Annals of Operations Research
https://doi.org/10.1007/s10479-025-06497-9

ORIGINAL RESEARCH

Draw procedures for balanced 3-team group rounds in sports competitions

Pablo Laliena[1] · **F. Javier López**[2]

Received: 9 April 2023 / Accepted: 13 January 2025

Abstract

We propose procedures for draws in sports tournaments that will achieve balanced groups of size 3. We start by developing an efficient algorithm that finds all possible balanced groupings; that is, allocations of teams to groups such that the strength of all groups lies between two bounds. The algorithm uses comparison results to discard large sets of unfeasible allocations and backtracking. We then define pot-and-ball draws that randomly pick one of the balanced groupings. Our draw procedures ensure that all groups are of (virtually) identical competitive levels, thereby enhancing the overall fairness and interest of the tournament.

Keywords Assignment problem · Backtracking algorithm · Integer programming · Sports

1 Introduction

Designing a sports tournament is a far from trivial problem. There are a variety of choices to be made, such as the structure (number of groups and/or knockout rounds), size of the groups, one/two leg confrontations, seeding of participants, the way in which participants are included in a group or matched in a knockout round and even the schedule of the matches. The overall aim of the tournament is to be as successful as possible. Although there is no established definition of a "successful tournament", the ingredients include having many interesting matches that attract large numbers of spectators and ensuring that some of the best participants play in the final rounds. There are also a number of constraints that must be respected. Aside from the obvious constraints of tournament duration and some logistic issues, the organizer will impose some other restrictions. These may include avoiding certain combinations of teams in the first rounds if, for instance, the opponents are from the same regional area. Moreover, the tournament should be fair, in the sense that better participants should have better chances of winning.

✉ F. Javier López
javier.lopez@unizar.es

Pablo Laliena
plaliena@iesmordefuentes.com

1 I.E.S. Mor de Fuentes, Av. Pueyo, 89, 22400 Monzón, Spain

2 Dpto. Métodos Estadísticos and BIFI, Facultad de Ciencias, Universidad de Zaragoza, C/ Pedro Cerbuna, 12, 50009 Zaragoza, Spain

Published online: 06 February 2025

While Operations Research has been used for decades to help organize good tournaments, there has been increasing interest in recent years in the use of analytical tools for the study and criticism of existing tournaments and the search for better designs (see Sect. 1.4 below).

Of all sports tournaments around the globe, the one that undoubtedly attracts most interest is the FIFA (Fédération Internationale de Football Association) Men's Soccer World Cup. Played every four years by national teams, it is the most-viewed sports tournament in the world; in 2022, it had a combined audience of around 5 billion viewers, with more than 1.5 billion spectators watching the final match between Argentina and France (FIFA, 2023c). Since 1950, the World Cup has generally had a group round with groups of 4 teams, followed by knockout rounds. The exceptions were the 1974, 1978 and 1982 editions, where different arrangements were adopted. The 2026 World Cup, to be held in Canada, Mexico, and the USA, will have 48 teams instead of the previous 32. Initially, there was a proposal for a group stage with 16 groups of three teams each, where two teams from each group would advance to the knockout stage (FIFA, 2017). This new organization of groups of 3 teams instead of 4 raised some concerns. For instance, there was a real chance of collusion in the last match of the group stage, where a specific result might lead to both contesting teams advancing, as discussed by Guyon (2020). Finally, in March 2023, FIFA decided on a format comprising 12 groups of 4 teams each (FIFA, 2023a), partly in response to concerns about collusion. The top two teams from each group, along with the eight best third-place teams, will progress to the knockout stage. This structure, however, introduces potential flaws such as group advantage, arbitrariness, and a lack of initiative, as highlighted by Guyon (2018b) in the analysis of the UEFA Euro 2016; see also Chapter 5 of Csató (2021b) and Guajardo and Krumer (2024).

FIFA's original decision to form 16 groups of 3 teams was our primary motivation for this paper. Although FIFA will not use this format for the 2026 World Cup, our work remains relevant as it can be applied to any sport and tournament featuring groups of 3 teams. Additionally, the solution we propose is valuable for future World Cups if FIFA considers the 3-team scheme again; it will become evident that a fair and balanced draw can be implemented. Moreover, after the results in Stronka (2024), which provide a method to greatly reduce the likelihood of collusion, the 3-team group option is indeed viable. It is important to note that a key advantage of this structure is that with $2 \times 16 = 32$ teams, the tournament maintains symmetry, avoiding the aforementioned issues inherent in the framework with 12 groups of 4 teams. Another advantage of the 3-group scheme is that there would be 80 matches in total instead of 104, with the maximal number of matches for a team being 7 instead of 8. This is important to have a shorter competition and avoid overstraining players, who already face a demanding game calendar, see (Guajardo and Krumer, 2024).

One of the aspects for which FIFA has been most criticized over the years is the draw to determine the groups for the opening stage. The draw employs a pot-and-ball arrangement, whereby the team names are placed in pots and then extracted sequentially to form the groups, in such a way that each group contains a team from each pot. Before the 2018 World Cup, the pots were formed using geographical criteria; thus, each group included teams from at least three different confederations (see, e.g., Cea et al. (2020)). This pot structure was unfair, since some teams had greater probabilities of ending up in a difficult group than other lower-ranking countries. In addition, the result was imbalanced, that is, there was a high degree of variability in the strength of the groups, a factor that was also regarded as unfair (see (Guyon, 2015)).

Following the recommendations of J. Guyon (see (Guyon, 2018a)), FIFA changed the draw system for the 2018 event. The pots are no longer filled on geographical criteria, but instead on the basis of the teams' relative rankings: the 8 teams occupying the highest positions in the world ranking are placed in the first pot; the next 8 highest-ranking teams are placed in the second pot, and so on, and the draw is done applying geographical criteria. While this has reduced the imbalance of the groups as compared to the previous arrangement, there are still major differences in the strength of the groups. For instance, in the 2022 edition, based on FIFA relative rankings as of March 31^{st}, 2022 used for the draw (FIFA, 2022), Group G, comprising Brazil (1), Switzerland (13), Serbia (21) and Cameroon (26) was a very strong group, with teams' rankings totaling 61, while Group H (Portugal (7), Uruguay (12), South Korea (23) and Ghana (32)) was much weaker, with the four rankings totaling 74. Furthermore, the draw was carried out when the identity of three teams was unknown, and they were mistakenly placed in the weakest pot (Csató, 2023); in particular, this gave rise to the strongest group: England (5), USA (14), Iran (18), Wales (16), totaling 53. The impact of imbalanced groups in FIFA World Cups has been analyzed by (Lapré and Palazzolo, 2022), Lapré and Palazzolo (2023). They demonstrate that even a small increase in the strength of group opponents can drastically decrease the probability of reaching the quarter-finals. Several proposals have appeared in the literature with draws that would reduce the differences in strength across the groups, e.g. Guyon (2015), Laliena and López (2019), Cea et al. (2020), but FIFA has not implemented any of them.

With 3-team groups, the problem of imbalance may be much more serious than in the 4-team arrangement. This is recognized, for instance, by Cea et al. (2020), who, speaking about the draw procedure for the group stage in the case of 3-team groups, state *"Yet the drawing procedure ought to change for the 2026 World Cup, with modifications not announced as of the time of this writing, thus research efforts in this respect remain relevant"*. For instance, if the draw uses three pots filled by team rank, then one group might be formed by the teams ranked 1, 17 and 33, totaling 51, and another group by the teams ranked 16, 32 and 48, totaling 96. One way of reducing this difference would be to use an "S-curve-type" constraint for the draw, similar to that suggested by Guyon (2015) for 4-team groups. In this S-curve arrangement, Pots 1, 2 and 3 would be split into Pots 1A, 1B, 2A, 2B, 3A and 3B, of 8 teams each, with the A-labeled sub-pots containing the strongest teams from their respective pot. The groups would then be formed by choosing one team from each of Pots 1A, 2B, 3A or one team from each of Pots 1B, 2A, 3B. A similar policy has been followed in the draw of the 2023 FIBA Basketball World Cup (FIBA, 2023). Even this arrangement might result in major differences between groups; for instance, two possible groups are 1-25-33, totaling 59 and 16-24-48 totaling 88.

1.1 Our goal

The goal of this paper is to propose draws that yield balanced groups of 3 teams. From here on, we assume that there are $3 \times g$ teams, where g is the number of groups, and each team has a score (based on its absolute or relative ranking in a list of teams or its Elo ranking) that is an (inverse) measure of its strength. By convention, we consider that lower scores correspond to stronger teams, which is consistent with the idea of using rankings. We also assume that the scores are integers and that two different teams cannot have the same score; the latter assumption may be dropped with a slight modification of the algorithm and draw procedures. If we use relative rankings as scores, each team will have an integer score of between 1 and $3g$; otherwise, each team will have an arbitrary positive integer score.

We also assume that there are g seeded teams, that is, each group contains one preassigned seeded team. Seeding is very common in sports tournaments, as a means of preventing top participants from being matched against one other in the opening rounds. For our draw, we assume that the best teams—those with the lowest scores—are seeded. Some tournaments include exceptions; for instance, the winner of the last tournament or the host team may be seeded, regardless of their strength. Nevertheless, it would be unusual for the reigning champion to be a weak participant and, in the case of the host team, playing all the matches at home would certainly give it some advantage (see, e.g., Clarke and Norman (1995)) and it would be reasonable to assign it a lower (i.e. stronger) score. We shall therefore maintain this assumption throughout the paper.

We define the overall score of a group as the sum of the scores of its three teams. The definition is quite reasonable and has been used, for instance, in Guyon (2015) and Cea et al. (2020) for groups of 4 teams. In Guyon (2015) and Laliena and López (2019), another measure was also considered when dealing with 4-team groups, namely the sum of the score of the three strongest teams in the group. The rationale is that 3 is one more than the number of teams that will qualify for the next round. See Remark 1 in Laliena and López (2019) for a discussion of this measure. Since in the present paper we stick to FIFA's initial setting of two teams qualifying for the next round, the two definitions—the sum of the scores of all teams and the sum of the scores of the 2+1 strongest teams—coincide. If only one team qualifies, the strength of the top two teams would probably be a better measure. Other measures, such as weighted sums of the strengths of the teams, have been considered in the literature; see Csató (2023).

Although our goal is to obtain balanced groups, sometimes an exact balance (i.e. all the groups having the same score), is not attainable. For instance, in the 2026 World Cup, with $g = 16$ and using relative rankings as scores, the sum of all scores is $48 \times 49/2 = 876$, which is not a multiple of 16, so it is not possible for all the groups to have exactly the same score. It is also likely that if we use other scores for the teams, such as absolute scores, it will not be possible to find groups totaling the same score. Another shortcoming of exact balance, when attainable, is that it substantially reduces the number of solutions and, consequently, the degree of randomness of the draw. Therefore, rather than imposing that all groups must have the same score, we set two bounds $\underline{s}$ and $\overline{s}$ and ask that all groups should have a score between these two values. In the particular case of using relative ranks with 48 teams, since $876/16 = 73.5$, we can set $\underline{s} = 73$ and $\overline{s} = 74$, but looser bounds are possible.

Mathematically speaking, the problem of finding all perfect solutions can be posed as follows. Let E be a set of $3g$ integers; find all the partitions of E in g sets of 3 elements such that the g lowest elements of E are in different subsets and the sum of the elements of each subset is between $\underline{s}$ and $\overline{s}$. The problem is similar to the classic equal-sum partition problem which looks for partitions of a set into subsets of equal sum. The latter problem is well-known in the combinatorics literature. See (Prodinger, 1982, 1984) and Cieliebak et al. (2008). Since in our case, each subset has already a fixed element (the seed), we can write the problem as the search for partitions of the set of the $2g$ largest elements of E into g subsets of cardinality 2, such that the sum of Subset i is between $\underline{s} - score(seed_i)$ and $\overline{s} - score(seed_i)$. With this in mind, our problem is close to the $k-$subset sum problem in Antonopoulos et al. (2021), which determines whether there exist k disjoint subsets with predefined sums. The aforementioned papers give algorithms to determine the existence of such solutions and, in some particular cases, results about the asymptotic behavior of the number of solutions. However, they do not give an algorithm to find all the partitions, which is our objective here. Note also that we do not want all groups to have exactly the same sum but to lie between two bounds.

Since there is a fixed element in each subset, the solutions to our problem are defined by choosing which other two integers are in each group. A trivial algorithm would be to enumerate all partitions of E in subsets of size 3 and retain those in which all the subsets add up between $\underline{s}$ and $\overline{s}$. However, even for moderate values of g, the number of possible partitions with the g lowest elements in different subsets is huge. For instance, if $g = 16$, each partition corresponds to organizing the 32 elements $\{17, \ldots, 48\}$ into 16 subsets of size 2. There are $\binom{32}{2}$ ways to choose the elements for first subset; for each of those, there are $\binom{30}{2}$ ways to choose the elements for the second subset, and so on. Therefore, the number of partitions is $\prod_{i=1}^{16} \binom{2i}{2} = \frac{32!}{2^{16}} \sim 4.01 \times 10^{30}$, indicating that this approach is unrealistic.

A better option is to write the problem as a binary linear problem, in a similar way to that used in Cea et al. (2020). While a single solution can be found by a computer in a few milliseconds, the complete set of solutions is much harder to find. In the particular case $E = \{1, \ldots, 48\}$ with $\underline{s} = 73$ and $\overline{s} = 74$, the number of partitions satisfying the conditions above is $\sim 3.11 \times 10^{11}$, but we failed to find more than 5×10^6 using this approach. Indeed, the left-hand panel of Table 1 shows the time needed by Gurobi 9.5.0 to find solutions running on an Intel(R) Core(TM) i5-6600 3.3 GHz. See also Fig. 1. Note that while the first solutions are found very quickly, the number of solutions per second decreases, and finding 4×10^6 solutions takes more than 55 h. Thus, in all likelihood, the complete set of 3.11×10^{11} solutions cannot be found using this approach.

In the mathematical problem stated above we have not considered that some partitions are not feasible solutions as groupings for the tournament because they do not satisfy the

Table 1 Time for finding solutions of the binary linear problem using Gurobi 9.5.0

Without geographical constraints			With geographical constraints		
Solutions	Time (sec)	Solutions/sec	Solutions	Time (sec)	Solutions/sec
10^4	4.6	2193.0	10^4	7.3	1373.6
10^5	108.8	919.5	10^5	141.2	708.1
10^6	12379.1	80.8	10^6	13996.2	71.4
2×10^6	50199.9	39.8	2×10^6	55692.3	35.9
3×10^6	116894.8	25.7	3×10^6	124528.3	24.1
4×10^6	198488.6	20.2	4×10^6	215852.8	18.5

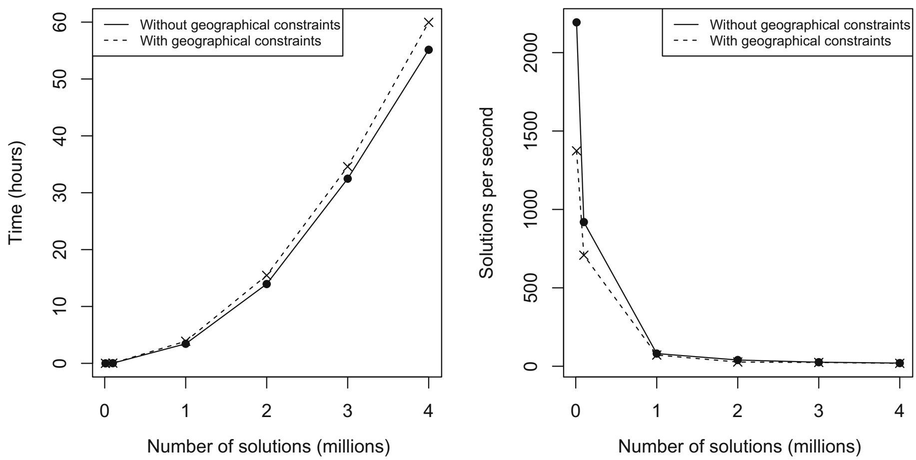

Fig. 1 Time for finding solutions of the binary linear problem using Gurobi 9.5.0

additional restrictions posed by the organizers, such as the geographical constraints imposed in FIFA World Cups. There are two ways of dealing with these restrictions. The first is to compute all the partitions, ignoring the restrictions, and later discard those which violate the constraints. The other is to include the restrictions when looking for the partitions. This second option can be implemented straightforwardly in the binary linear problem. As we show below, in the particular case of the 2026 FIFA World Cup, there are $\sim 3.13 \times 10^9$ out of 3.11×10^{11} solutions that satisfy the geographical constraints we consider in Sect. 1.3. Inclusion of the restrictions in the binary linear problem is not of great help in finding all the solutions by an optimizer, as the right-hand panel of Table 1 shows. The binary linear programming option must therefore be discounted.

1.2 Our proposal

In this paper, we start from the ideas of Laliena and López (2019) to develop a new algorithm that efficiently lists all perfect solutions., i.e., all groupings of teams that satisfy a balancedness condition, together with other restrictions imposed by the organizer, such as seeding or geographical constraints. Once all perfect solutions have been found, we present three different pot-and-ball draw procedures that will pick one of these solutions at random. Both the algorithm for listing all perfect solutions and the draw procedures are developed in a general context and we then show their application to a particular case (see Sect. 1.3 below for details).

The algorithm for finding all perfect solutions is given in Sect. 2 and has two parts. First, it uses comparison results to discard large subsets of partitions which cannot contain any perfect solution; it then uses backtracking to find all the solutions in the feasible region. Application of the algorithm to the 2026 FIFA World Cup is shown in Sect. 2.3. If the geographical constraints are not included, the number of perfect solutions is 311,229,091,808 and the algorithm takes slightly less than 1,000 h to find all of them on a Intel(R) Core(TM) i5-6600 3,3 GHz, or approximately 87,000 solutions per second. If the geographical constraints are included, the number of perfect solutions is 3,128,263,466 and the time taken by the algorithm to find them is around 24 h (37,000 solutions per second).

Once all perfect solutions have been found by the algorithm, a draw procedure must be performed to randomly pick one of them (Sect. 3). One possibility is to use a rejection sampler, as described in Roberts and Rosenthal (2024), which does not require the list of all perfect solutions. This method involves randomly generating groupings from the entire (unrestricted) set of possible groupings until one satisfies the restrictions. While this rejection sampler produces a grouping uniformly distributed over all perfect solutions, it is impractical for use—specially in live shows—if the conditions for valid groupings are highly restrictive. In such cases, the proportion of groupings that satisfy the constraints is very low, leading to a high number of unsuccessful attempts before finding a suitable solution.

A more realistic approach, using the list of all perfect solutions, consists of picking a number between 1 and the number of solutions. While this procedure is fair, since it assigns equal probability to each solution, it lacks the emotion of pot-and-ball draws of the kind used by FIFA and UEFA, which are broadcast live worldwide to a large audience. We therefore define three different draw procedures based on pots and balls, all of which extract one solution from the list of solutions found by the algorithm. These three procedures are similar to those that have been used by FIFA and UEFA for many years, and any of them could therefore be adopted by the tournament organizers in the future. We assess the fairness of the procedures using Monte Carlo simulation in the 2026 World Cup by comparing the estimated probability of pairs of teams being in the same group with the proportion of perfect solutions in which the teams are in the same group.

Since the present paper is directly related to our previous work (Laliena and López, 2019), we highlight here the main differences between the two papers:

- Group size and score. In Laliena and López (2019) the size of the groups was 4 and the score of a group was defined as the sum of the scores of the three strongest teams. In the present paper, groups are of size 3 and their score is defined as the sum of the scores of the three teams.
- Goal. In Laliena and López (2019) all groups were required to have the same score, while here the scores of the groups must fall within specified upper and lower bounds.
- Algorithm. When we applied the algorithm in Laliena and López (2019) to the case of 8 groups of 4 teams in the 2014 World Cup, it took $1'45''$ to find all perfect solutions. In that problem, the number of possible partitions of 32 teams into 8 groups with 8 seeded teams was $24!/3^8 \sim 9.5 \times 10^{19}$. We now want to build an algorithm for groups of 3 teams and apply it to the case of the 2026 World Cup with 16 groups. The number of possible partitions of 48 teams into 16 groups of 3 with 16 seeded teams is approximately 4.01×10^{30}. Thus, the algorithm here needs to be much more efficient in order to solve the problem in a reasonable amount of time. The present algorithm improves on Laliena and López (2019) in all its parts, from the conditions for discarding sets of partitions to the backtracking algorithm for finding particular solutions. We note that the ideas behind our algorithm can be translated to the setting of groups of 4 teams, yielding a much faster algorithm than that proposed there. In particular, the definition of a valid change (Definition 9 below) can be used to broaden the concept of an invalid change in Definition 5 of Laliena and López (2019). This, combined with a reformulation of their Proposition 2, would enable a much faster elimination of unfeasible allocations. Furthermore, the backtracking algorithm developed here could be applied as an alternative to their enumerative approach.
- Draw procedures. The pot-and-ball draw procedure proposed in Laliena and López (2019) for groups of size 4, called the 3-2-3 draw system, did not use the list of all perfect solutions and did not guarantee that the result was one of them. Rather, it was a heuristic algorithm which provided good results. In contrast, the draw procedures in the present paper guarantee that the result is a perfect solution.

1.3 The case of 2026 world cup

In order to show the application of our algorithm and drawing procedures to a particular case, we choose the original setting of the 2026 World Cup, where 48 teams had to be allocated to 16 groups. Although, as of March 2023, the groups will have a size of 4, we believe this example is valuable. It includes all the key elements of the general problem, such as seeded teams and geographical constraints, and it is familiar to most readers.

We use the FIFA ranking to define the strength of teams, and the score of a team will be its relative ranking (among the 48 participants). Other rankings, such as the World Football Elo Rating (available at http://eloratings.net/) have been used in the literature; see (Chater et al., 2021; Csató, 2023; Lapré and Palazzolo, 2023) or Stronka (2024). The use of absolute or relative rankings for defining the strength of a team in a tournament is controversial but we believe that relative rankings are better; as (Guyon, 2015) argues, the use of absolute FIFA rankings can be misleading since the ranks can vary from 1 to more than 200. The ranking we use is the one released on December 23, 2021, the last published when we started considering the problem (see FIFA, 2021).

Since the qualifying tournaments for the 2026 World Cup have yet to be held, the names of the teams that will be included are still not known. However, the number of teams from each geographical confederation has already been established, except for the last two places. These two places will be assigned in a playoff tournament played by one team from each confederation except UEFA (see FIFA (2018)). Thus, for this particular case, we take the highest ranked teams in each confederation to be included in the World Cup; for the last two places, we choose the best teams in confederations other than UEFA that have been left out. The list of participants using these criteria, together with their absolute and relative ranks are shown in Table B.1 of the Online Appendix.

The national teams from the three host countries will play in the World Cup (FIFA, 2023b), and we assume that, as in previous editions, they will be seeded. As commented above, we use relative ranks as scores; there is a slight change in the case of Canada, since it is a host country and therefore assumed to be seeded, but it is not among the 16 strongest teams. We set the score of Canada at 16 and the score of each team from Colombia to Nigeria as being equal to their respective relative ranks plus 1. The geographical constraints FIFA used in the last World Cups specify that no group can have two teams from the same confederation, except for UEFA, and that every group must have at least one and at most two UEFA teams. We adhere to these rules but remove the requirement of having at least one UEFA team per group. This adjustment is justified because there are exactly 16 UEFA teams, matching the number of groups, and enforcing this rule would significantly reduce randomness. We also applied our algorithm using the rankings released on December 22, 2022 and December 21, 2023; see Tables B.2−3 in the Online Appendix.

1.4 Literature review

Operations research (OR) has been widely applied to sports for decades. Probably the best overviews of OR applications in sports are provided by Wright (2009), Wright (2014), Csató (2021b), Lenten and Kendall (2022) and Devriesere et al. (2024). Recent research published in OR journals has focused on the comparison and analysis of tournaments design. Scarf et al. (2009) define several notions of success for a tournament and use simulations to evaluate different tournament arrangements, e.g. with vs without seeds, group vs knockout stages. Sziklai et al. (2022) compare tournament designs when the objective is to achieve the best ranking of the participants. Csató et al. (2025) analyze the fairness of the allocation of berths to each continent in FIFA World Cups. Csató (2021a) compares several arrangements for the group stage of handball tournaments by studying the number of interesting matches and the probability that good teams reach the final round. The role of seeding, which prevents certain participants from playing one another in early stages of the tournament, has been analyzed profusely. Scarf and Yusof (2011) investigate the effect of seeding policies for several tournament structures and illustrate their results in the FIFA World Cup; Dagaev and Suzdaltsev (2018) pose a discrete optimization problem to find a seeding that maximizes spectators' interest; Corona et al. (2019) use Monte Carlo simulation to study the effect of the different seeding arrangements in the UEFA Champions League.

One interesting aspect in the design of tournaments is their strategyproofness. A tournament is strategy-proof if participants cannot benefit from "tanking" (deliberately losing a game in order to gain some advantage). Csató (2020) and Csató (2022) show that the UEFA Champions League and the European qualifying tournament for FIFA World Cup are not strategy-proof and give suggestions for avoiding this problem. A closely related problem is the risk of collusion, which may occur if a particular result of a game in the group stage classifies the two opponents for the next round. Chater et al. (2021) analyze this problem and propose several ways to minimize its occurrence, both when the groups are formed by 3 and by 4 teams. In groups of 3 teams, the best strategy is to leave the strongest team out of the final match. Guyon (2020) and Stronka (2024) also show that the risk of collusion is high for the 3-team scheme and give potential solutions to reduce it. In particular, the solution provided in Stronka (2024) could reduce the expected number of matches with high risk of collusion in FIFA's initial proposal for the 2026 World Cup from 5.5 to just 0.26.

Particular attraction has been paid to the issue of fairness of tournaments. Arlegi and Dimitrov (2020) define fairness as letting stronger participants have better opportunities to win the tournament. They give conditions for knockout tournaments to be fair and show real examples of fair and unfair tournaments. Della Croce et al. (2022) use nonlinear mathematical programming to define fair Grand Slam tennis tournaments, where fairness is related to the number of times unseeded participants have been paired with seeded players in the first round of recent tournaments. Boczoń and Wilson (2023) focus on the draw system for knockout stages of a tournament and declare a draw procedure to be fair if teams have similar probabilities of being matched with any potential partner. They consider the particular case of the UEFA Champions League and quantify the distortions in match likelihoods caused by the constraints imposed by UEFA. These distortions were previously analyzed for earlier seasons of the tournament by Kiesl (2013) and Klößner and Becker (2013).

Fairness of draw procedures is also related to uniform distribution, that is, all the valid results are equally likely. Roberts and Rosenthal (2024) argue that the sequential draw procedures used by UEFA and FIFA for group stages are not equally distributed. Actually, there seems to be no straightforward sequential procedures based on pots and balls where all the solutions are equally likely. They propose several sequential methods to extract a solution uniformly based on multiple balls schemes or Metropolis algorithms. Csató (2024b) also analyzes the problem of unequal probability of the outcome of the UEFA draw procedures. He proposes several easy-to-implement modifications, such as changing the order of the pots, so that the results are closer (although not equal) to the uniform distribution.

Closely related to the present work are papers in which the concept of fairness in a group stage is based on balance, i.e., all the groups have similar strength. Imbalanced groups are seen as being unfair, because teams in stronger groups have lower probabilities of advancing to the next stage; for instance, Lapré and Palazzolo (2022), Lapré and Palazzolo (2023) quantify the effect of imbalance in both the Women's and Men's FIFA World Cup. Guyon (2015) proposes a draw procedure which uses an "S-curve-type" constraint to achieve balanced groups. Although the procedure does not guarantee that all groups have the same ranking total, it far outperforms FIFA's procedure. Laliena and López (2019) study the formation of 4-team groups and build an algorithm which finds all the possible groupings where all the groups have the same score, defined as the sum of the scores of the three best teams. Once all the possible groupings are found, one of them is selected using a random number generator. They also devise an alternative draw procedure based on pots and balls which produces groups with a similar score. Cea et al. (2020) propose the use of linear integer programming to find balanced groups. They minimize the difference between the maximum and the minimum range of the groups subject to the difference between the maximum and minimum ranking sums being lower than a fixed value. They propose solving the problem several times and picking one solution randomly.

The paper is organized as follows. The algorithm for listing all perfect solutions is described in Sect. 2: the first part of the algorithm, which finds all valid configurations together with its mathematical justification, is in Sect. 2.1; the second part, which consists of finding all the solutions arising from each valid configuration, is given in Sect. 2.2; an example of the application of the algorithm to the particular case $E = \{1, \ldots, 48\}$ is provided in Sect. 2.3. Section 3 is devoted to the proposal of draw procedures, which randomly pick one of the perfect solutions (Sects. 3.1, 3.2) and their comparison (Sect. 3.3).

2 A fast algorithm for enumerating all perfect solutions

Our goal is to devise a fair draw which produces groups with similar scores, defined as the sum of the individual scores of the three teams in the group. To this end, we will proceed as in Laliena and López (2019), by first enumerating all the possible groupings which fulfill certain conditions, namely, the seeding and geographical constraints and a balance condition. Then, a draw will be used to pick one of the perfect solutions. In this section we build an algorithm to find all perfect solutions. First we establish some theoretical results which reduce substantially the number of groupings to be checked. Our results improve on those of Laliena and López (2019), finding sharper conditions which allow many more groupings to be discarded in advance.

2.1 The first part of the algorithm: finding all valid configurations

Recall that the cardinal of E is $3g$, which is the number of teams that must be accommodated to g groups of 3 teams each. Since we are assuming a seeding rule, the g smallest elements of E must be in different groups. We begin with the definition of grouping:

Definition 1 A *grouping* $\mathbf{x} = (x_1, \ldots, x_g)$ is a partition of E in g 3-tuples, $x_j = (\ell_j, m_j, n_j)$, for $j = 1, \ldots, g$, such that:

(i) $\ell_1 < \cdots < \ell_g$ are the g smallest elements of E,
(ii) $m_j < n_j$, for $j = 1, \ldots, g$.

Each element x_j of $\mathbf{x}$ is a *group* and the *score* of Group x_j is defined as $s_j = \ell_j + m_j + n_j$, $j = 1, \ldots, g$.

Ideally, all groups should have the same score, but in many occasions, this is not possible since the sum of the scores of all the teams may not be a multiple of the number of groups g. More generally, we will search for draws such that the scores of all the groups are between two bounds. This leads to the following definition.

Definition 2 Given two integers $\underline{s} \leq \overline{s}$, a grouping is a $(\underline{s}, \overline{s})$-*perfect solution* if $\underline{s} \leq s_j \leq \overline{s}$, for all $j = 1, \ldots, g$.

In what follows, we assume that $\underline{s}$ and $\overline{s}$ are fixed. Therefore, to avoid overburden notation, we will write *perfect solution* instead of $(\underline{s}, \overline{s})$-*perfect solution* when there is no risk of confusion.

In a similar way to Laliena and López (2019), we divide set E into three subsets of size g so the groups will be formed by taking an element from each subset. Obviously, due to condition (i) of Definition 1, one of the subsets will be formed by the g smallest elements of E. The objective of the next definitions and results is to determine which of these partitions of E cannot define any perfect solution. This will reduce the number of the sets where perfect solutions must be searched for.

Definition 3 A *configuration* is a partition of E into three sets (B_1, B_2, B_3) with $card(B_1) = card(B_2) = card(B_3) = g$ and such that B_1 is formed by the g smallest elements of E. A configuration is *admissible* if there exists a grouping $\mathbf{x} = (x_1, \ldots, x_g)$ such that $\ell_j \in B_1$, $m_j \in B_2$ and $n_j \in B_3$, for all $j = 1, \ldots g$.

The set B_1 is formed by $\ell_1, \ldots, \ell_g$, the g smallest elements of E. This set is fixed throughout the paper and we assume $\ell_1 < \cdots < \ell_g$. Note that condition (ii) in Definition 1 may prevent a configuration from being admissible. Since B_1 is fixed, a configuration is admissible if and only if there is a matching between the elements of B_2 and B_3 such that each element of B_2 is smaller than the element of B_3. We will relate this property with Hall's marriage problem, see (Hall, 1935), in Lemma 1 below. The relationship between the marriage problem and draws in football has already been highlighted in Kiesl (2013) and Wallace and Haigh (2013).

Definition 4 Let (B_1, B_2, B_3) be an admissible configuration. We say that a grouping $\mathbf{x}$ is *generated* by (B_1, B_2, B_3) if $\ell_j \in B_1$, $m_j \in B_2$ and $n_j \in B_3$, for $j = 1, \ldots, g$.

Note that every grouping is generated by a unique configuration, namely, $B_1 = \{\ell_1, \ldots, \ell_g\}$, $B_2 = \{m_1, \ldots, m_g\}$, $B_3 = \{n_1, \ldots, n_g\}$. Therefore, in order to find all perfect solutions, we may find all admissible configurations and, for each one, all groupings it generates. We now define the concepts of basic configuration and change of teams.

Definition 5 The *basic configuration* is (B_1, B_2^0, B_3^0) where $\max(B_2^0) < \min(B_3^0)$.

That is, B_2^0 is formed by the g smallest elements of $E \backslash B_1$ and B_3^0 by the g largest elements of E. The basic configuration is admissible since it generates, for instance, the grouping with $x_j = (e_{(j)}, e_{(g+j)}, e_{(2g+j)})$, $j = 1, \ldots, g$, where $e_{(k)}$ is the k smallest element of E.

Definition 6 Let $r \in \{1, \ldots, g\}$. A *change* of r teams from the basic configuration (B_1, B_2^0, B_3^0) is $c = \big(\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\}\big)$, where $\{j_1, \ldots, j_r\} \subseteq B_2^0$ and $\{k_1, \ldots, k_r\} \subseteq B_3^0$. The *result of the change* c is the configuration (B_1, B_2^c, B_3^c) given by $B_2^c = B_2^0 \setminus \{j_1, \ldots, j_r\} \cup \{k_1, \ldots, k_r\}$ and $B_3^c = B_3^0 \setminus \{k_1, \ldots, k_r\} \cup \{j_1, \ldots, j_r\}$.

Definition 7 A change c is *admissible* if the configuration (B_1, B_2^c, B_3^c) is admissible.

A change is a swap of r teams between B_2^0 and B_3^0. For convenience we write $\emptyset$ for a change of 0 teams (that is, no change). The result of change $\emptyset$ is $(B_1, B_2^{\emptyset}, B_3^{\emptyset}) = (B_1, B_2^0, B_3^0)$. We now define a relation between changes. The relation will be defined explicitly for very similar changes and then extended to the general case by transitivity.

Definition 8 (a) Let $c = \big(\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\}\big)$ and $c' = \big\{(j'_1, \ldots, j'_r), (k'_1, \ldots, k'_r)\big\}$ be two changes such that either (i) there exists $\ell \in \{1, \ldots, r\}$ with $j'_\ell < j_\ell$, $j'_i = j_i$, for all $i \neq \ell$ and $k'_i = k_i$ for all $i \in \{1, \ldots, r\}$ or (ii) $j'_i = j_i$ for all $i \in \{1, \ldots, r\}$ and there exists $\ell \in \{1, \ldots, r\}$ with $k'_\ell > k_\ell$, $k'_i = k_i$, for all $i \neq \ell$. We say that c' is *stronger* than c (and write $c < c'$).

(b) Let $c = \big(\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\}\big)$ and $c' = \big\{(j'_1, \ldots, j'_{r+1}), (k'_1, \ldots, k'_{r+1})\big\}$ be two changes such that $\{j_1, \ldots, j_r\} \subset \{j'_1, \ldots, j'_{r+1}\}$ and $\{k_1, \ldots, k_r\} \subset \{k'_1, \ldots, k'_{r+1}\}$. We say that c' is *stronger* than c (and write $c < c'$).

(c) Let c and c' be two changes such that there exists a sequence of changes $c = c_1 < \cdots < c_n = c'$, with the order $<$ defined in (a) and (b). We say that c' is *stronger* than c (and write $c < c'$).

In the definition above, the order between changes is defined first (part (a)) for changes which are equal except that a team taken from B_2^0 in c is replaced by a smaller (better) team in c', or a team taken from B_3^0 in c is replaced by a higher (worse) team in c'. Then it is defined (part (b)) for changes which are equal except that change c' has one more element than change c. Last the relation is extended to other changes by using transitivity. It is clear that any change c can be reached from change $\emptyset$ with a sequence of changes of the type (a) and (b) such that $\emptyset = c_1 < \cdots < c_n = c$, so every change is stronger than the change $\emptyset$.

Lemma 1 *The configuration* (B_1, B_2, B_3) *is admissible if and only if*

$$card\left(p^{\uparrow}\right) \leq card\left(G\left(p^{\uparrow}\right)\right), \qquad \forall\, p \in B_2, \tag{1}$$

where, for $p \in B_2$, $p^{\uparrow} = \{k \in B_2 : k \geq p\}$ *and, for* $A \subseteq B_2$, $G(A) = \{k \in B_3\ \exists j \in A \ with\ j < k\}$.

Proof It is a consequence of Hall's marriage problem, see (Hall, 1935). By Definition 3, a configuration is admissible if and only if each $j \in B_2$ can be matched with a $k \in B_3$ with $k > j$. In other words, the configuration is admissible if and only if a marriage between B_2 and B_3 is possible, where each element j in B_2 can marry any element k of B_3 with $k > j$. Following Hall's theorem, the necessary and sufficient condition for the existence of a marriage is

$$card(A) \leq card(G(A)), \qquad \forall\, A \subseteq B_2. \tag{2}$$

It is obvious that if (2) holds, then so does (1), by taking $A = p^{\uparrow}$. Conversely, let $A \subseteq B_2$ and let $p = \min(A)$. Then, $card(A) \leq card\left(p^{\uparrow}\right) \leq card\left(G\left(p^{\uparrow}\right)\right) = card(G(A))$ and the result is proved. □

We now give necessary and sufficient conditions for a change to be admissible and a property to discard changes in our search for perfect solutions. For the next two results, without loss of generality, we assume that $E = \{1, \ldots, 3g\}$ and $B_1 = \{1, \ldots, g\}$, $B_2^0 = \{g+1, \ldots, 2g\}$, $B_3^0 = \{2g+1, \ldots, 3g\}$. We use the following notation: for a change $c = (\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\})$ let $y_p = card\,(\{i \in \{j_1, \ldots, j_r\} : i < p\})$ for $p = g+1, \ldots, 2g$, and $z_p = card\,(\{i \in \{k_1, \ldots, k_r\} : i \geq p\})$ for $p = 2g+1, \ldots, 3g$.

Proposition 1 *(a) The change* $c = (\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\})$ *is admissible if and only if the following conditions hold:*

(i) for all $p = g+1, \ldots, 2g$, $y_p \leq (p-g-1)/2$;
(ii) for all $p = 2g+1, \ldots, 3g$, $z_p \leq (3g-p+1)/2$.

(b) If c *and* c' *are two changes with* $c < c'$ *and* c *is not admissible, then* c' *is not admissible.*

Proof (a) We first show that c is admissible if and only if the following conditions hold:

(i') for all $p \in \{g+1, \ldots, 2g\} \setminus \{j_1, \ldots, j_r\}$, $y_p \leq (p-g-1)/2$;
(ii') for all $p \in \{k_1, \ldots, k_r\}$, $z_p \leq (3g-p+1)/2$.

In order to prove it, we apply Lemma 1 to $B_2^c = \{g+1, \ldots, 2g\} \setminus \{j_1, \ldots, j_r\} \cup \{k_1, \ldots, k_r\}$ and $B_3^c = \{2g+1, \ldots, 3g\} \setminus \{k_1, \ldots, k_r\} \cup \{j_1, \ldots, j_r\}$. Let $p \in \{g+1, \ldots, 2g\} \setminus \{j_1, \ldots, j_r\}$ and note that $p^{\uparrow}$ is formed by $\{k_1, \ldots, k_r\}$ and the elements in $\{g+1, \ldots, 2g\}$ which are greater than or equal to p and are not in $\{j_1, \ldots, j_r\}$. Thus

$card\left(p^{\uparrow}\right) = r + (2g - p + 1) - (r - y_p)$. Also, $card\left(G\left(p^{\uparrow}\right)\right) = g - y_p$. Then, condition (1) is

$$2g - p + 1 + y_p \leq g - y_p \Leftrightarrow y_p \leq \tfrac{p-g-1}{2}.$$

Let now $p \in \{k_1, \ldots, k_r\}$ and note that $p^{\uparrow}$ is formed by the elements of $\{k_1, \ldots, k_r\}$ greater than or equal to p, whose cardinal is z_p. We also have $card\left(G\left(p^{\uparrow}\right)\right) = 3g - p + 1 - z_p$. Thus, condition (1) is

$$z_p \leq \tfrac{3g-p+1}{2},$$

and we have proved that c is admissible if and only if (i') and (ii') hold. Since (i) and (ii) trivially imply (i') and (ii'), it only remains to prove the converse. Suppose then that (i') and (ii') hold. In particular, taking $p = \min\{k_1, \ldots, k_r\}$ we have

$$r = z_p \leq \frac{3g - p + 1}{2} \leq \frac{g}{2}. \tag{3}$$

Let now $p \in \{y_1, \ldots, y_r\}$. Suppose first $p < \max\left(\{g+1, \ldots, 2g\} \setminus \{j_1, \ldots, j_r\}\right)$ and define $p^+ = \min\{i \in \{g+1, \ldots, 2g\} \setminus \{j_1, \ldots, j_r\} : i > p\}$. We have $y_{p^+} = y_p + p^+ - p$; thus

$$y_p = y_{p^+} - (p^+ - p) < y_{p^+} - \frac{p^+ - p}{2} \leq \frac{p - g - 1}{2}$$

by (i'). Otherwise, if $p > \max\left(\{g+1, \ldots, 2g\} \setminus \{j_1, \ldots, j_r\}\right)$, we have $r = y_p + 2g + 1 - p$ so, from (3),

$$y_p = r - (2g + 1 - p) \leq r - \frac{2g + 1 - p}{2} \leq \frac{p - g - 1}{2}.$$

Let now $p \in \{2g+1, \ldots, 3g\} \setminus \{k_1, \ldots, k_r\}$. If $p > \max\{k_1, \ldots, k_r\}$ then $z_p = 0$. For $p < \max\{k_1, \ldots, k_r\}$ define $p^+ = \min\{i \in \{k_1, \ldots, k_r\} : i > p\}$; we have $z_p = z_{p^+} \leq (3\,g - p^+ - 1)/2 < (3\,g - p - 1)/2$.

(b) It suffices to prove that if c is not admissible, then any change c' as in parts (a) and (b) of Definition 8 is not admissible. Consider (a), and assume that the changes are of the form given in part (i), that is, there exists $\ell \in \{1, \ldots, r\}$ with $j'_\ell < j_\ell$, $j'_i = j_i$, for all $i \neq \ell$ and $k'_i = k_i$ for all $i \in \{1, \ldots, r\}$. In this case, we have $y_p(c) \leq y_p(c')$ for all $p \in \{g+1, \ldots, 2g\}$ and $z_p(c) = z_p(c')$ for all $p = 2g+1, \ldots, 3g$, so if the change c violates either condition (i) or (ii) of the Proposition, so does c'. An analogous proof works for the case $(a)(ii)$ of Definition 8. For (b), it is immediate that $y_p(c) \leq y_p(c')$ for all $p \in \{g+1, \ldots, 2\,g\}$ and $z_p(c) \leq z_p(c')$ for $p = \{2g+1, \ldots, 3g\}$ and the proof is complete.

□

Proposition 1 gives an easy way to check if a change is admissible. If a change c is not admissible, then the configuration (B_1, B_2^c, B_3^c) does not generate any solution nor does $(B_1, B_2^{c'}, B_3^{c'})$ with $c' > c$.

By (3), if (B_1, B_2^c, B_3^c) is an admissible configuration, then $r \leq g/2$, so we assume $r \leq g/2$ in the rest of the section. We now drop the assumption of $B_1 = \{1, \ldots, g\}$, $B_2^0 = \{g+1, \ldots, 2g\}$, $B_3^0 = \{2g+1, \ldots, 3g\}$; that is, E is an arbitrary set of $3g$ integers and (B_1, B_2^0, B_3^0) is the basic configuration as in Definition 5. Given a change $c = (\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\})$, we assume, without loss of generality that $j_1 < \cdots < j_r$ and $k_1 < \cdots < k_r$.

Definition 9 An admissible change $c = (\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\})$ is $(\underline{s}, \overline{s})-$*valid* if, for all $n = 1, \ldots, r$, the following conditions hold:

$$\sum_{i=g-n+1}^{g} \ell_i + \sum_{i=1}^{n} j_i + S(\{n\ largest\ elements\ of\ B_2^0 \setminus \{j_1, \ldots, j_r\}\ smaller\ than\ j_n\}) \geq n\underline{s}, \tag{4}$$

$$\sum_{i=1}^{n} \ell_i + \sum_{i=r-n+1}^{r} k_i + S(\{n\ smallest\ elements\ of\ B_3^0 \setminus \{k_1, \ldots, k_r\}\ greater\ than\ k_{r-n+1}\}) \leq n\overline{s}, \tag{5}$$

where $S(A)$ stands for the sum of the elements of A. We say that an admissible change is $(\underline{s}, \overline{s})-$*invalid* if it is not $(\underline{s}, \overline{s})-$valid. A configuration (B_1, B_2^c, B_3^c) is $(\underline{s}, \overline{s})-$*valid* if the change c is $(\underline{s}, \overline{s})-$ valid.

By convention, the change Ø is valid and the configuration (B_1, B_2^0, B_3^0) is valid.

Remark 1 The sets inside the S symbol in Definition 9 are well defined since the change c is admissible. As above, in the rest of the paper we will omit the prefix $(\underline{s}, \overline{s})-$ when speaking about valid or invalid changes when there is no risk of confusion.

Proposition 2 *If an admissible change c is invalid, then the configuration* (B_1, B_2^c, B_3^c) *does not generate any perfect solution.*

Proof Suppose that c is invalid due to a violation of (a) in Definition 9 for some n. Let $\mathbf{x}$ be a grouping generated by the configuration (B_1, B_2^c, B_3^c). In the groups where the teams $j_1, \ldots, j_n \in B_3^c$ are, the teams of B_2^c must have been chosen from the set $B_2^0 \setminus \{j_1, \ldots, j_r\}$ which are smaller than j_n, due to condition (ii) in Definition 1. An upper bound for the sum of the scores of these n groups is achieved by choosing the n largest elements of the set above together with the n largest elements of B_1. This upper bound is the LHS of (4), so, as the condition is violated, the upper bound does not reach $n\underline{s}$, so it is impossible that all n groups have scores at least $\underline{s}$. The proof for condition (b) in Definition 9 is analogous. □

Proposition 3 *Let* $c < c'$ *be two admissible changes. If c is invalid, then* c' *is invalid.*

Proof We show it for $c < c'$ in the setting of (a) and (b) in Definition 8 and the result for the rest of changes follows from transitivity.

Suppose the changes c and c' are as in condition $(a)(i)$ of Definition 8. We consider two particular choices for c and c'; the rest of the cases can be proved by transitivity. In the first choice we assume that, for some $q \in \{2, \ldots, r\}$, there is a gap between j_{q-1} and j_q (they are not consecutive values of the set B_2^0) and change c' is equal to change c except for the element j_q, which is included in change c and it is substituted by $p(j_q)$ in c', where $p(j_q)$ is the maximum of the elements in B_2^0 which are smaller than j_q. Note that $p(j_q)$ is not equal to j_{q-1}. Thus, in this case the changes are $c = (\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\})$ and $c' = (\{j_1, \ldots, j_{q-1}, p(j_q), j_{q+1}, \ldots, j_r\}, \{k_1, \ldots, k_r\})$. In the second choice, $q = 1$ (note that y_1 cannot be the minimum of B_2^0 since c is admissible) and change c' is equal to change c except for the element j_1, which is included in change c and it is substituted by $p(j_1)$.

Let us see how the LHS of (4) is modified from c to c'. If $n < q$, then $j'_1 + \cdots + j'_n = j_1 + \cdots + j_n$ and the set of elements in $B_2^0 \setminus \{j'_1, \ldots, j'_r\}$ smaller than j'_n is the same as the set of elements in $B_2^0 \setminus \{j_1, \ldots, j_r\}$ smaller than j_n, so the LHS of (4) is not changed. If $n \geq q$, then $j_1' + \cdots + j_n' = j_1 + \cdots + j_n - j_q + p(j_q)$ and the set of elements in $B_2^0 \setminus \{j_1', \ldots, j_r'\}$ smaller than j_n' is the same as $B_2^0 \setminus \{j_1, \ldots, j_r\}$ smaller than j_n except for, perhaps (depending on the particular values of $j_1, \ldots, j_r, n$ and q), the swapping of $p(j_q)$ for j_q. Thus, the LHS of (4) for c' is either equal to the LHS for c or it is decreased by the difference $j_q - p(j_q)$. Obviously, the LHS of (5) is not modified. Therefore, if c is invalid, then c' is also invalid. A similar proof applies for c and c' in the conditions of $(a)(ii)$ of Definition 8.

Suppose now that c and c' are as in the conditions of (b) of Definition 8. Let $c = (\{j_1, \ldots, j_r\}, \{k_1, \ldots, k_r\})$ and $c = (\{j_1', \ldots, j_{r+1}'\}, \{k_1', \ldots, k_{r+1}'\})$ and let q be the index of the element in $\{j_1', \ldots, j_{r+1}'\} \setminus \{j_1, \ldots, j_r\}$. Then,

$$j_1' = j_1 < \cdots < j_{q-1}' = j_{q-1} < j_q' < j_{q+1}' = j_q < \cdots < j_{r+1}' = j_r.$$

The LHS of (4) is unchanged when moving from c to c' if $n < q$. When $n \geq q$, we have

$$j_1' + \cdots + j_n' = j_1 + \cdots + j_{q-1} + j_q' + j_q + \cdots + j_{n-1} < j_1 + \cdots + j_n.$$

Also,

$$\left\{i \in B_2^0 \setminus \{j_1', \ldots, j_r'\} \text{ smaller than } j_n'\right\} \subseteq \left\{i \in B_2^0 \setminus \{j_1, \ldots, j_r\} \text{ smaller than } j_n\right\}$$

so the third term in the LHS of (4) for c' is smaller than or equal to the corresponding one for c. Therefore, if change c does not satisfy condition (4) neither does c'. A similar reasoning works for (5). □

Once we have established the theoretical results that allow us to discard many groupings (all those not coming from a valid configuration), we describe the first step of the algorithm.

This first step identifies all the valid configurations (B_1, B_2, B_3) using the results above. Recall that every configuration can be written as (B_1, B_2^c, B_3^c) for a certain change c, and all the changes can be reached from Ø using an increasing sequence of changes as in parts (a) or (b) of Definition 8. Also, if a change is not admissible or invalid, then all changes greater than it are also not admissible or invalid. The algorithm starts with configuration (B_1, B_2^0, B_3^0), which is admissible and (by convention) valid, and starts by checking which changes of one element $(\{j_1\}, \{k_1\})$ are admissible and valid. We begin with $j_1 = \max(B_2^0)$, $k_1 = \min(B_3^0)$ and continue by decreasing j_1 and increasing k_1. Note that if a change $(\{j_1\}, \{k_1\})$ is not admissible or invalid, there is no need to check any change $(\{j_1'\}, \{k_1'\})$ with $j_1' \leq j_1$ and $k_1' \geq k_1$. Once we have identified all the admissible and valid one-element changes, we turn to the two-element changes $(\{j_1, j_2\}, \{k_1, k_2\})$ such that all the changes $(\{j_1\}, \{k_1\})$, $(\{j_2\}, \{k_1\})$, $(\{j_1\}, \{k_2\})$, $(\{j_2\}, \{k_2\})$ are admissible and valid. This procedure is repeated with three, four,... elements with a maximum of $g/2$. Of course, if there are no admissible and valid changes of r elements, then there is no need to check the changes of more than r elements.

2.2 The second step of the algorithm: finding all perfect solutions within each valid configuration

Having identified all valid configurations, we have to find, for each one, all perfect solutions, that is, all groupings where each group has a score between $\underline{s}$ and $\overline{s}$. Note that different configurations imply different groupings, so there is no risk of getting duplicated groupings. Note also that, while no perfect solution can be obtained from a configuration which is not valid, there is no guarantee that a valid configuration generates any perfect solution.

Let (B_1, B_2, B_3) be a valid configuration and let us see how to find all perfect solutions it generates. First, and without loss of generality, we assume that the i-th element of B_1 is in Group i, $i = 1, \dots, g$. One possible way of finding all perfect solutions is to enumerate all possible assignments of elements of B_2 and B_3 to the groups and check whether the resulting groups have a score between $\underline{s}$ and $\overline{s}$. However, if we are considering $g = 16$, the cardinals of B_2 and B_3 are 16, so we need to check $16! \times 16! \sim 1.4 \times 10^{26}$ possibilities. Although many of these are not actually groupings, due to condition (ii) of Definition 1, we cannot identify them in advance. Thus, the direct enumeration procedure is discarded; this contrasts with the situation in Laliena and López (2019), where a much lower number of possibilities arose, namely $(8!)^3 \sim 6.6 \times 10^{13}$.

Our algorithm to find all perfect solutions generated by a valid configuration is based on backtracking. We need a preliminary step.

2.2.1 A preliminary step

Note that, if $\underline{s} = \overline{s}$, then all groups must have a score equal to $\underline{s}$, but if $\underline{s} < \overline{s}$, then the groups may have different scores, with the constraint that the sum of all scores must be equal to $S(E)$, the sum of the elements of E. So the preliminary step consists of computing, by direct enumeration or dynamic programming, all possible decompositions of $S(E)$ into g integer values between $\underline{s}$ and $\overline{s}$. Once this is done, for each decomposition we assign the terms to the groups and, for each assignment, we run the backtracking algorithm below.

2.2.2 The backtracking algorithm

In this step, we have B_2, B_3 and the scores of all groups fixed (s_i, $i = 1, \dots, g$). Thus, a solution will be defined by the assignment of each element of B_2 to a Group $i = 1, \dots, g$, since the element of B_3 in the group will be the one needed for the score of the group to be equal to s_i. We assume that the elements of B_2 are arranged in increasing order. We define an array with g rows representing the groups and g columns representing the elements of B_2. The element (i, j) of the array is the element of B_3 such that, if its score is added to ℓ_i (the score of the i-th team of B_1) and the score of the j-th team of B_2, it yields s_i, so long as the element of B_3 is greater than the j-th element of B_2, in order to fulfill condition (ii) in Definition 1. If no such element of B_3 exists, then the element of the array is 0. Here, we can include the geographical constraints: if Group i formed by the i-th team of B_1, the j-th team of B_2 and the corresponding team of B_3 whose score adds up to s_i violates a geographical constraint, then the element of the array is 0. In Laliena and López (2019), the geographical constraints were not included in the algorithm, which searched for all perfect solutions and checked one by one whether they satisfied the constraints. Although we can proceed in the same way here, the algorithm will be much faster if we include the constraints in this step, since it will allow us to discard many unfeasible solutions directly.

Once the array is built, our problem is similar to the classic n-queens problem. If we think of the array as a $g \times g$ board, such that there is a number (the value in the array) on each square, we must place g queens on the board. Placing a queen on square (i, j) means that Group i will be formed by the i-th element of B_1, the j-th element of B_2 and the corresponding team of B_3. The following constraints apply:

1. No two queens can be in the same row (only one team of B_2 per group),
2. No two queens can be in the same column (no team of B_2 can be in more than one group),

3. Queens can only be placed on squares with a nonzero value (there must be a compatible team of B_3 such that the group formed by the i-th team of B_1, the j-th team of B_2 and the team of B_3 has score s_i and satisfies the geographical constraints),
4. All the queens must be placed on squares with different values (no team of B_3 can be in more than one group).

Our backtracking algorithm works very much like the one for the n-queens problem. It goes as follows. Select the first row and place a queen on the leftmost available square (a square with a nonzero value). Then let the values of all squares in the row and column and in all squares that share the same value with the placed queen, be equal to zero. Proceed to the next row and repeat the procedure; if on reaching the end of the board we have placed g queens, then we have found a perfect solution. In such case, or if we have not reached the end of the board, move one step (row) back and select the next available (to the right) column to place the queen. When there are no more squares in which to place the queen in the first row, we will have obtained all perfect solutions in the array. The differences with the classic n-queens problem are: here there are some forbidden squares (those with a zero value) and the condition about the diagonals in the n-queens problem is substituted by the condition of squares having different values. A detailed example of the algorithm in a simple case of 4 groups of 3 teams each is shown in Online Appendix A. It is worth pointing out that backtracking can be used to replicate the draw procedure in FIFA World Cup since the 2018 edition, as at each draw the group where the selected team is placed depends on the existence of a feasible arrangement from that moment on; see (Roberts and Rosenthal, 2024) and Csató (2024a).

2.3 The particular case $E = \{1, \ldots, 48\}$ with $\underline{s} = 73, \overline{s} = 74$

In this section we show how we applied our algorithm to the 2026 World Cup Finals, with $g = 16$ groups, using relative rankings as team scores, $\underline{s} = 73$, $\overline{s} = 74$ and the geographical constraints as detailed in Sect. 1.3.

We apply the first part of the algorithm to identify all valid configurations. In this case, $B_1 = \{1, \ldots, 16\}$, $B_2^0 = \{17, \ldots, 32\}$, $B_3^0 = \{33, \ldots, 48\}$. It is easy to see that the only admissible and valid one-element changes are $(\{j_1\}, \{k_1\})$ with $j_1 \in \{29, 30, 31, 32\}$ and $k_1 \in \{33, 34, 35, 36\}$. Indeed, for $j_1 = 29$, condition (4) is $16 + 29 + 28 \geq 73$, but for $j_1 = 28$, the condition cannot be satisfied since $16 + 28 + 27 \ngeq 73$. Similar computations work for k_1.

In the case of two-element changes, they must be formed by pairs of elements in $\{29, \ldots, 32\}$ and $\{33, \ldots, 36\}$. It is straightforward to show that all changes taking two elements from the first set and two elements from the second are valid, except for the changes where $\{j_1, j_2\} = \{29, 30\}$ or $\{k_1, k_2\} = \{35, 36\}$. For instance, if $\{j_1, j_2\} = \{29, 31\}$, condition (4) is

$$(n = 1) \quad 16 + 29 + 28 \geq 73; \qquad (n = 2) \quad 15 + 16 + 29 + 31 + 28 + 30 \geq 146.$$

However, for $\{j_1, j_2\} = \{29, 30\}$ condition (4) cannot be satisfied since

$$(n = 1) \quad 16 + 29 + 28 \geq 73; \qquad (n = 2) \quad 15 + 16 + 29 + 30 + 27 + 28 \ngeq 146.$$

Changes of three elements must be formed by three elements of $\{29, 30, 31, 32\}$ which do not include the pair $(29, 30)$ and three elements of $\{33, 34, 35, 36\}$ which do not include the pair $(35, 36)$. The valid changes are those where $\{j_1, j_2, j_3\}$ is either $\{29, 31, 32\}$ or $\{30, 31, 32\}$ and those where $\{k_1, k_2, k_3\}$ is either $\{33, 34, 35\}$ or $\{33, 34, 36\}$. Lastly, there are no four-element changes since the only candidate is ({29, 30, 31, 32}, {33, 34, 35, 36}) which is invalid because it includes {29, 30} and {35, 36}. Therefore, in the particular case $E = \{1, \ldots, 48\}$ with $\underline{s} = 73$ and $\overline{s} = 74$ there are $1 + 4 \times 4 + 5 \times 5 + 2 \times 2 = 45$ valid configurations.

Once we have all valid configurations, we proceed to the second step of the algorithm, which is carried out 45 times, one for each valid configuration.

Fix a valid configuration. All decompositions of $S(E) = 1176$ into the sum of 16 integers between 73 and 74 are given by 8 values of 73 and 8 values of 74. Therefore, the possible assignments are defined by the selection of 8 groups with score 73 and 8 groups with score 74. This can be done in $\binom{16}{8} = 1{,}024$ ways, so the backtracking part of the algorithm must be repeated 1,024 times, one for each assignment.

Each valid configuration and assignment of the s_i to the groups defines a different array for the backtracking algorithm. For instance, taking $B_2 = \{17, \ldots, 31, 33\}$, $B_3 = \{32, 34, \ldots, 48\}$, $s_i = 73$ for $i = 1, \ldots, 8$ and $s_i = 74$ for $i = 9, \ldots, 16$, the array is

	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	33
1	0	0	0	0	0	0	0	$\underline{48}$	47	46	45	44	0	0	0	39
2	0	0	0	0	0	0	0	47	0	$\underline{45}$	44	43	0	0	40	38
3	0	0	0	0	0	48	$\underline{47}$	46	45	44	43	42	0	40	0	0
4	0	0	0	0	48	0	46	45	$\underline{44}$	43	42	41	40	0	38	36
5	0	0	0	48	47	0	0	44	0	42	$\underline{41}$	40	0	38	37	0
6	0	0	48	47	46	45	44	43	42	41	0	39	38	37	0	$\underline{34}$
7	0	48	47	46	45	44	43	42	41	40	39	38	$\underline{37}$	0	35	0
8	48	47	$\underline{46}$	45	44	43	42	41	40	39	38	37	0	35	34	0
9	48	47	46	45	44	43	42	41	40	39	38	37	0	$\underline{35}$	0	0
10	47	46	45	44	0	42	41	40	39	38	0	$\underline{36}$	35	0	0	0
11	46	0	0	43	0	41	40	39	0	37	36	35	0	0	$\underline{32}$	0
12	45	44	43	$\underline{42}$	0	0	39	38	37	36	35	34	0	32	0	0
13	44	43	42	41	40	$\underline{39}$	38	37	36	35	34	0	32	0	0	0
14	$\underline{43}$	42	41	40	0	0	37	36	0	34	0	32	0	0	0	0
15	42	41	40	39	$\underline{38}$	0	36	35	34	0	0	0	0	0	0	0
16	41	$\underline{40}$	39	0	37	36	0	34	0	32	0	0	0	0	0	0

where the row numbers are the teams in B_1, the column numbers are the teams in B_2 and the numbers inside the array are the teams in B_3. Note that for nonzero values the row number i plus the column number j plus the value of a_{ij} add up to 73 for $i = 1, \ldots, 8$ and 74 for $i = 9, \ldots, 16$. The zeros in the array come from three different sources: (a) impossibility to meet the score: for instance, the element (1, 1) corresponds to Team 1 from B_1, Team 17 from B_2, so there is no value in B_3 such that the sum of scores is 73; (b) condition (ii) of Definition 1: for instance, the element (9,16) corresponds to Team 9 from B_1, Team 33 of B_2 so the element of B_3 to give a score of 74 would be 32, which is smaller than 33; (c) geographical constraints: for instance, the element (1,29) is 0 because it corresponds to Team 1 (Belgium), Team 29 (Morocco) and Team 43 (Côte d'Ivoire), having two countries from

Table 2 First solution found by the algorithm in the array defined by $B_2 = \{17, \ldots, 31, 33\}$, $B_3 = \{32, 34, \ldots, 48\}$, $s_i = 73$ for $i = 1, \ldots, 8$ and $s_i = 74$ for $i = 9, \ldots, 16$.

A: BEL-SRB-NZL	E: ARG-JPN-GHA	I: DEN-ALG-PAR	M: SUI-IRN-CMR
B: BRA-UKR-PAN	F: ITA-AUS-NGA	J: NED-POL-EGY	N: MEX-COL-CIV
C: FRA-PER-CHN	G: ESP-MAR-QAT	K: USA-TUN-KOR	O: CRO-SEN-CRC
D: ENG-CHI-JAM	H: POR-SWE-UAE	L: GER-WAL-MLI	P: CAN-URU-KSA

CAF. Underlined numbers are the first solution found by the backtracking algorithm, which is shown in Table 2. There are 65,594 perfect solutions arising from this array.

This procedure is repeated for the 1,024 assignments of s_i within the configuration, and the procedure is repeated for each valid configuration. The number of perfect solutions arising from each valid configuration is shown in Table 3. The numbers in the 'Changes' column of the table represent the teams that are swapped between B_2^0 and B_3^0. For instance, '29,33' means that team 29 (Morocco) is exchanged with team 33 (Australia), resulting in $B_2 = \{17, \ldots, 28, 30, 31, 32, 33\}$ and $B_3 = \{29, 34, 35, \ldots, 48\}$.

In the case of the rankings released in December 2022 and December 2023 (Tables B.2−3 of the Online Appendix), the number of perfect solutions are shown in Tables B.4−5. It is worth noting how small changes in the list of countries can cause large changes in the number of perfect solutions from each valid configuration; in particular, in the latter tables there are valid configurations which do not give rise to any perfect solution, because of the geographical constraints.

2.4 An improvement to the backtracking algorithm

Our choice of taking the rows as the groups $i = 1, \ldots, g$, in increasing order, and the elements of B_2 as the columns also in increasing order, is arbitrary and may affect the speed of the backtracking algorithm, especially due to the geographical constraints. As a general idea, an array with fewer nonzero values in the first rows will be faster to analyze than other with many nonzero values in the first rows. Consider this simplified example, with $g = 6$, such that the array is represented by

$$\begin{pmatrix} * & * & * & * & * & * \\ * & * & * & * & * & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 \\ * & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where the values $*$ are all different nonzero numbers. It is clear that the only possible solution is the placement of a queen on each of the elements $(i, 7 - i)$, $i = 1, \ldots, 6$. However, the backtracking algorithm will revise several options before finding this solution. If we had used a different ordering of the rows, things would have been different. For instance, by reversing the order, the array is

Table 3 Number of solutions from each valid configuration. The configuration is defined by the change of the corresponding teams from the basic configuration. The total number of perfect solutions is 3,128,263,466. Total run time: 85,219 sec (23.67 h)

Changes	Perfect solutions	Changes	Perfect solutions	Changes	Perfect solutions
-	487,236,414	29,31,33,34	3,154,100	30,32,33,34	90,784,948
29,33	38,081,689	29,31,33,35	1,705,287	30,32,33,35	48,280,340
29,34	32,197,651	29,31,33,36	685,203	30,32,33,36	17,056,834
29,35	17,848,952	29,31,34,35	855,801	30,32,34,35	25,669,030
29,36	8,301,349	29,31,34,36	271,731	30,32,34,36	8,192,323
30,33	136,347,600	29,32,33,34	27,088,748	31,32,33,34	149,393,763
30,34	109,353,398	29,32,33,35	18,376,037	31,32,33,35	66,922,893
30,35	50,360,271	29,32,33,36	7,539,607	31,32,33,36	27,800,922
30,36	21,142,659	29,32,34,35	6,776,789	31,32,34,35	49,020,624
31,33	182,390,167	29,32,34,36	2,581,899	31,32,34,36	16,406,078
31,34	184,041,647	30,31,33,34	21,484,003	29,31,32,33,34,35	683,156
31,35	72,011,648	30,31,33,35	8,728,652	29,31,32,33,34,36	214,913
31,36	32,500,884	30,31,33,36	3,313,621	30,31,32,33,34,35	2,904,035
32,33	470,400,619	30,31,34,35	4,513,047	30,31,32,33,34,36	914,749
32,34	423,478,776	30,31,34,36	1,399,884		
32,35	169,223,977				
32,36	80,626,748				

$$\begin{pmatrix} * & 0 & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & * & * & 0 \\ * & * & * & * & * & * \end{pmatrix}$$

and the algorithm directly checks the only possible solution (and no other options). Moreover, in our original version we decided to take B_1 as the rows, B_2 as the columns and B_3 as the elements in the array. However, we can exchange the roles of B_1, B_2 and B_3 so that the number of nonzero elements in the first rows of the array becomes small. This flexibility is analogous to choosing the ordering of the pots in the proposal of Csató (2024b) for the FIFA World Cup draw, designed to improve fairness. However, we change the roles of the pots only to speed up the search for perfect solutions; the original order is maintained for our draw proposals in Sect. 3.

We applied this improvement to our example of the FIFA 2026 World Cup. In this case, we chose B_2 as the rows, B_3 as the columns and B_1 as the numbers inside the array; moreover, the elements in B_2 are rearranged so the number of zeros in the rows is decreasing. Application of this modification in the backtracking algorithm resulted in a 73% decrease in the time needed to find all perfect solutions as compared to our original setting (B_1 as rows, B_2 as columns and B_3 inside, with rows arranged in the natural order).

3 Draw procedures

Once we have obtained all perfect solutions, we need to devise a draw procedure to randomly select one of them. Actually, in theory, there is no need to know all the perfect solutions in advance, because the classic rejection method, as described in Sect. 2.1 of Roberts and Rosenthal (2024), could be used. This involves drawing a number uniformly between 1 and the number of unrestricted groupings (meaning that we do not take into account the geographical constraints or the balancedness condition). Then, check if the selected grouping fulfills the requirements: if it does, we have the solution; otherwise, reject it and draw the grouping again. However, in our case, this method cannot be used as, out of 4.01×10^{30} possible solutions of the unrestricted draw, only 3,128,263,466 satisfy the conditions, that is, only one out of 1.28×10^{21} draws would be accepted.

Using the list of all perfect solutions, a simple draw would be to choose a random number from 1 to N, the number of perfect solutions. This can be achieved easily by choosing a number between 0 and 9 for units, tens, hundreds, and so on, and rejecting the final number if it is greater than N. We call this draw 'Uniform Draw'. It is true that with the Uniform Draw, all perfect solutions have the same probability of being selected so we can regard this draw as fair, but it lacks the appeal of draws in which teams, represented by balls, are drawn from pots in a globally broadcast spectacle. One way to add extra interest to this draw is to use the Metropolis (swap) method (see Sect. 5.3 of Klößner and Becker (2013), and Sect. 7.1 of Roberts and Rosenthal (2024)). It is an iterative method that starts with a perfect solution chosen uniformly. Then, at each step, it extracts two teams randomly from a pot; if the swap of the chosen teams results in another perfect solution, then the swap move is performed; otherwise, a new pair of teams is selected. This procedure is repeated until a pre-specified number of swap moves has been made. If the initial solution was drawn uniformly, the result of the procedure is also uniform over all perfect solutions.

However, the Metropolis (swap) method is very different from what FIFA has been doing so far. Moreover, in our particular setting, due to the stringent conditions defining the perfect solutions, only a few pairs of teams can be swapped from a perfect solution to result in another perfect solution. More specifically, the probability of a successful swap is smaller than $2/31$. This bound arises because, once the first team (with score i, say) is chosen, the second team must have a score of $i-1$ or $i+1$ to be a candidate for a successful swap. In practice, the probability is much smaller. For instance, from the perfect solution in Table 2, the only successful swaps are POL↔JPN, GHA↔MLI, MAR↔ALG, QAT↔CRC and SWE↔WAL—five swaps out of $\binom{32}{2} = 496$ possible two-team extractions. This results in approximately a 1% likelihood of a successful swap, meaning that many unsuccessful extractions of pairs of teams may be necessary before achieving a valid result. Therefore, we believe that a draw method closer to the classic FIFA draws will be more appealing to the general public.

We propose three different pot-and-ball drawing procedures that will extract one of the perfect solutions. In these procedures, a complete list of all perfect solutions is needed. We will describe the procedures in the particular case of $E = \{1, \ldots, 48\}$, $\underline{s} = 73$, $\overline{s} = 74$ and with the teams and geographical constraints set out in Sect. 1.3.

We compare our procedures by analyzing their closeness to the uniform distribution (the Uniform Draw), as this closeness is considered a measure of fairness (Csató, 2024a, b). Other measures of fairness, such as using estimations of the probability of teams advancing to the next stage, have been proposed in the literature (Csató, 2024a), but we do not consider them here.

3.1 Drawing procedure 1

For this procedure we only need one pot, with 32 balls labeled with the numbers $17, \ldots, 48$. Teams 1 to 16 are assigned to Groups A to P respectively. (A modification of this draw, which may increase its closeness to the uniform distribution, would be to choose the groups for teams 1 to 16 randomly.) Then the balls are extracted sequentially from the pot and each team extracted is included in the first "available" group at that moment, starting with Group A. Here an available group is a group with less than 3 teams and such that there is a perfect solution with all the previously extracted teams in their corresponding groups plus the newly extracted team in its group.

The practical implementation consists of extracting one ball and checking if it can be in Group A, by seeing whether there is at least one perfect solution in which the team is in Group A. If so, the team is included in Group A and all the perfect solutions in which the team is not in Group A are discarded. If there is no solution with the team in Group A, then we move to Group B, continuing until a group is found for the team. This procedure is repeated until the pot is emptied.

3.2 Drawing procedures 2A and 2B

We call the two procedures in this section Draw 2A and Draw 2B because the latter is only a slight variation of the former. The procedures are similar to that used heretofore by FIFA. We have three pots (1, 2 and 3) whose composition is explained below. Groups will be formed by one team from each pot in such a way that the score of the team from Pot 1 is smaller than that of the team from Pot 2 which is in turn smaller than that of the team from Pot 3. Pot 1 will be formed by teams 1 to 16, which will be in groups A to P respectively. The procedure requires two sequential draws: a draw to decide the composition of Pots 2 and 3 and a draw to form the groups from the pots.

The first draw yields a composition of Pots 2 and 3 which is compatible with the existence of perfect solutions. These pots will act as the sets B_2 and B_3 in Sect. 2.1. Therefore, the composition of the pots must be given by configurations (B_1, B_2, B_3) which are valid and generate perfect solutions. In our particular case, the valid configurations are those defined in Table 3. Note that, in this case, all valid configurations generate perfect solutions. The difference between Draw 2A and Draw 2B is only in the first draw defining the composition of Pots 2 and 3.

The first draw should pick one of the valid configurations. This can be done in a number of ways, but if we want to keep a pot-and-ball arrangement, one possibility is to introduce balls labeled from 29 to 36 in a pot and extract four of them sequentially. These four teams will be in Pot 2 together with teams $17, \ldots, 28$. Since not all possible 4-ball extractions are feasible (that is, yield a valid configuration according to Table 3), a slight modification must be included. It is easy to see that all extractions are feasible except for those where teams 29 and 30 are both in Pot 3 or teams 35 and 36 are both in Pot 2. Therefore, if at any extraction Ball 35 or 36 is selected, then the other ball should be removed from the pot for the remaining extractions; also, if after three extractions neither Ball 29 nor Ball 30 have been extracted, then only those two balls must be in the pot for the last extraction. This procedure, which uses pots and balls to define the composition of Pots 2 and 3 will be called Draw 2A.

Due to the nature of the draw defining the composition of Pots 2 and 3 in Draw 2A, all the possible compositions of the pots will have similar probabilities. This contrasts with the number of perfect solutions arising from the configurations (see Table 3). This draw procedure may be expected to favor some solutions, especially those coming from configurations with relatively few perfect solutions and so the procedure will be far removed from the Uniform Draw, where all perfect solutions are equally likely. One way to overcome this problem is to change the form in which Pots 2 and 3 are formed. Instead of using the pot-and-ball arrangement described above, we can choose a composition of the pots with a probability that is proportional to the number of perfect solutions in Table 3. We can do that using a random number generator. This procedure, which uses a random number generator to define the composition of Pots 2 and 3, will be called Draw 2B.

Once we have defined the composition of Pots 2 and 3, we move to the draw to form the groups. This draw has 16 repetitions, one for each group. At each repetition a group is formed in the following way. A ball i is extracted from Pot 1, determining the group to be formed (A for $i = 1$, B for $i = 2$ and so on). Then balls in Pot 2 that are not compatible with the extraction from Pot 1 are removed, i.e., the balls with teams which are not in the same group as Team i in any perfect solution. A ball j is extracted from Pot 2. Now all balls in Pot 3 with teams which are not in the same group together with i and j in any perfect solution are removed from Pot 3. Note that in the particular example we are working with, as $\underline{s} = 73$ and $\overline{s} = 74$, there will be only one or two balls remaining in Pot 3. A ball k is then extracted from Pot 3, thus forming the group with Teams i, j, k. We now move to the next repetition, for which all the balls, except for i, j, k are returned to their original pots and all perfect solutions where Teams i, j, k are not in the same group are discarded. This procedure is repeated until all groups are formed.

3.3 Comparison of the drawing procedures

We now compare the drawing procedures 1, 2A and 2B described in Sects. 3.1 and 3.2 with the Uniform Draw, in which all perfect solutions are equally likely. Ideally, we should compute the probability of each perfect solution under the drawing procedures and compare it with $1/N$, where N is the number of perfect solutions. This is not possible due to the large number of perfect solutions, and because of the existence of multiple ways of obtaining a specific solution, which makes the computations for each solution lengthy. We can therefore rely on Monte Carlo simulation. However, we cannot estimate the probability of each perfect solution, because this would require a very large number of simulation runs in order to approximate the probability by the frequency of occurrence of each one of the solutions (3,128,263,466 in this particular case).

Since a direct comparison of the probabilities of each solution with $1/N$ is not possible, we can compare the draws by using a proxy. A good option (see Csató (2024a), Csató (2024b)) is to compare the proportion of times each pair of teams are in the same group under a particular drawing procedure with the corresponding proportion using the Uniform Draw. For the Uniform Draw, this is done by listing all the perfect solutions. See Table C.1 in the Online Appendix; for instance Belgium and Paraguay are in the same group in approximately 11.2% of the perfect solutions. Note that for reasons of space, the entire array of pairs is not shown in the table but since it is a symmetric array, all values can be read in it. An estimation of the corresponding values for Draws 1, 2A and 2B is made using Monte Carlo simulation. For each draw, we have performed 50,000 simulation runs and computed the proportion of times each pair of teams are in the same group. The simulation error for a probability p is given by $\sqrt{p(1-p)/n}$ and it is bounded above by $1/(2\sqrt{n}) = 0.0024$. The results are in Tables C.2−4 of the Online Appendix.

We compare Draws 1, 2A and 2B with the Uniform Draw by graphical and numerical means. Figure 2 shows three scatterplots, one for each drawing procedure. Each point is a pair of teams represented by the proportion of times the pair is in the same group under the Uniform Draw (X-axis) and under Draws 1, 2A and 2B (Y-axis). The closer the points are to the line $y = x$, the more similar the draw is to the Uniform Draw. It is evident that Draw 2B is the best, followed by Draw 1 and, a long way behind, by Draw 2A. Additionally, see the heatmaps in Tables C.2−C.4 (inspired by Csató (2024a)), where green indicates that the value under a given draw is greater than under the Uniform Draw and red indicates that it is smaller, with darker tones implying larger differences. This is corroborated numerically, by computing the distance between the values of the arrays represented in Tables C.2−4 with the corresponding values in Table C.1 of the Online Appendix. We use, as a measure of dissimilarity between the arrays, the chi-square statistic, defined as

$$\sum_{i \sim j} \frac{(\text{prop times } i, j \text{ together in Draw } k - \text{prop times } i, j \text{ together in Uniform Draw})^2}{\text{prop times } i, j \text{ together in Uniform Draw}},$$

where $i \sim j$ means that Teams i and j are in the same group in at least one perfect solution, and k =1, 2A, 2B. The values of this statistic are 10.77 for Draw 1, 33.41 for Draw 2A and 1.25 for Draw 2B, corroborating the conclusion that Draw 2B is the closest to the Uniform Draw.

Other measures of dissimilarity between the matrices, such as the sum of the absolute or relative deviations and the maximum of the absolute differences have also been computed, all indicating the superiority of Draw 2B.

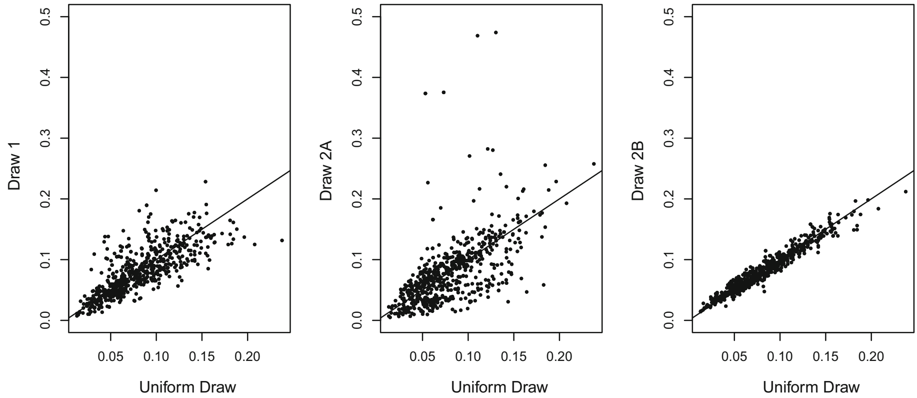

Fig. 2 Scatterplots comparing the probabilities of every pair of teams playing together in Draws 1, 2A and 2B with the Uniform Draw

3.4 Potential limitations of our procedures

Our algorithm provides the groupings that are nedeed to create balanced draws. The algorithm is fast and the draw procedures are well suited for a TV spectacle format. There are also some potential drawbacks to our proposal:

- The number of perfect solutions is sensitive to the geographical constraints and the choice of lower and upper bounds, which can be seen as arbitrary. Note, however, that the degree of freedom given by the choice of bounds lets the organizer adjust them to obtain an appropriate number of perfect solutions, based on the geographical constraints and the allocation of teams in the tournament.
- Enumerating all perfect solutions can be computationally challenging. Nevertheless, our algorithm is efficient, finding more than 10^9 solutions within 24 h. As the number of solutions depends on the chosen lower and upper bounds, these can be adjusted to prevent excessively long running times.
- Transparency could be compromised by the large number of solutions. The pot-and-ball draw procedures involve assigning teams to groups based on the existence of compatible perfect solutions, which requires a computer-assisted search. This process is not transparent for the spectators watching the show. To mitigate this issue, a complete list of perfect solutions and the corresponding code should be made publicly available in advance.

4 Conclusion

We have defined several draw procedures which give balanced solutions for the group stage of tournaments where groups are of size 3. First, a fast algorithm is given for obtaining a complete list of all perfect solutions (groupings where the sum of the scores of the teams of all groups are between two bounds and where all groups satisfy the constraints imposed by the organizer). This algorithm builds on that given in Laliena and López (2019) but it is much faster since it checks sharper conditions, in order to discard large subsets of infeasible groupings. One important aspect of having a list of all perfect solutions is that it can be refined by imposing further conditions. For instance, in consonance with (Cea et al., 2020), we can keep only those which minimize the difference between the ranges of the groups.

Once the complete list of solutions is obtained, several methods are defined for randomly picking one of them. The fairest method is to use a random number generator to obtain a number between 1 and the number of perfect solutions. This method is fair, since all the solutions have the same probability, but it is not exciting for the audience. We then provide three different drawing procedures based on the sequential extraction of balls from pots which maintain the traditional draw scheme used by FIFA and aim to align with transparency principles, a key issue in the draws of sports tournaments (see (Boczoń and Wilson, 2023; Roberts and Rosenthal, 2024)). The random parts of the draws are transparent because they involve the extraction of balls from pots during a broadcast event. Transparency is less clear for the deterministic steps of the draws–such as selecting which balls are placed in the pots for each extraction and determining the group for the extracted team–as they rely on the whole set of perfect solutions, which can be very large. This can be mitigated by publishing a list of all perfect solutions before the draw, along with the algorithm's code, allowing sophisticated users to replicate the procedure and confirm that they obtain the same solution given the balls extracted from the pots.

FIFA initially decided to use the 3-team scheme for the 2026 World Cup but later changed this due to concerns about collusion. However, the adopted scheme of 12 groups of 4 teams also has its problems. Stronka (2024) demonstrated how the likelihood of collusion can be reduced for groups of 3, and the present paper shows how balanced groups can be achieved. In light of these developments, we believe the 3-team scheme should be preferred for future World Cups.

Our draw procedures can be implemented in a similar way to that currently used by FIFA and other tournament organizers. In particular, they are very well suited for a TV spectacle format, and could therefore be adopted by FIFA. Of the three proposed methods, the one we call 2B is the least biased, since it gives probabilities that are the closest to the uniform case.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10479-025-06497-9.

Acknowledgements We are grateful to the referees for their thorough comments, which have improved the paper in many aspects. We also thank Tim Nicholson for proof-reading and copy-editing the manuscript.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was partially funded by grants PID2020-116873GB-I00 and TED2021-130702B-I00 of the Ministry of Science and Innovation. F. Javier López is a member of the research group Modelos Estocásticos of DGA (Aragón, Spain).

Declarations

Conflict of interest Both authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Antonopoulos, A., Pagourtzis, A., Petsalakis, S., & Vasilakis, M. (2021). Faster Algorithms for k-SUBSETSUM and Variations. *International Workshop on Frontiers in Algorithmics* (pp. 37–52). Cham: Springer International Publishing.

Arlegi, R., & Dimitrov, D. (2020). Fair elimination-type competitions. *European Journal of Operational Research, 287*(2), 528–535.

Boczoń, M., & Wilson, A. J. (2023). Goals, constraints, and transparently fair assignments: A field study of randomization design in the UEFA Champions League. *Management Science, 69*(6), 3474–3491.

Cea, S., Durán, G., Guajardo, M., Sauré, D., Siebert, J., & Zamorano, G. (2020). An analytics approach to the FIFA ranking procedure and the World Cup final draw. *Annals of Operations Research, 286*(1), 119–146.

Chater, M., Arrondel, L., Gayant, J. P., & Laslier, J. F. (2021). Fixing match-fixing: Optimal schedules to promote competitiveness. *European Journal of Operational Research, 294*(2), 673–683.

Cieliebak, M., Eidenbenz, S. J., Pagourtzis, A., & Schlude, K. (2008). On the complexity of variations of equal sum subsets. *Nordic Journal of Computing, 14*(3), 151–172.

Clarke, S. R., & Norman, J. M. (1995). Home ground advantage of individual clubs in English soccer. *Journal of the Royal Statistical Society: Series D (The Statistician), 44*(4), 509–521.

Corona, F., Forrest, D., Tena, J. D. D., & Wiper, M. (2019). Bayesian forecasting of UEFA Champions League under alternative seeding regimes. *International Journal of Forecasting, 35*(2), 722–732.

Csató, L. (2023). Group draw with unknown qualified teams: A lesson from the 2022 FIFA World Cup draw. *International Journal of Sports Science & Coaching*, 18(2), 539-551.

Csató, L. (2024a). The fairness of the group draw for the FIFA World Cup. arXiv preprint arXiv:2103.11353

Csató, L. (2024b). Fairness versus transparency: The optinal design of a randomised mechanism for constrained assignment arXiv preprint arXiv:2109.13785.

Csató, L. (2020). The UEFA Champions League seeding is not strategy-proof since the 2015/16 season. *Annals of Operations Research, 292*(1), 161–169.

Csató, L. (2021a). A simulation comparison of tournament designs for the World Men's Handball Championships. *International Transactions in Operational Research, 28*(5), 2377–2401.

Csató, L. (2021b). *Tournament Design: How Operations Research Can Improve Sports Rules. Palgrave Pivots in Sports Economics*. Cham, Switzerland: Palgrave Macmillan.

Csató, L. (2022). Quantifying incentive (in)compatibility: A case study from sports. *European Journal of Operational Research, 302*(2), 717–726.

Csató, L., Kiss, L. M., & Szádoczki, Z. (2025). The allocation of FIFA World Cup slots based on the ranking of confederations. *Annals of Operations Research, 344*, 153–173.

Dagaev, D., & Suzdaltsev, A. (2018). Competitive intensity and quality maximizing seedings in knock-out tournaments. *Journal of Combinatorial Optimization, 35*(1), 170–188.

Della Croce, F., Dragotto, G., & Scatamacchia, R. (2022). On fairness and diversification in WTA and ATP tennis tournaments generation. *Annals of Operations Research, 316*, 1107–1119.

Devriesere, K., Csató, L., & Goossens, D. (2024). Tournament design: A review from an operational research perspective. *European Journal of Operational Research, In press.* https://doi.org/10.1016/j.ejor.2024.10.044

FIBA (2023). The FIBA Basketball World Cup 2023 draw principles explained. https://www.fiba.basketball/basketballworldcup/2023/news/the-fiba-basketball-world-cup-2023-draw-principles-explained, Last accessed on 2024-07-16.

FIFA (2017). Unanimous decision expands FIFA World Cup to 48 teams from 2026. https://inside.fifa.com/about-fifa/organisation/fifa-council/media-releases/fifa-council-unanimously-decides-on-expansion-of-the-fifa-world-cuptm--2863100, Last accessed on 2024-07-15.

FIFA (2018). Guide to the bidding process for the 2026 FIFA World Cup. https://digitalhub.fifa.com/m/5730ee56c15eeddb/original/hgopypqftviladnm7q90-pdf.pdf, Last accessed on 2023-01-14.

FIFA (2021). Mens' ranking: 23 Dec 2021. https://www.fifa.com/es/fifa-world-ranking/men?dateId=id13505, Last accessed on 2022-06-30.

FIFA (2022). Mens' ranking: 31 Mar 2022. https://www.fifa.com/es/fifa-world-ranking/men?dateId=id13603, Last accessed on 2023-02-23.

FIFA (2023a). FIFA Council approves international match calendars. https://inside.fifa.com/about-fifa/organisation/fifa-council/media-releases/fifa-council-approves-international-match-calendars, Last accessed on 2024-07-15.

FIFA (2023b). FIFA Council highlights record breaking revenue in football. https://inside.fifa.com/about-fifa/organisation/fifa-council/media-releases/fifa-council-highlights-record-breaking-revenue-in-football, Last accessed on 2024-07-17.

FIFA (2023c). One Month On: 5 billion engaged with the FIFA World Cup Qatar 2022. https://inside.fifa.com/tournaments/mens/worldcup/qatar2022/news/one-month-on-5-billion-engaged-with-the-fifa-world-cup-qatar-2022-tm, Last accessed on 2024-07-15.

Guajardo, M., & Krumer, A. (2024). Tournament design for a FIFA World Cup with 12 four-team groups: Every win matters. In *The Palgrave Handbook on the Economics of Manipulation in Sport* (pp. 207-230). Palgrave Macmillan, Cham, Switzerland.

Guyon, J. (2015). Rethinking the FIFA World Cup™ final draw. *Journal of Quantitative Analysis in Sports, 11*(3), 169–182.

Guyon, J. (2018a). Pourquoi la Coupe du monde est plus équitable cette année. *The conversation*. https://theconversation.com/pourquoi-la-coupe-du-monde-est-plus-equitable-cette-annee-97948, Last accessed on 2024-07-29.

Guyon, J. (2018b). 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), 259–279.

Hall, P. (1935). On representatives of subsets. *Journal of the London Mathematical Society, 10*(1), 26–30.

Kiesl, H. (2013). Match me if you can-Mathematische Gedanken zur Champions-League-Achtelfinalauslosung. *Mitteilungen der Deutschen Mathematiker-Vereinigung, 21*(2), 84–88.

Klößner, S., & Becker, M. (2013). Odd odds: the UEFA Champions League Round of 16 draw. *Journal of Quantitative Analysis in Sports, 9*(3), 249–270.

Laliena, P., & López, F. J. (2019). Fair draws for group rounds in sport tournaments. *International Transactions in Operational Research, 26*(2), 439–457.

Lapré, M. A., & Palazzolo, E. M. (2022). Quantifying the impact of imbalanced groups in FIFA Women's World Cup tournaments 1991–2019. *Journal of Quantitative Analysis in Sports, 18*(3), 187–199.

Lapré, M. A., & Palazzolo, E. M. (2023). The evolution of seeding systems and the impact of imbalanced groups in FIFA Men's World Cup tournaments 1954–2022. *Journal of Quantitative Analysis in Sports, 19*(4), 317–332.

Lenten, L. J., & Kendall, G. (2022). Scholarly sports: Influence of social science academe on sports rules and policy. *Journal of the Operational Research Society, 73*(12), 2591–2601.

Prodinger, H. (1982). On the number of partitions of $\{1, \ldots, n\}$ into two sets of equal cardinalities and equal sums. *Canadian Mathematical Bulletin, 25*(2), 238–241.

Prodinger, H. (1984). On the number of partitions of $\{1, \ldots, n\}$ into r sets of equal cardinalities and sums. *Tamkang Journal of Mathematics, 15*(161–164), 1984.

Roberts, G. O., & Rosenthal, J. S. (2024). Football group draw probabilities and corrections. *Canadian Journal of Statistics, 52*(3), 659–677.

Scarf, P. A., & Yusof, M. M. (2011). A numerical study of tournament structure and seeding policy for the soccer World Cup Finals. *Statistica Neerlandica, 65*(1), 43–57.

Scarf, P., Yusof, M. M., & Bilbao, M. (2009). A numerical study of designs for sporting contests. *European Journal of Operational Research, 198*(1), 190–198.

Stronka, W. (2024). Demonstration of the collusion risk mitigation effect of random tie-breaking and dynamic scheduling. *Sports Economics Review, 5*, 100025.

Sziklai, B. R., Biró, P., & Csató, L. (2022). The efficacy of tournament designs. *Computers & Operations Research, 144*, 105821.

Wallace, M., & Haigh, J. (2013). Football and marriage-and the UEFA draw. *Significance, 10*(2), 47–48.

Wright, M. B. (2009). 50 years of OR in sport. *Journal of the Operational Research Society, 60*(sup1), S161–S168.

Wright, M. B. (2014). OR analysis of sporting rules-A survey. *European Journal of Operational Research, 232*(1), 1–8.