfinalists. In 2015 and 2019, the knockout stage started with the round of 16. In addition to the winners and runners-up, the four best teams among the third-placed finishers in the group stage advanced to the round of 16. Article 28 in Regulations FIFA Women's World Cup Canada 2015 specifies where the third-placed finishers play in the bracket.

To achieve sporting fairness and competitive balance in a World Cup tournament, it is imperative that all groups have similar competitive levels (Guyon 2015).³ We illustrate how FIFA creates groups using the 2015 World Cup as an example. FIFA created 4 pots of 6 teams (see Table 1a) and protected most of the strongest teams as well as the host by assigning them as “seeds” in pot 1. Pots 2, 3, and 4 were based on continents. FIFA then placed the seeds from pot 1 into groups. Subsequently, for each group, FIFA randomly drew one team from each of the other pots. FIFA's creation of groups is problematic in terms of ensuring competitive balance. The first problem concerns the allocation of slots to the different confederations. Quite a few teams ranked in the top 24 in the world could not participate, whereas several substantially lower ranked teams could, such as teams ranked 30 (AFC), 40 (CONCACAF), 31 and 49 (COMMEBOL), and 51 and 64 (CAF). Consequently, the allocation of confederation slots is not aligned with the distribution of the best teams in the world.⁴

The second problem concerns the allocation of seeds: instead of randomly allocating seeds to the six groups, FIFA intentionally placed seeds into specific groups for ticketing and promotion reasons.

“For the 2015 Women's World Cup, . . . , FIFA almost decided that France and Germany would meet in quarterfinals, as they placed Germany in Group B and France in Group F, which implied that if both teams won their group and advanced to the quarterfinals, they would play each other – which is exactly what happened. This was, of course, a terrible way of organizing the tournament, and proved how difficult it is for FIFA to cope with sporting fairness” (Guyon 2018, p.315).

The third problem concerns drawing the remaining teams. FIFA does not impose restrictions on the draw to prevent creating groups of unequal strength. As Table 1b shows, Group C ended up with one of the two worst teams from pot 2 (ranked 51), the worst team from pot 3 (ranked 49), and the worst team from pot 4 (ranked 18). Conversely, Group D ended up with the third team from pot 2 (ranked 35), the best team from pot 3 (ranked 10), and the best team from pot 4 (ranked 5). So, Group D became a very tough group, whereas Group C became a very weak group.

³ The Women's World Cup knockout brackets are built with the assumption of balance. For example, Csató (2021) discusses how the 2015 World Cup satisfies balance. In contrast, Csató (2020a) uses the example of handball to demonstrate how intentionally imbalanced groups can increase the quality of all matches without sacrificing fairness.

⁴ Stone and Rod (2016) and Csató (2022a) investigate the allocation of slots to confederations in the Men's World Cup and advocate for a more transparent allocation process.

In Section 2, we use match data from the World Cups to empirically assess competitive balance across groups in every Women's World Cup. We use least squares to determine ratings for all teams (Koning 2017; Winston 2009). For each team, we average the ratings of the opponents in the group to calculate group opponents rating. We find that the range in group opponents rating varies between 2.5 and 4.5 goals indicating substantial competitive imbalance. In Section 3, logistic regression shows that one goal less in group opponents rating can increase the probability of reaching the quarterfinal by 33%. Lastly, we discuss several policy recommendations to reduce competitive imbalance in Section 4.

1.2 Related research

Related research has focused on building structures for single elimination tournaments (Horen and Riezman 1985, Scarf and Yusof 2011) and analyzing seeding systems in the UEFA Champions League (Corona et al. 2019; Csató 2020b; Dagaev and Rudyak 2019; Engist, Merkus, and Schafmeister 2021). For the FIFA Men's World Cups between 1982 and 2006, Monks and Husch (2009) study the impact of seeding, home continent, and hosting on World Cup success. The authors find that teams benefit from being seeded and playing in their home continent, but host teams do not enjoy any advantages beyond the benefits of seeding and the home continent effect.

Several papers have studied FIFA's procedure to allocate teams to groups for the group stage in the Men's World Cup. Jones (1990) shows that the draw for the 1990 Men's World Cup was not mathematically fair. Similarly, Rathgeber and Rathgeber (2007) explain that in the 2006 Men's World Cup, Germany was likely to play in a tough group whereas Italy – another seeded team – was not. Guyon (2015) identifies three flaws in FIFA's final draw procedure used for 32-team Men's World Cup tournaments. First, the draw is imbalanced, meaning the resulting groups are not at the same competitive level. Second, the draw is unfair, meaning that some teams have a greater chance of ending up in a tough group. Third, the draw is unevenly distributed, meaning not all possible outcomes of the draw are equally likely. Guyon (2015) then proposes a new draw procedure that addresses all three flaws. Building on Guyon (2015), Laliena and López (2019) develop two evenly distributed designs for the draw with geographical restrictions that produce groups with similar (or equal) competitive levels. In the same context of 32-team Men's World Cup tournaments, Cea et al. (2020) propose an alternative draw procedure using an optimization method to minimize the difference between the maximum and

Table 1a: Pots for FIFA World Cup 2015.

Pot 1 (Seeds)	Pot 2 (CAF, CONCACAF, OFC)	Pot 3 (AFC, CONMEBOL)	Pot 4 (UEFA)
Canada (8)	Cameroon (51)	Australia (10)	England (7)
Brazil (6)	Ivory Coast (64)	China PR (14)	Netherlands (15)
France (4)	Nigeria (35)	South Korea (17)	Norway (9)
Germany (2)	Costa Rica (40)	Thailand (30)	Spain (16)
Japan (3)	Mexico (25)	Colombia (31)	Sweden (5)
United States (1)	New Zealand (19)	Ecuador (49)	Switzerland (18)

FIFA rankings at the time of the draw (December 6, 2014) in parentheses.

Table 1b: The draw for the FIFA World Cup 2015.

Group A	Group B	Group C	Group D	Group E	Group F
Canada (8)	Germany (2)	Japan (3)	United States (1)	Brazil (6)	France (4)
China PR (14)	Norway (9)	Cameroon (51)	Australia (10)	South Korea (17)	England (7)
Netherlands (15)	Thailand (30)	Switzerland (18)	Sweden (5)	Costa Rica (40)	Colombia (31)
New Zealand (19)	Ivory Coast (64)	Ecuador (49)	Nigeria (35)	Spain (16)	Mexico (25)

FIFA rankings at the time of the draw (December 6, 2014) in parentheses.

minimum ranking sums of each group's members. Another source of competitive imbalance is the order in which pots are emptied during the draw procedure. For the 2018 Men's World Cup, Csató (2022b) shows that FIFA's draw order distorts the probability of advancing to the knockout stage.

Guyon (2018) studies the UEFA Men's European Championship which has the same 24-team structure as the FIFA Women's World Cup tournament in 2015 (Figure 1). One flaw of this 24-team structure is group advantage, i.e., from some groups it is easier to advance than from others. Another flaw is lack of win incentive, i.e., for some groups it is unclear whether it is better to finish first or second, or – if the third-placed team moves on – whether it is better to finish second or third in the group. Problematically, a lack of win incentive can lead to tanking or collusion (Chater et al. 2021; Guyon 2020; Stronka 2020). To overcome these flaws, Guyon's (2018) proposed solution ranks the teams qualified for the knockout stage and seeds these teams in the draw after the group stage. Building on Guyon (2018), Csató (2021) analyzes soccer tournaments with 24 teams. The author concludes that UEFA's redesign of the 2020 Men's European Championship dominates FIFA's redesign of the 2019 Women's World Cup (Figure 1). For example, in the 2019 bracket, the winner and runner-up of groups A and E could face each other again in the semifinals, whereas it is preferable to prevent a repeated matchup as such before the final.

Guyon (2015, 2018), Laliena and López (2019) and Cea et al. (2020) all address the draw and propose better ways to execute the draw; however, these papers do

not empirically assess imbalance at the World Cup. Our contribution is to use game outcomes at the Women's World Cup tournaments to (i) empirically assess imbalance between groups, (ii) quantify the extent of imbalance, (iii) quantify the impact of imbalance on the probability of success in the World Cup, and (iv) quantify the spread in probability of success across teams as a result of imbalance.

2 Assessing imbalance in groups

Prior research has investigated models to predict scores for domestic soccer matches within a national league. Maher (1982), for example, used a bivariate Poisson model to predict soccer scores for English leagues. International soccer – between national teams – is different (McHale and Davies 2007). Many matches are qualifying matches for major tournaments, and within tournaments top teams are seeded to avoid playing each other too early in a tournament. As a result, many international matches have low competitive balance. Because low competitive balance leads to negative correlation between the two teams' scores, the bivariate Poisson model is not appropriate for international soccer (McHale and Davies 2007). Lasek, Szlávik, and Bhulai (2013) compare the predictive capabilities of different ranking systems. The authors find that several ranking systems outperform FIFA's method to rank men's teams. One of the methods with better predictive performance is the least squares method that we will use in this paper. Least squares can easily

accommodate incomplete schedules (i.e., when a team does not play all other teams) and draws which are common in soccer. For recent reviews of rating soccer teams see Koning (2017), Van Eetvelde and Ley (2019) and Groll, Schauburger, and Van Eetvelde (2020). Other papers have extended rating methods to reflect teams' current strength (Ley, Van de Wiele, and Van Eetvelde 2019) and to find the all-time greatest teams (Baker and McHale 2018).

2.1 Rating teams

To assess whether groups in a World Cup are balanced, we first determine team ratings and subsequently use the ratings to calculate group strength. We follow the least squares procedure from Winston and Albright (1997) and Winston (2009) to determine team ratings also referred to as team qualities (Haan, Koning, and van Witteloostuijn 2007, Koning 2017). This least squares procedure has also been used in English soccer (Clarke and Norman 1995, Koning 2017) and in several national European soccer leagues to assess competitive balance (Haan, Koning, and van Witteloostuijn 2007). Let r_{it} be the rating for team i in World Cup t . The rating represents the expected scoring margin against an average team in World Cup t . If $r_{it} > 0$, then team i is expected to beat an average team by r_{it} goals in World Cup t . If $r_{it} < 0$, then team i is expected to lose to an average team by r_{it} goals in World Cup t . A great advantage of the least squares method is the clear interpretation of the difference in ratings, which is lacking in other methods (Groll, Schauburger, and Van Eetvelde 2020).

We use the actual match results to calculate the scoring margin for each match. By convention, the team listed first is called the home team, and the team listed second is called the away team. (Note: this only affects the choice of kits used by the teams, as each team has a home kit and an away kit.) For each match, the scoring margin is the number of goals scored by the home team minus the number of goals scored by the away team. For example, in 2019 the result for Australia–Italy was 1–2. So, the scoring margin for this particular match was -1 . Formally, let m_{ijt} represent the scoring margin for the match between home team i and away team j in World Cup t . The number of teams in World Cup t is N_t . For World Cups 1991 and 1995, $N_t = 12$, for 1999 through 2011, $N_t = 16$, and for 2015 and 2019, $N_t = 24$. Lastly, let $S(t)$ denote the set of all matches that were played at World Cup t .

Note that the forecasted margin for a match between home team i and away team j in World Cup t would be $r_{it} - r_{jt}$, and the forecasted error for this match would

be $(r_{it} - r_{jt}) - m_{ijt}$. For each World Cup t , we solve for the team ratings by minimizing the sum of squared errors subject to the average rating being equal to zero (Winston 2009):

$$\begin{aligned} \min \quad & \sum_{(i,j) \in S(t)} ((r_{it} - r_{jt}) - m_{ijt})^2 \\ \text{s.t.} \quad & \frac{1}{N_t} \sum_{i=1}^{N_t} r_{it} = 0 \end{aligned}$$

This rating problem is a quadratic programming problem. For each World Cup t , we use Excel Solver to solve this rating problem. Since we minimize a convex quadratic function in the team ratings subject to a linear constraint, Solver is guaranteed to find the global minimum (Winston and Albright 1997).

2.2 Group strength

Less variation in the ratings r_{it} means better competitive balance (Haan, Koning, and van Witteloostuijn 2007). We use the empirically determined ratings in Section 2.1 to calculate group strength. Following Winston (2009), for each group G_t in World Cup t we average the ratings of the teams: $gs_{G_t} = \frac{1}{4} \sum_{i \in G_t} r_{it}$. Figure 2 shows the group strength for all groups in each World Cup. Note that balance across all groups in a World Cup means $gs_{G_t} = 0$ for all groups. Clearly, Figure 2 shows that groups are not of equal strength. In fact, for each World Cup, the strongest group is at least one goal tougher than the weakest group. One goal makes a big difference in soccer (Winston 2009). In the 2003 World Cup, the range in group strength is more than three goals.

2.3 Group opponents rating

We adopt the FIFA 2011 TSG's implied benchmark of absolute scoring margins of two goals or less. To assess whether – on average – teams face opponents in the group stage within the benchmark of absolute scoring margins of two goals or less, we calculate group opponents rating for team i in World Cup t as $gopr_{it} = \frac{1}{3} \sum_{j \in G_{it}} r_{jt}$, where G_{it} is the set of three opponents for team i in the group stage of World Cup t . Figure 3 shows the range in group opponents rating for each World Cup. This range is more than the two-goal benchmark for each World Cup. In 2003, the range was more than 4.5 goals. Clearly, the range in quality of teams has been high for every World Cup – too many teams participate in the World Cup which cannot compete with the best teams.

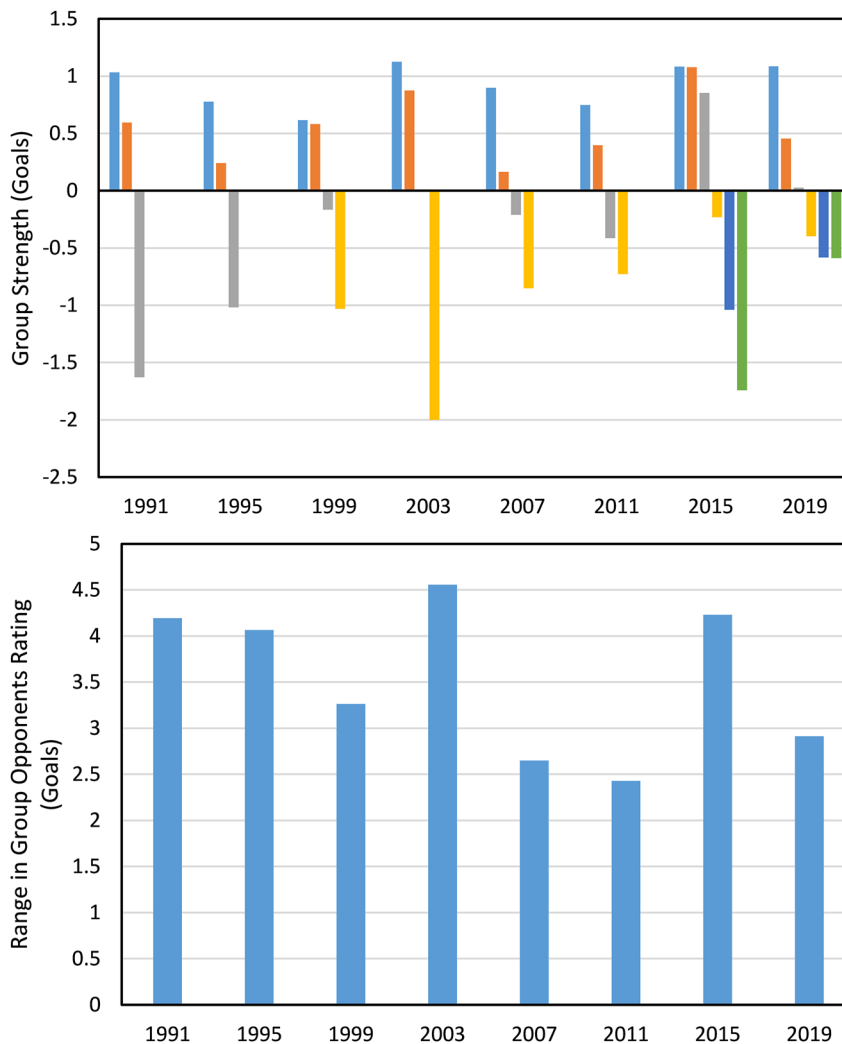


Figure 2: Group strength. Number of groups in 1991–1995: 3, 1995–2011: 4, 2015–2019: 6. Group strength represents the average scoring margin (in goals) when an average team in the group plays against an average team in the World Cup.

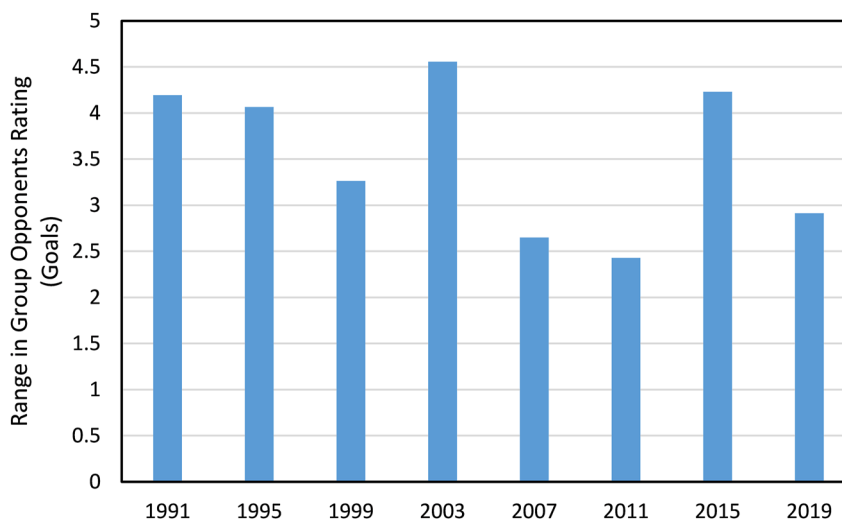


Figure 3: Range in group opponents rating.

3 Impact of imbalanced groups

Next, we investigate the impact of group opponents rating on the probability of reaching the quarterfinal. We choose the quarterfinal for several reasons. First, the quarterfinal is the first common round in the knockout phase across all World Cup tournaments from 1991 through 2019 (Figure 1). Second, if the highest ranked teams win their groups, the brackets in Figure 1 do not prevent the best eight teams from advancing to the quarterfinal. However, the brackets do prevent some of the best four teams to advance to the semifinals. For example, the bracket in Figure 1 pre-determined that in 2015, Germany (1B, ranked 2nd at the time of the draw) and France (1F, ranked 4th) would meet in

the quarterfinals if they each won their group – which they did. Similarly, the bracket pre-determined that in 2019, the USA (1F, ranked 1st) and host France (1A, ranked 3rd) would meet in the quarterfinals if they each won their group – which also materialized. In both examples, if the best ranked teams won their groups, then at least one team ranked in the top four of the world was guaranteed to be eliminated before the semifinals. Third, for our estimation method, logistic regression, Hosmer, Lemeshow, and Sturdivant (2013) recommend at least 10 observed events per explanatory variable to avoid overfitting. Since we include four independent variables, we need at least 40 events. We meet this recommendation with 64 quarterfinalists from all eight World Cups combined. With only 32 semifinalists

from all World Cups, analyses for probabilities of reaching the semifinal would not meet the recommended 40 events for four explanatory variables.

3.1 Impact of group opponents rating

Our main dependent variable is $QF_{it} = 1$ if team i reached the quarterfinal in World Cup t , 0 otherwise. Our main independent variable is the average strength of the three opponents in the group of team i in World Cup t : $gopr_{it}$. We control for the number of teams, N_t , in World Cup t . Soccer teams playing at home have an advantage (Clarke and Norman 1995). At the World Cup, only the hosting nation plays in their home country. Any home advantage for the host will be included in the host rating. However, teams from the same continent as the host (including the host itself) can also benefit from climatic conditions and cultural circumstances (Groll, Schauburger, and Tutz 2015). In addition, teams from the same continent and their fans benefit from shorter travel distances (Monks and Husch 2009).⁵ We control for home continent advantage with $HC_{it} = 1$ if World Cup t was held in the home continent of team i . Lastly, we control for the rating of team i in World Cup t , r_{it} .

We use logistic regression to assess the impact of group opponents rating on the probability of reaching the quarterfinal:

$$\ln \frac{\Pr(QF_{it} = 1)}{1 - \Pr(QF_{it} = 1)} = \beta_0 + \beta_1 N_t + \beta_2 HC_{it} + \beta_3 r_{it} + \beta_4 gopr_{it} + e_{it}$$

A negative value for β_1 would indicate that the probability of reaching the quarterfinal is reduced when there are more teams participating in the World Cup. A positive value for β_2 would indicate that teams playing in their home continent enjoy a higher probability of reaching the quarterfinal. A positive value for β_3 would indicate that higher rated teams have a higher probability of reaching the quarterfinal. A negative value for β_4 would indicate that playing against tougher opponents in the group stage reduces the probability of reaching the quarterfinal.

Table 2 shows the logistic regression results when we add the independent variables sequentially. In Model (1),

the estimate for β_1 is negative and statistically significant. So, the probability of reaching the quarterfinal is reduced when there are more teams participating in the World Cup. In Model (2), the estimate for β_2 is positive and statistically significant supporting the notion that teams playing in their home continent have a higher probability of reaching the quarterfinal. In Model (3), the estimate for β_3 is positive and statistically significant. As expected, better teams have a higher probability of reaching the quarterfinal. In Model (4), the estimate for β_4 is negative and statistically significant. Furthermore, the estimates for β_1 , β_2 and β_3 in Model (4) are consistent with the previous models.

The LR χ^2 tests whether at least one of the estimated coefficients show a statistically significant difference from zero. Clearly, for all models, the LR χ^2 rejects the null hypothesis that all estimated coefficients are zero. The Pseudo R^2 reported is the McFadden Pseudo R^2 . Higher values indicate better fit. For datasets with fewer than 200 observations and percent observations of success between 38% and 62%, Hemmert et al. (2018) note that McFadden Pseudo R^2 values between 0.2 and 0.4 suggest good model fit and values greater than 0.4 suggest excellent model fit. With 136 observations in Table 2 and 64 quarterfinalists representing 47% of the observations, we can use these Pseudo R^2 benchmarks. As the Pseudo R^2 increases with every variable added in Table 2, the model fit improves with each additional variable. The Pseudo R^2 of 0.644 for Model (4) suggests excellent model fit. So, even when we control for the number of teams, home continent advantage, and the quality of the teams, the probability of reaching the quarterfinal is reduced when teams have to play against tougher opponents in the group stage. While we expect β_4 to be negative, we can now quantify the impact of the variation in group opponents rating from Figure 3 on the probability of reaching the quarterfinal.

Figure 3 showed tremendous range in group opponents rating. We can use the estimates in Table 2 to illustrate the impact of group opponents rating on the probability of reaching the quarterfinal. Let $\hat{\beta}_s$ be the estimate for β_s ($s = 0, \dots, 4$). We can solve the estimated logistic regression model for the estimated probability of reaching the quarterfinal, $\Pr(QF_{it} = 1)$:

$$\begin{aligned} \ln \frac{\Pr(QF_{it} = 1)}{1 - \Pr(QF_{it} = 1)} &= \hat{\beta}_0 + \hat{\beta}_1 N_t + \hat{\beta}_2 HC_{it} + \hat{\beta}_3 r_{it} + \hat{\beta}_4 gopr_{it} \\ &\Leftrightarrow \widehat{\Pr(QF_{it} = 1)} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 N_t + \hat{\beta}_2 HC_{it} + \hat{\beta}_3 r_{it} + \hat{\beta}_4 gopr_{it}}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 N_t + \hat{\beta}_2 HC_{it} + \hat{\beta}_3 r_{it} + \hat{\beta}_4 gopr_{it}}} \end{aligned}$$

⁵ For the Women's World Cups from 1991 to 2019, there were 34 seeded teams of which 31 seeds advanced to the quarterfinal. So, unlike the Men's World Cup data used by Monks and Husch (2009), there is not enough variation in success among the seeded teams in the Women's World Cup to include a dummy variable for seeds.

Table 2: Logistic regression models: 1991–2019.

	(1)	(2)	(3)	(4)
Constant, β_0	1.762* (0.728)	1.526* (0.769)	4.087** (1.296)	4.904** (1.605)
Number of teams, β_1	−0.104** (0.039)	−0.112** (0.042)	−0.277*** (0.072)	−0.325*** (0.089)
Home continent, β_2		1.564*** (0.456)	2.258*** (0.705)	2.347** (0.783)
Team rating, β_3			1.724*** (0.328)	2.015*** (0.396)
Group opponents rating, β_4				−1.364*** (0.366)
LR χ^2	7.34**	20.61***	101.45***	121.07***
Pseudo R^2	0.039	0.110	0.539	0.644
Number of observations	136	136	136	136

Standard errors in parentheses. * Significant at 0.05, ** at 0.01, and *** at 0.001. For World Cups 1991 and 1995, $N_t = 12$, for 1999 through 2011, $N_t = 16$, and for 2015 and 2019, $N_t = 24$. So, we have $2 \times 12 + 4 \times 16 + 2 \times 24 = 136$ observations.

We calculate for each World Cup the estimated probability of reaching the quarterfinal for an average team (i) facing the highest observed group opponents rating and (ii) facing the lowest observed group opponents rating. Specifically, to calculate $\Pr(\widehat{QF}_{it} = 1)$ for each World Cup t , we substitute the average home continent $\overline{HC}_t = 1/N_t \sum_{i=1}^{N_t} HC_{it}$ and the average team rating $\bar{r}_t = 1/N_t \sum_{i=1}^{N_t} r_{it}$ which is 0 by construction, as well as the largest and smallest values for $goppr_{it}$ observed in World Cup t . Figure 4 shows the resulting range in $\Pr(\widehat{QF}_{it} = 1)$ for each World Cup. The range in $\Pr(\widehat{QF}_{it} = 1)$ is huge. If teams of equal strength participate in World Cup t , $\Pr(\widehat{QF}_{it} = 1)$ should be around $8/N_t$, the number of quarterfinal spots divided by the number of teams in World Cup t . Clearly, this is not the case.

In Figure 5, we illustrate the relationship between group opponents rating and the probability of reaching the quarterfinal for an average team. Because $\Pr(\widehat{QF}_{it} = 1)$ depends on N_t and HC_{it} , we plot $\Pr(\widehat{QF}_{it} = 1)$ for 12, 16,

and 24 team World Cups both for average teams playing in their home continent as well as away from their home continent. For example, the range in group opponents rating in 24 team World Cups, for teams playing in their home continent (24-HC) was -2.88 to 1.35 which translates to an estimated probability of reaching the quarterfinal from 0.97 down to 0.08. Going from a group opponents rating of 0 to a group opponents rating of -1 changes the estimated probability of reaching the quarterfinal from 0.36 to 0.69. These variations in group opponents rating are typical and the implied changes in probability of reaching the quarterfinal are huge.

We can also use the example of an average team playing in their home continent in a 24 team World Cup with a group opponents rating of 0 to compare the effect of group opponents rating to the effects of home continent and number of teams. If an average team in a group with a group opponents rating of 0 in a 24 team World Cup were to play in an away continent rather than their home continent, the estimated probability of reaching the quarterfinal would drop from 0.36 to 0.05. This effect is

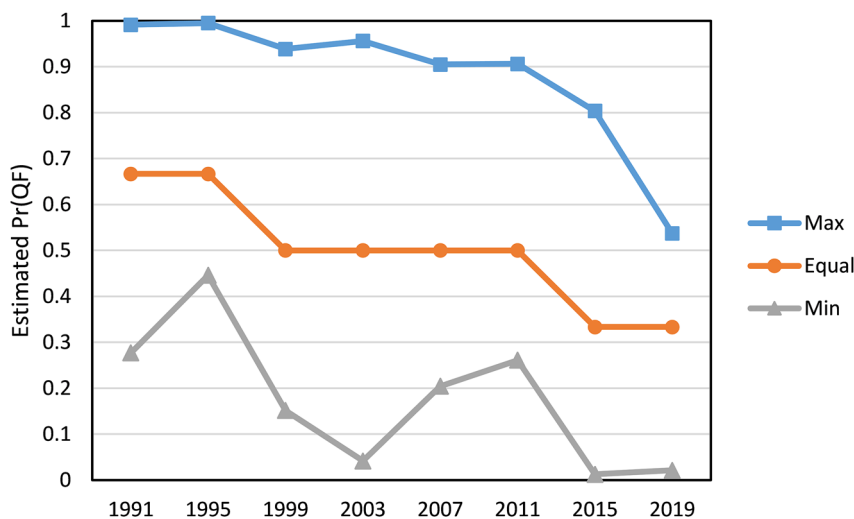


Figure 4: Impact of range in group opponents rating on reaching the quarterfinal. Max (Min) is the estimated probability of reaching the quarterfinal for an average team facing the lowest (highest) observed group opponents rating in each World Cup. Equal is $8/N_t$.

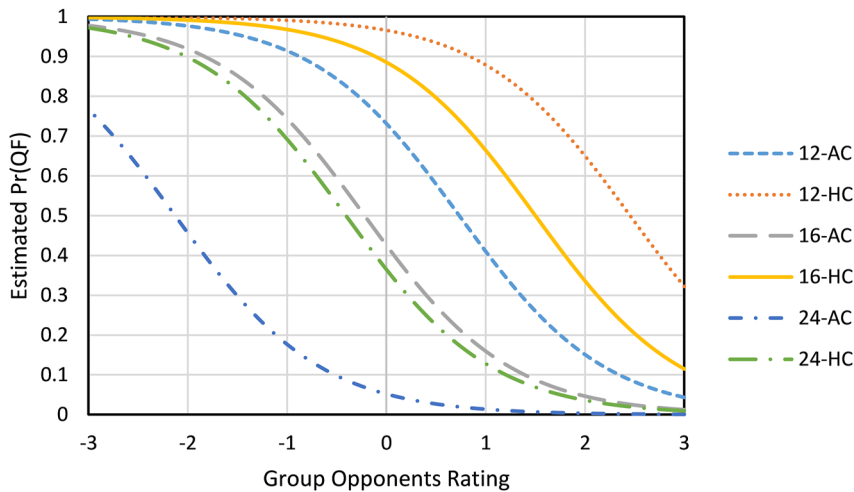


Figure 5: Impact of group opponents rating on reaching the quarterfinal. Estimated probability of reaching the quarterfinal in 12, 16, and 24 team World Cups for an average team playing in their home continent (HC) as well as in an away continent (AC). The observed ranges in group opponents rating for 12, 16, and 24 team World Cups were -2.46 to 1.87 , -2.27 to 2.29 , and -2.88 to 1.35 respectively.

roughly equal to playing in a group with a 1.7 goals higher group opponents rating. If an average team playing in a group with a group opponents rating of 0 in their home continent were to play in a 16 team World Cup rather than a 24 team World Cup, the estimated probability of reaching the quarterfinal would increase from 0.36 to 0.89 which is roughly equal to playing in a group with 1.9 goals lower group opponents rating.

3.2 Impact of group opponents FIFA points

FIFA started ranking Women's teams in 2003. Four times a year, FIFA updates and publishes its Women's World Ranking, which is based on a points system, see FIFA Rankings Procedure (FIFA 2019a). The scaling "is chosen in such a way that the very best in the world can have rating points exceeding 2000, while the absolute beginners score around 1000 rating points" (FIFA 2019a, p.3). Because the ranking method is an adjustment to the previous ranking, the rankings are somewhat sticky.⁶ Our ratings determined in Section 2 are based solely on the matches played during the World Cup tournament. So, it is worthwhile to assess how sensitive our finding for group opponents rating is when we use group opponents' performance in the recent past rather than group opponents' performance at the World Cup. We re-run our analysis using the most recent FIFA Women's World Ranking prior to the World Cup (FIFA 2019b). Because FIFA started the Women's World Ranking in 2003, we have to omit the World Cups from

1991, 1995, and 1999 in this analysis. Let FIFA Points_{it} be the points in the FIFA ranking for team i immediately prior to World Cup t (FIFA Women's Ranking 2019). We calculate group opponents FIFA points for team i in World Cup t as $\text{goppFIFAp}_{it} = \frac{1}{3} \sum_{j \in G_{it}} \text{FIFA Points}_{jt}$, where G_{it} is the set of three opponents for team i in the group stage of World Cup t . Next, we estimate:

$$\ln \frac{\Pr(\text{QF}_{it} = 1)}{1 - \Pr(\text{QF}_{it} = 1)} = \beta_0 + \beta_1 N_t + \beta_2 \text{HC}_{it} + \beta_3 r_{it} + \beta_4 \text{goppFIFAp}_{it} + e_{it}$$

Table 3 shows the estimation results. For Models (2)–(4), the LR χ^2 rejects the null hypothesis that all estimated coefficients are zero. As the Pseudo R^2 increases with every variable added in Table 3, the model fit improves with each additional variable. The Pseudo R^2 of 0.553 for Model (4) suggests excellent model fit. Similarly to Table 2, we find that (i) the probability of reaching the quarterfinal is reduced when there are more teams participating in the World Cup, (ii) teams playing in their home continent enjoy a higher probability of reaching the quarterfinal, as do higher quality teams, and (iii) the probability of reaching the quarterfinal is reduced when teams have to play tougher opponents in the group stage.

3.3 Impact on group winners

One alternative explanation for the huge variation in $\Pr(\text{QF}_{it} = 1)$ is that each group has the same range in group opponents rating, i.e., each group has some really good teams of similar level and some really bad teams of similar level. While Figure 2 suggests this is not the case, we can formally address this alternative explanation. The team which finishes first in the group after round-robin play is

⁶ Several scholars have criticized the (former) FIFA Men's World Ranking (Cea et al. 2020; Lasek, Szlávik, and Bhulai 2013, Lasek et al. 2016). In contrast, the FIFA Women's World Ranking – based on the Elo rating system – is a competitive ranking method for prediction, just like the least squares method (Lasek, Szlávik, and Bhulai 2013).

designated the group winner. In Figures 6 and 7, we plot goppr_{it} and goppFIFApts_{it} for the group winners. Clearly, in each World Cup, there was tremendous variation in group opponents rating across the group winners – typically

more than 2 goals, and in 2015 more than 3.5 goals as shown in Figure 6. The large variation in group opponents rating across group winners should have been anticipated based on the group opponents FIFA points shown in Figure 7.

Table 3: Logistic regression models: 2003–2019.

	(1)	(2)	(3)	(4)
Constant, β_0	1.386 (1.061)	1.326 (1.145)	2.564 (1.622)	26.754** (9.567)
Number of teams, β_1	−0.087 (0.053)	−0.107 (0.058)	−0.205* (0.083)	−0.338*** (0.106)
Home continent, β_2		1.921*** (0.556)	1.911** (0.719)	1.995** (0.754)
Team rating, β_3			1.746*** (0.416)	1.769*** (0.438)
Group opponents FIFA points, β_4				−0.012** (0.004)
LR χ^2	2.76	16.60***	63.32***	72.14***
Pseudo R^2	0.021	0.127	0.486	0.553
Number of observations	96	96	96	96

Standard errors in parentheses. * Significant at 0.05, ** at 0.01, and *** at 0.001. For World Cups 2003 through 2011, $N_t = 16$, and for 2015 and 2019, $N_t = 24$. So, we have $3 \times 16 + 2 \times 24 = 96$ observations.

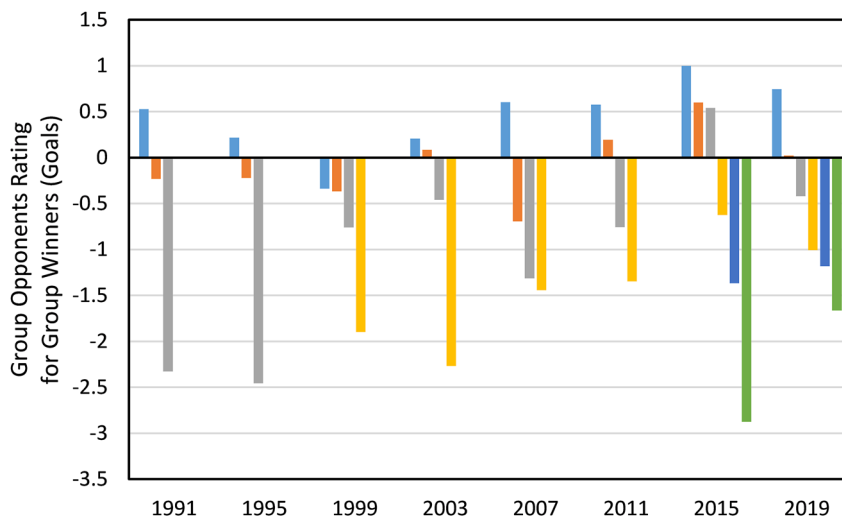


Figure 6: Group opponents rating for the group winners. For each group winner, the bar represents the average rating for the three opponents faced in the group stage.

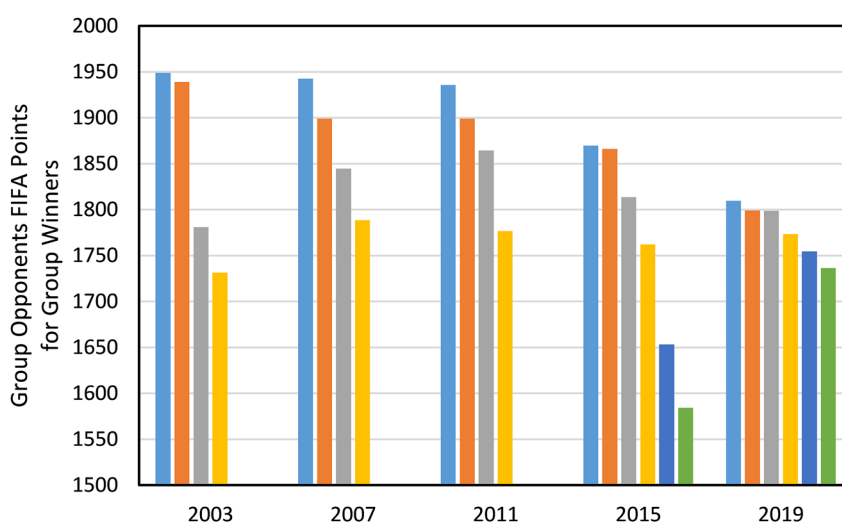


Figure 7: Group opponents FIFA points for the group winners. For each group winner, the bar represents the average FIFA points for the three opponents faced in the group stage.

4 Concluding remarks

We have empirically assessed the competitive imbalance across groups in the FIFA Women's World Cup tournaments. The range in group strength is more than 1 goal in each World Cup and was as high as 3.5 goals in 2003. The range in group opponents rating is more than 2 goals in each World Cup and was as high as 4.5 goals in 2003. Using logistic regression, we show that 1 goal less in group opponents rating can increase the probability of reaching the quarterfinal by 33%. Clearly, there is substantial competitive imbalance across groups which should be reduced to increase competitive excitement. Routs such as USA – Thailand (13–0) in 2019 and Germany – Ivory Coast (10–0) in 2015 are sorely lacking in competitive excitement. Several avenues can be pursued to reduce competitive imbalance.

First, the draw procedure should be changed to better balance the groups. Recent research has proposed several ways to improve draw procedures (Cea et al. 2020; Csató 2022b; Guyon 2015, 2018, Laliena and López 2019).

Second, the number of teams at the World Cup could be reduced to have a more homogenous, competitive pool of participating teams. We realize that this avenue is the opposite of what FIFA intends to do, but it is an obvious way to reduce competitive imbalance.

Third, the allocation of the number of teams to confederations should better match the actual distribution of top teams in the world. For example, in 2019, the teams ranked 17 and 18 (Denmark and Switzerland) could not participate in the 24 team World Cup because the number of European teams had been limited to 9 whereas 12 European teams were ranked in the top 24. Yet, several less competitive teams such as Thailand (34), South Africa

(49) and Jamaica (53) did participate. We do not know the method FIFA uses to allocate the number of confederation slots. Geographic diversity is desirable, but the number of confederation slots could better reflect the FIFA rankings prior to each World Cup while still having some geographic diversity.

These avenues are not mutually exclusive. For example, FIFA could improve the allocation of confederation slots and improve the draw procedure to better balance the groups. We have empirically demonstrated that competitive imbalance across groups in the Women's World Cup tournaments is substantial. Moreover, competitive imbalance significantly affects the probability of success at the World Cup. We hope FIFA pursues avenues to reduce competitive imbalance to increase the level of fairness and consequently boost competitive excitement.

Acknowledgments: The authors thank Megan Lawrence and Brian McCann for helpful comments on an earlier version of this paper. The authors also thank two anonymous reviewers for the thoughtful and constructive feedback during the review process.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix

Table 4 shows that the pairwise correlations between any two explanatory variables used in our logistic regression models are not statistically significant. Consequently, multicollinearity is not a concern in our estimations.

Table 4: Correlation matrix.

		1	2	3	4	5
1	Quarterfinal QF_{it}	–				
2	Number of teams N_t	–0.230**	–			
3	Home continent HC_{it}	0.306***	–0.030	–		
4	Team rating r_{it}	0.626***	–0.000	0.139	–	
5	Group opponents rating $gopr_{it}$	–0.317***	0.000	–0.018	–0.101	–

Number of observations is 136. * Significant at 0.05, ** at 0.01, and *** at 0.001.

References

- Baker, R. D., and I. G. McHale. 2018. "Time-Varying Rating for International Football Teams." *European Journal of Operational Research* 267 (2): 659–66.
- Cea, S., G. Durán, M. Guajardo, D. Sauré, J. Siebert, and G. Zamorano. 2020. "An Analytics Approach to the FIFA Ranking Procedure and the World Cup Final Draw." *Annals of Operations Research* 286 (1–2): 199–46.
- Chater, M., L. Arrondel, J.-P. Gayant, and J.-F. Laslier. 2021. "Fixing Match-Fixing: Optimal Schedules to Promote Competitiveness." *European Journal of Operational Research* 294 (2): 673–83.
- Clarke, S. R., and J. M. Norman. 1995. "Home Ground Advantage of Individual Clubs in English Soccer." *The Statistician* 44 (4): 509–21.
- Corona, F., D. Forrest, J. D. Tena, and M. Wiper. 2019. "Bayesian Forecasting of UEFA Champions League under Alternative Seeding Regimes." *International Journal of Forecasting* 35 (2): 722–32.
- Csató, L. 2020a. "Optimal Tournament Design: Lessons from the Men's Handball Champions League." *Journal of Sports Economics* 21 (8): 848–68.
- Csató, L. 2020b. "The UEFA Champions League Seeding is not Strategy-Proof since the 2015/16 Season." *Annals of Operations Research* 292 (1): 161–9.
- Csató, L. 2021. *Tournament Design: How Operations Research Can Improve Sports Rules*. Cham, Switzerland: Palgrave Macmillan.
- Csató, L. 2022a. "Quantifying the Unfairness of the 2018 FIFA World Cup Qualification." *International Journal of Sports Science & Coaching*. in press. <https://doi.org/10.1177/174795412110734>.
- Csató, L. 2022b. "On the Fairness of the Restricted Group Draw in the 2018 FIFA World Cup." Manuscript. arXiv: 2103.11353.
- Dagaev, D., and V. Y. Rudyak. 2019. "Seeding the UEFA Champions League Participants: Evaluation of the Reform." *Journal of Quantitative Analysis in Sports* 15 (2): 129–40.
- Engist, O., E. Merkus, and F. Schafmeister. 2021. "The Effect of Seeding on Tournament Outcomes: Evidence from a Regression-Discontinuity Design." *Journal of Sports Economics* 22 (1): 115–36.
- FIFA. 2011. *FIFA Women's World Cup Germany 2011: Technical Report and Statistics*. <https://resources.fifa.com/image/upload/germany-2011-1508769.pdf?cloudid=pcahcegynzgmcyv4mdu> (accessed July 20, 2019).
- FIFA. 2015a. *FIFA Women's World Cup Canada 2015: Technical Report and Statistics*. <https://resources.fifa.com/image/upload/fifa-women-s-world-cup-canada-2015-technical-report-and-statistics-26708-2670891.pdf?cloudid=jaeq2lvmczqjofxcj3u> (accessed July 20, 2019).
- FIFA. 2015b. *FIFA Women's World Cup Canada 2015: Television Audience Report*. https://resources.fifa.com/mm/document/affederation/tv/02/74/59/85/fwwccanada2015tvaudiencereport_neutral.pdf (accessed August 2, 2019).
- FIFA. 2019a. "FIFA Rankings Procedure." In *FIFA/Coca-Cola Women's World Ranking*. <https://img.fifa.com/image/upload/rxqyxdjhbs2qdtstluy6.pdf> (accessed July 20, 2019).
- FIFA. 2019b. *FIFA Women's Ranking*. <https://www.fifa.com/fifa-world-ranking/ranking-table/women/> (accessed July 20, 2019).
- FIFA. 2019c. *FIFA Women's Football Member Associations Survey Report*. <https://resources.fifa.com/image/upload/fifa-women-s-survey-report-confederations-global-mas.pdf?cloudid=nq3ensohyxpuxovcovj0> (accessed August 2, 2019).
- Groll, A., G. Schauburger, and G. Tutz. 2015. "Prediction of Major International Soccer Tournaments Based on Team-specific Regularized Poisson Regression: An Application to the FIFA World Cup 2014." *Journal of Quantitative Analysis in Sports* 11 (2): 97–115.
- Groll, A., C. Ley, G. Schauburger, H. Van Eetvelde, and A. Zeileis. 2019. "Hybrid Machine Learning Forecasts for the FIFA Women's World Cup 2019." Manuscript. arXiv: 1906.01131.
- Groll, A., G. Schauburger, and H. Van Eetvelde. 2020. "Ranking and Prediction Models for Football Data." In *Science Meets Sports: When Statistics Are More than Numbers*, edited by C. Ley, and Y. Dominicy, 95–122. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Guyon, J. 2015. "Rethinking the FIFA World Cup Final Draw." *Journal of Quantitative Analysis in Sports* 11 (3): 169–82.
- Guyon, J. 2018. "What a Fairer 24 Team UEFA Euro Could Look like." *Journal of Sports Analytics* 4 (4): 297–317.
- Guyon, J. 2020. "Risk of Collusion: Will Groups of 3 Ruin the FIFA World Cup?" *Journal of Sports Analytics* 6 (4): 259–79.
- Haan, M., R. H. Koning, and A. van Witteloostuijn. 2007. "Competitive Balance in National European Soccer Competitions." In *Statistical Thinking in Sports*, edited by J. Albert, and R. H. Koning, 63–75. Boca Raton: CRC Press.
- Hemmert, G. A. J., L. M. Schons, J. Wieseke, and H. Schimmelpfennig. 2018. "Log-Likelihood-Based Pseudo-R² in Logistic Regression: Deriving Sample-Sensitive Benchmarks." *Sociological Methods & Research* 47 (3): 507–31.
- Horen, J., and R. Riezman. 1985. "Comparing Draws for Single Elimination Tournaments." *Operations Research* 33 (2): 249–62.
- Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*, 3rd ed. Hoboken, New Jersey: John Wiley & Sons.
- Jones, M. C. 1990. "The World Cup Draw's Flaws." *The Mathematical Gazette* 74 (470): 335–8.
- Koning, R. H. 2000. "Balance in Competition in Dutch Soccer." *The Statistician* 49 (3): 419–31.
- Koning, R. H. 2017. "Rating of Team Abilities in Soccer." In *Handbook of Statistical Methods and Analyses in Sports*, edited by J. Albert, M. E. Glickman, T. B. Swartz, and R. H. Koning, 355–71. Boca Raton: CRC Press.
- Laliena, P., and F. J. López. 2019. "Fair Draws for Group Rounds in Sport Tournaments." *International Transactions in Operational Research* 26 (2): 439–57.
- Lasek, J., Z. Szlávik, and S. Bhulai. 2013. "The Predictive Power of Ranking Systems in Association Football." *International Journal of Applied Pattern Recognition* 1 (1): 27–46.

- Lasek, J., Z. Szlávík, M. Gagolewski, and S. Bhulai. 2016. "How to Improve a Team's Position in the FIFA Ranking? A Simulation Study." *Journal of Applied Statistics* 43 (7): 1349–68.
- Ley, C., T. Van de Wiele, and H. Van Eetvelde. 2019. "Ranking Soccer Teams on the Basis of Their Current Strength: A Comparison of Maximum Likelihood Approaches." *Statistical Modelling* 19 (1): 55–73.
- Maher, M. J. 1982. "Modelling Association Football Scores." *Statistica Neerlandica* 36 (3): 109–18.
- McHale, I., and S. Davies. 2007. "Statistical Analysis of the Effectiveness of the FIFA World Rankings." In *Statistical Thinking in Sports*, edited by J. Albert, and R. H. Koning, 77–90. Boca Raton, Florida : CRC Press.
- Monks, J., and J. Husch. 2009. "The Impact of Seeding, Home Continent, and Hosting on FIFA World Cup Results." *Journal of Sports Economics* 10 (4): 391–408.
- Rathgeber, A., and H. Rathgeber. 2007. "Why Germany was Supposed to be Drawn in the Group of Death and Why it Escaped." *Chance* 20 (2): 22–4.
- Scarf, P. A., and M. M. Yusof. 2011. "A Numerical Study of Tournament Structure and Seeding Policy for the Soccer World Cup Finals." *Statistica Neerlandica* 65 (1): 43–57.
- Stone, C., and M. Rod. 2016. "Unfair Play in World Cup Qualification? An Analysis of the 1998–2010 FIFA World Cup Performances and the Bias in the Allocation of Tournament Berths." *Soccer and Society* 17 (1): 40–57.
- Stronka, W. 2020. "Anti-Tanking Pair Matching before an Elimination Phase of a Two-phase Tournament." *Economies* 8 (3): 66.
- Van Eetvelde, H., and C. Ley. 2019. "Ranking Methods in Soccer." In *Wiley StatsRef: Statistics Reference Online*, edited by R. S. Kenett, T. N. Longford, W. Piegorsch, and F. Ruggeri, 1–9. Hoboken, New Jersey: Springer.
- Winston, W. L. 2009. *Mathletics*. Princeton: Princeton University Press.
- Winston, W. L., and S. C. Albright. 1997. *Practical Management Science*. Belmont: Duxbury Press.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/jqas-2021-0052>).