

Optimal tournament design: lessons from the men's handball Champions League

László Csató*

Institute for Computer Science and Control, Hungarian Academy of Sciences (MTA SZTAKI)
Laboratory on Engineering and Management Intelligence, Research Group of Operations
Research and Decision Systems

Corvinus University of Budapest (BCE)
Department of Operations Research and Actuarial Sciences

Budapest, Hungary

20th July 2020

Der Beweis, den wir fordern, ist überall nötig, wo der Vorzug des vorgeschlagenen Mittels nicht so evident ist, daß er keinen Zweifel zuläßt, und er besteht darin, daß jedes der beiden Mittel seiner Eigentümlichkeit nach untersucht und mit dem Zweck verglichen werde.¹

(Carl von Clausewitz: *Vom Kriege*)

Abstract

Many sports tournaments are organised in a hybrid design consisting of a round-robin group stage followed by a knock-out phase. The traditional seeding regime aims to create balanced groups roughly at the same competition level but may result in several uneven matches when the quality of the teams varies greatly. Our paper is the first challenging this classical solution through the example of the men's EHF (European Handball Federation) Champions League, the most prestigious men's handball club competition in Europe, which has used unbalanced groups between the 2015/16 and 2019/20 seasons. Its particular design is compared to an alternative format with equally strong groups, as well as to the previous scheme of the EHF Champions League. We find that it is possible to increase the quality of all matches played together with raising the uncertainty of outcome, essentially without sacrificing fairness. Our results have useful implications for the governing bodies of major sports.

JEL classification number: C44, C63, Z20

MSC class: 62F07, 68U20

Keywords: OR in sports; tournament design; handball; simulation; competitive balance

* E-mail: laszlo.csato@sztaki.mta.hu

¹ “The demonstration we require is always necessary when the superiority of the means propounded is not so evident as to leave no room for doubt, and it consists in the examination of each of the means on its own merits, and then of its comparison with the object desired.” (Source: Carl von Clausewitz: *On War*, Book 2, Chapter 5—Criticism. Translated by Colonel James John Graham, London, N. Trübner, 1873. <http://clausewitz.com/readings/OnWar1873/TOC.htm>)

1 Introduction

The choice of an appropriate format for sports tournaments poses an important question in economics and operations research because “*designing an optimal contest is both a matter of significant financial concern for the organisers, participating individuals, and teams, and a matter of consuming personal interest for millions of fans*” (Szymanski, 2003, p. 1137). In addition, any sports tournament can be considered a kind of selection mechanism with a number of managerial applications such as recruitment strategies (Ryvkin, 2010).

Many sports competitions are organised in a hybrid design consisting of a round-robin group stage followed by a knock-out phase. Examples include the [FIFA World Cup](#), the [UEFA Champions League](#), the [UEFA European Championship](#) (all in association football), the [FIBA Basketball World Cup](#), the [FIVB Volleyball World Championship](#), the [EHF European Handball Championship](#), or the [IHF World Handball Championship](#).

Since only the top teams from each group qualify for the next stage, the probability of advancing is strongly influenced by the opponents. Therefore, the allocation of the teams into groups is governed by certain rules, and almost all systems use seeds. The classical solution is to rank the contestants based on their past performance. Then, in the case of k groups, each group gets one team from the first pot of the best k teams, one team from the second pot consisting of the next k teams, and so on. Further considerations may play a role, for instance, clubs from the same national association could not be drawn against each other in the UEFA Champions League. Similarly, FIFA strives for creating geographically diverse groups in the World Cup.

The seeding procedure above aims to provide balanced groups roughly at the same competitive level, but may inevitably lead to several uneven matches when the quality of the teams varies greatly, which is the usual case. This can be against the interest of the administrators because higher contest quality and greater uncertainty of outcome are usually associated with higher attendance (Forrest and Simmons, 2002; Borland and MacDonald, 2003). Despite that, while there are some recent results on the draw procedures of the groups (Guyon, 2015; Laliena and López, 2019; Cea et al., 2020), we do not know any work challenging the traditional seeding regime.

Therefore, the current paper attempts to outline and evaluate an alternative to the well-established format of balanced groups. In particular, we will compare the design applied by the men’s handball [EHF Champions League](#) from the 2015/16 season with a traditional variant and the previous format of this tournament. Simulations show that the quality of all matches played can be increased and the uncertainty of outcome can be raised, essentially without sacrificing fairness. The latter is crucial because pre-tournament tanking cannot be excluded if strong teams prefer to be drawn into a weak group of a tournament containing unbalanced groups. Thus our results have useful implications for the governing bodies of major sports.

The paper has the following structure. Section 2 describes the three tournament designs to be analysed. The simulation experiment and the metrics used for the evaluation of the competition formats are presented in Section 3. Section 4 demonstrates the results and provides a robustness check. Finally, Section 5 summarises our main findings.

2 An instructive competition design

The [EHF Champions League](#), organised by the European Handball Federation (EHF) since the 1993/94 season, is the most prestigious men’s handball club competition of the

continent. The tournament uses a hybrid design mixing knock-out and round-robin stages. As home advantage is a well-documented phenomenon in this sport (Meletakos and Bayios, 2010; Lago-Peñas et al., 2013), teams usually play both at home and away against each other, with the exception of the Final Four, which takes place at the Lanxess Arena in Cologne, Germany from the 2009/10 season onwards.

On 21 March 2014, the 119th meeting of the EHF Executive Committee decided to introduce a new competition format from the 2015/16 season (EHF, 2014). The number of competing teams was increased in order to open up the tournament to more nations and new markets across the continent. The reform also guaranteed more top matches between the leading clubs for the sake of making the competition more attractive to spectators, sponsors, and the media. At the same time, it was ensured that all clubs playing in the group phase retain the chance to qualify for the Final Four.

This format of the EHF Champions League starts with 28 competing teams. They play in four round-robin groups, groups A and B with eight teams each, and groups C and D with six teams each such that:

- in groups A and B, the top team directly qualifies for the quarter-finals, the bottom two clubs are eliminated, while the remaining teams advance to the first knock-out phase;
- in groups C and D, the bottom four teams drop out of the tournament, while the top two teams in both groups contest a play-off to determine the two teams that advance to the first knock-out phase.

Consequently, the first knock-out phase involves 12 teams, five from group A, five from group B, and two from groups C and D. The six winners qualify for the quarter-finals, where they join the group winners of groups A and B. The winners of the quarter-finals participate in the Final Four.

Figure A.1 presents the tournament format in detail until the teams of the Final Four are selected. The design remains deterministic after the groups are drawn. For the semi-finals in the Final Four, there is a new draw with all teams being in the same pot.

This competition format will be denoted by $D(8 + 6)$ in the following.

Design $D(8 + 6)$ seems to be rather strange because the groups are treated differently. An axiomatic criterion of fairness can be *equal treatment of equals* (Palacios-Huerta, 2012; Brams and Ismail, 2018), that is, if all clubs are equally strong then each of them should have the same probability of being the final winner even *ex post*, after the drawing of the groups. However, in the case of format $D(8 + 6)$:

- a team from groups A and B will be the group winner with a probability of $1/8$ and will be eliminated with a probability of $1/4$, thus it qualifies for the quarter-finals with a probability of $1/8 + 5/8 \times 1/2 = 7/16$; while
- a team from groups C and D qualifies for the play-off with a probability of $1/3$, for the first knock-out phase with a probability of $1/6$, and for the quarter-finals with a probability of $1/12$.

Hence, a “lucky” team has an $84/16 = 5.25$ times higher chance to win the competition than another with the assumption of homogeneous teams. However, this unrealistic scenario has rather only a theoretical relevance since sports contests are usually unbalanced, which justifies an investigation via Monte-Carlo simulations.

To understand the main characteristics of format $D(8 + 6)$, two alternative designs will be considered.

First, we have devised a candidate with balanced groups and the same number of teams for comparative purposes. The alternative format $D(4 \times 7)$ is outlined in Figure A.2: the four group winners and the four runners-up qualify directly for the Round of 16, while the third-, fourth-, fifth-, and sixth-placed teams play against each other in the first knock-out phase, where the winners advance to the Round of 16.

Second, the previous format of the EHF Champions League, used between the seasons 2009/10 and 2014/15, is also included in the analysis.² It is shown in Figure A.3: there are four groups of six teams each, the top four clubs qualify for the Round of 16, where a standard knock-out stage starts with drawn brackets—its details will be discussed later. The design is denoted by $D(4 \times 6)$. Note that $D(4 \times 6)$ contains only 24 competitors, which might increase the probability of winning compared to the other two designs, but the latter are organised with more matches and the top teams can skip a knockout round.

3 Methodology

The three tournaments designs, the innovative $D(8 + 6)$, as well as the classical $D(4 \times 7)$ and $D(4 \times 6)$ with balanced groups, will be analysed via simulation techniques.

3.1 The simulation of match outcomes

Most numerical studies of tournament designs apply specific models for simulating match results (Scarf et al., 2009; Scarf and Yusof, 2011; Goossens et al., 2012; Lasek and Gagolewski, 2018; Corona et al., 2019; Dagaev and Rudyak, 2019), but we do not follow this approach due to several reasons.

First, general works comparing different competition formats (Appleton, 1995; McGarry and Schutz, 1997; Marchand, 2002) or ranking methods (Mendonça and Raghavachari, 2000) avoid the use of specific prediction models. Second, while there exists a number of such models for football matches (Maher, 1982; Dixon and Coles, 1997; Koning et al., 2003; Tutz and Schauburger, 2015), handball seems to be a more difficult sport with respect to forecasting since it is a fast, dynamic, and high-scoring game. Significant differences can be observed between the total number of goals scored per match across the leading men’s handball national leagues together with an increasing trend in all countries (Meletakos and Bayios, 2010). Furthermore, the dynamics of handball matches violate independence and identical distribution, sometimes showing a non-stationary behaviour (Dumangane et al., 2009). Unsurprisingly, Groll et al. (2020) is probably the first project to forecast handball results through a collaboration of professional statisticians and handball experts. Third, it is almost impossible to adequately address all issues influencing match outcomes. For instance, the schedule of round-robin tournaments may result in a substantial advantage for some contestants as recent analytical (Krumer et al., 2017a, 2020; Sahm, 2019) and empirical works (Krumer and Lechner, 2017) show. Similarly, even the kick-off time can affect various aspects of games such as the home advantage of the underdog team (Krumer, 2020).

Finally, the main message of the current paper is the consideration of competition designs with a non-traditional round-robin group stage. Since this may be relevant in

² We thank an anonymous referee for this suggestion.

other sports, it makes no sense to fit a particular prediction model on the results of the EHF Champions League matches because it would not contribute much to the general applicability of our suggestion.

Thus, the probability with which a given team would beat another team is fixed *a priori*, and it neither changes during the competition (stationarity) nor is influenced by the previous results (independence). While these conditions clearly do not hold in practice, they can offer a good approximation of long-run averages (McGarry and Schutz, 1997). In addition, a variety of models within reason could be taken to determine the winners for comparative purposes (Appleton, 1995).

Therefore, each match of a tournament is organised as a Tullock contest such that:

$$p_{ij} = \frac{(57 - i)^r}{(57 - i)^r + (57 - j)^r} \quad (1)$$

gives the probability of team i winning against team j , where $r \geq 0$ reflects the discriminatory power of the contest and $1 \leq i, j \leq 28$ is the *identifier* of the teams, which can be called their *pre-tournament rank*.

Consequently, draws are not allowed in any match, although it is not a rare event in handball: in the 2017/18 EHF Champions League, there were 10 in group A, 8 in group B (from 56 matches, respectively), as well as 0 in group C and 3 in group D (from 30 matches, respectively). This assumption is also relatively standard in theoretical papers comparing tournament formats (Appleton, 1995; McGarry and Schutz, 1997; Marchand, 2002; Csató, 2019a).

The formula $57 - i$ has been chosen as the effort of the teams because (1) it should be a non-decreasing function of the pre-tournament rank, and (2) the straightforward solution of $1 \leq 29 - i \leq 28$ would lead to a much higher competitive balance among the top teams compared to the competitive balance among the bottom teams, which is implausible. Throughout the paper, two different values of r ($= 3, 5$) will be studied to check the robustness of our results.

Naturally, teams ranked 25–28 are not considered in the design $D(4 \times 6)$.

Figure 1: The probability that team i beats its opponent

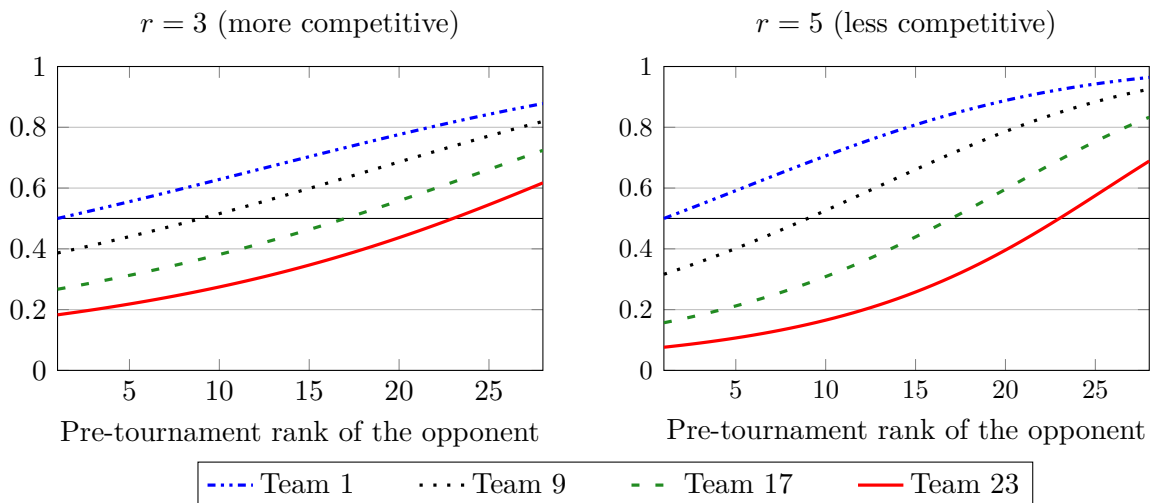


Figure 1 depicts the probability of winning for certain clubs based on formula (1). It can be seen that adjacent teams are closely matched, team $k - 1$ defeats team k with a

probability of no more than 55% even if $r = 5$. On the other hand, there is a substantial difference between a top club and an underdog: the strongest team has more than 70% chance to win against the bottom half of the teams even if $r = 3$.

3.2 Draws in the group and knock-out stages

Seeding may play a substantial role in knock-out tournaments (Hwang, 1982; Schwenk, 2000; Marchand, 2002; Groh et al., 2012; Karpov, 2016; Dagaev and Suzdaltsev, 2018; Karpov, 2018). Although the knock-out stage of competition formats $D(8+6)$ and $D(4 \times 7)$ is predetermined by the previous group stage (see Figures A.1–A.2) – with the exception of the Final Four when there is only one pot –, the clubs should be drawn into groups before the start of the tournament, which may affect its outcome (Guyon, 2015; Boczoń and Wilson, 2018; Guyon, 2018; Dagaev and Rudyak, 2019; Laliena and López, 2019; Cea et al., 2020). The knock-out brackets of format $D(4 \times 6)$ are also random.

Compared to the more popular UEFA Champions League, the composition of the pots in the EHF Champions League seems to be less regulated, it depends heavily on the decisions of the EHF Executive Committee (EHF, 2019a,b). Nevertheless, the administrators obviously intend to place the strongest clubs into groups A and B, including the titleholder, the champions, and the runners-up of the strongest associations, while the champions of low-ranked associations and the runners-up of middle-ranked associations go to groups C and D. Furthermore, groups A and B (C and D) are drawn from eight (six) pots such that the best teams are coming from the first pots.

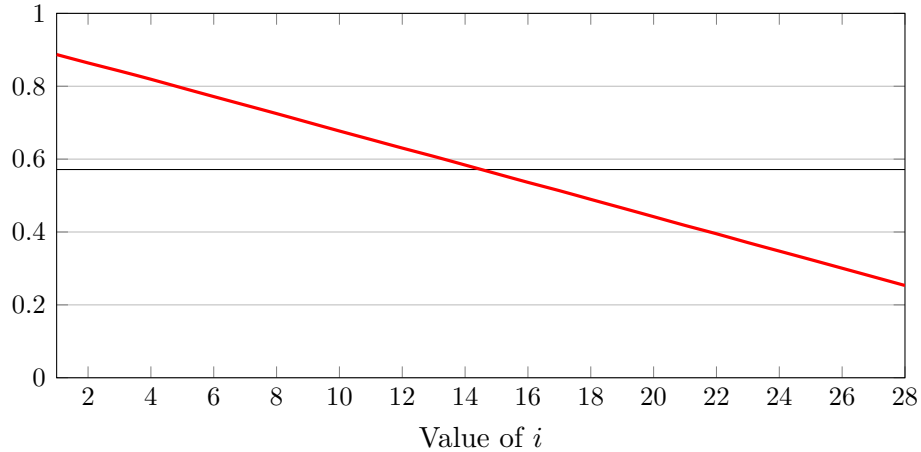
Therefore, two variants of each tournament design, called *seeded* and *random*, will be examined. They can be considered as two extreme cases: in the real-world, it is impossible to perfectly identify the strength of a team, however, the organiser has some information on it because past performance, the composition of the squad, etc. are known before the draw. The actual scenario lies somewhere between these limits, and if a tournament design has more favourable metrics with both assumptions, then it would probably be better in practice, too.

In the seeded version of $D(8+6)$, the two teams having the highest pre-tournament rank are placed in Pot 1, the next two teams are placed in Pot 2, and so on. Groups A and B get a club from each of Pots 1–8 randomly, while groups C and D get a club from each of Pots 9–14 randomly. Analogously, in the seeded versions of $D(4 \times 7)$ and $D(4 \times 6)$, the four teams with the highest pre-tournament rank are placed in Pot 1, the next four teams are placed in Pot 2, and so on. The four groups get a club from each pot randomly. In the following, these formats are denoted by $D(8+6)/S$, $D(4 \times 7)/S$, and $D(4 \times 6)/S$, respectively.

The random variants work along similar lines but assume some uncertainty in the identification of the teams, while their strength remains unchanged. In particular, the teams are reranked on the basis of the stochastic values given by the formula $44 \times Rnd + (28 - i)$, where i is the teams' pre-tournament rank and Rnd is a random number drawn uniformly from the semi-open range $[0, 1)$. The seeding into pots is based on these estimated coefficients. The appropriate versions are denoted by $D(8+6)/R$, $D(4 \times 7)/R$, and $D(4 \times 6)/R$, respectively.

Figure 2 shows the probability that a club goes to groups A and B in the design $D(8+6)/R$, in other words, it is classified among the strongest 16. The best team has more than 88% chance to achieve this (compared to the 57.14% of full randomness) and the lowest-ranked team still has around 25% chance to be drawn into the two top groups.

Figure 2: The probability that team i is allocated to the top groups
Competition design $D(8 + 6)/R$



Finally, there is some randomness in the knock-out stage of the previous EHF Champions League design $D(4 \times 6)$. In its Round of 16, the group winners play against the fourth-placed teams such that teams from the same pot cannot be drawn against each other (EHF, 2015a). Contrary to the UEFA Champions League, there is no protection for the teams from the same country. The same rules apply for the runners-up and third-placed teams, which are drawn against each other. In the quarter-finals, the first pot contains the winners of group winners versus fourth-placed teams, and the second pot contains the winners of runners-up versus third-placed teams (EHF, 2015b). There are no restrictions, even the winner and the runner-up of the same group can meet as happened in the 2012/13 season of the EHF Champions League for Groups A, B, and C.

Table 1: Feasible matchings in the Round of 16 draw of design $D(4 \times 6)$

Group of the first team		Group of the second team							
A	B	B	B	C	C	C	D	D	D
B	A	C	D	A	D	D	A	C	C
C	D	D	A	D	A	B	B	A	B
D	C	A	C	B	B	A	C	B	A

In the Round of 16 draw, there are $4! = 24$ possible scenarios to pair the groups. Among them, 15 violates the restriction, hence we chose from the remaining 9 matchings with a uniform probability of $1/9$ for the four clashes between group winners and fourth-placed teams, as well as for the four matches between runners-up and third-placed teams. They are listed in Table 1. In the absence of any restrictions, the draw of the quarter-finals is straightforward after the pots are determined.

3.3 Tournament metrics and details of the simulation procedure

Choosing a particular design and a prediction model for match outcomes, the competition can be simulated repeatedly in order to obtain any metrics of interest.

We will analyse the following success measures, which are widely used in evaluating tournament designs via simulations (Scarf et al., 2009; Dagaev and Rudyak, 2019):

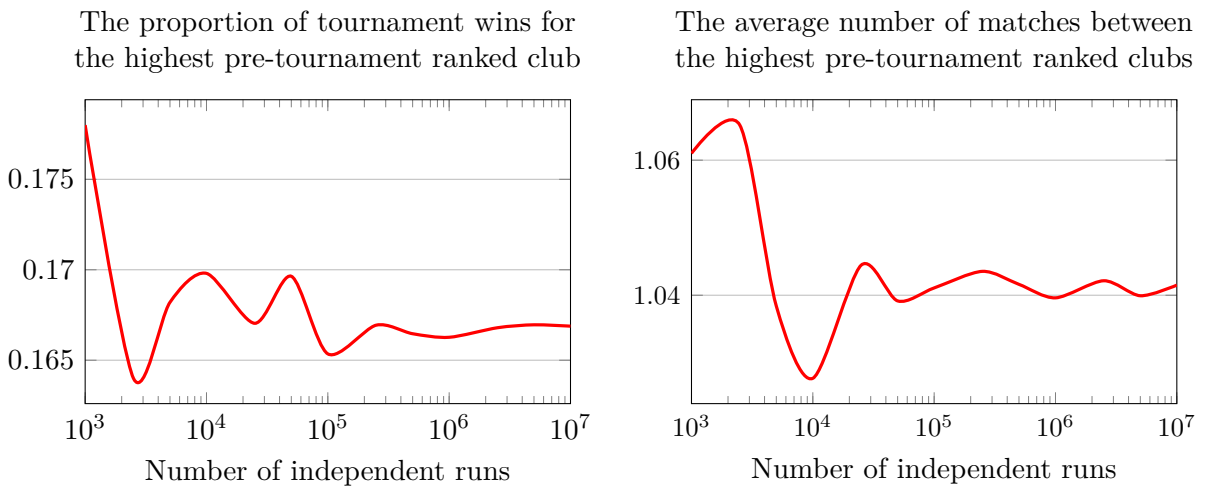
- the average pre-tournament ranks of the clubs in the Final Four, that is, the winner, the second-, third-, and fourth-placed teams;
- the expected quality of all matches, measured by the sum of the playing teams’ pre-tournament ranks;
- the expected competitive balance of all matches, measured by the difference between the playing teams’ pre-tournament ranks.

In the case of the first metric, our focus is on the first four places because the Final Four of the Champions League is promoted by the EHF separately. For example, this event has its own website (<http://www.ehffinal4.com/>), and tickets can be bought for the whole weekend, which offers an “*indisputable highlight of the European club handball season*”, instead of the individual matches (EHF, 2019c).

Hence, the first four clubs, as well as the number of matches played by any two clubs, and the winning percentage of each club have been recorded. Note also that a lower value is preferred for all tournament metrics.

According to Section 3.1, draws between the teams are not allowed in any match. This is not to be confused with ties in the ranking of round-robin groups, resolved in our simulations with an equal-odds coin toss. Furthermore, both formats contain a knock-out stage with home-away matches before the Final Four. If one team wins the first and the other wins the second match, then the qualifying team is chosen randomly with the probability given by formula (1), which is equivalent to the assumption that the clash is decided by three matches. It is used because resolution by a pure coin toss would be an inappropriate solution in the case of a fixed winning probability p : the chance to qualify with two matches is $p^2 + 2 \times p(1 - p) \times 0.5 = p$ as the probability of a tied contest is $2 \times p(1 - p)$, therefore, a knock-out played over two legs would be the same as a knock-out played over one leg.

Figure 3: Dependence of some tournament metrics on the number of iterations
Competition design $D(8 + 6)/R$; $\alpha = 4$



To get a reasonable estimate of all tournament metrics despite the stochastic nature of the simulations, we have determined the required number of independent runs on the basis of the random variant of design $D(8 + 6)$. Figure 3 shows that both success measures analysed for this purpose, the proportion of tournament wins for the strongest club, and

the average number of matches between the two strongest clubs in one iteration, remain practically unchanged after one million (10^6) iterations. Since, in the view of model limitations, there is no need to further reduce the statistical error, it has been decided that all subsequent simulations will be implemented with one million runs ($N = 1,000,000$).

The validity of the simulation procedure has been investigated in several ways. First, the assumption of equally strong teams ($p_{ij} = 0.5$ for all combinations of i and j) has led to, as expected, an outcome where all teams are placed first to fourth equally often in format $D(4 \times 7)$, but not in $D(8 + 6)$, see Section 2. However, the chances of the first 16 and the last 12 clubs are the same according to design $D(8 + 6)/S$. Second, we have analysed a fully deterministic matrix ($p_{ij} = 1$ if $i < j$), which implies that the top team wins, the fourth team is the fourth, while the second strongest occupies the second position with a probability of $2/3$ (when it does not play against the best team in the semi-final) and the third position with a probability of $1/3$ (when it plays against the best team in the semi-final) in the seeded variant of all tournament formats. Finally, changing certain values of the fully deterministic matrix has been checked to modify the outcome concerning the first four places accordingly.

3.4 The main limitations of our approach

No theoretical model can be a perfect substitute for real-world experiments. Any interpretation of our results should be careful due to the following reasons:

- The strength of all teams is exogenously given and fixed during the whole tournament.
- Draws are not allowed, which is not in line with the rules of handball.
- Home advantage is disregarded both in the group and elimination stages. This can be a problem if playing the second leg at home means an advantage. In addition, German teams might benefit from organising the Final Four in Cologne.
- It is assumed that teams exert full effort in all games. This is not necessarily true, especially if a team has already qualified from its group. In particular, according to the theoretical result of [Vong \(2017\)](#), if more than one team qualify from a group, a team may potentially benefit from exerting zero effort in some matches even when this is costless.
- Goal difference is taken into account neither in the group stage nor in the knock-out phase. However, the choice of the winner when both teams have won one match each in the elimination stage favours the better team, which somewhat mitigates this problem.

To conclude, the use of a particular probabilistic model always implies certain limitations. Nonetheless, some efforts will be made to minimise this weakness by carrying out robustness checks with respect to competitive balance and seeding, and a wide range of models may be suitable for comparative purposes ([Appleton, 1995](#)).

4 Results

In the following, a detailed analysis of the three competition formats is provided.

Table 2: The number of matches in a season

Tournament format	$D(8 + 6)$	$D(4 \times 7)$	$D(4 \times 6)$
Groups A and B	112	84	60
Groups C and D	60	84	60
Play-off	4	—	—
First knock-out phase	12	16	—
Round of 16	—	16	16
Quarter-finals	8	8	8
Final Four	4	4	4
Total	200	212	148

The tournament designs differ in the number of matches played according to Table 2. The innovative format $D(8 + 6)$ requires fewer matches than its traditional pair $D(4 \times 7)$. While $D(4 \times 6)$ contains only 24 teams, the reduction in the number of matches is more substantial.

Figure 4: The distribution of matches played in a season

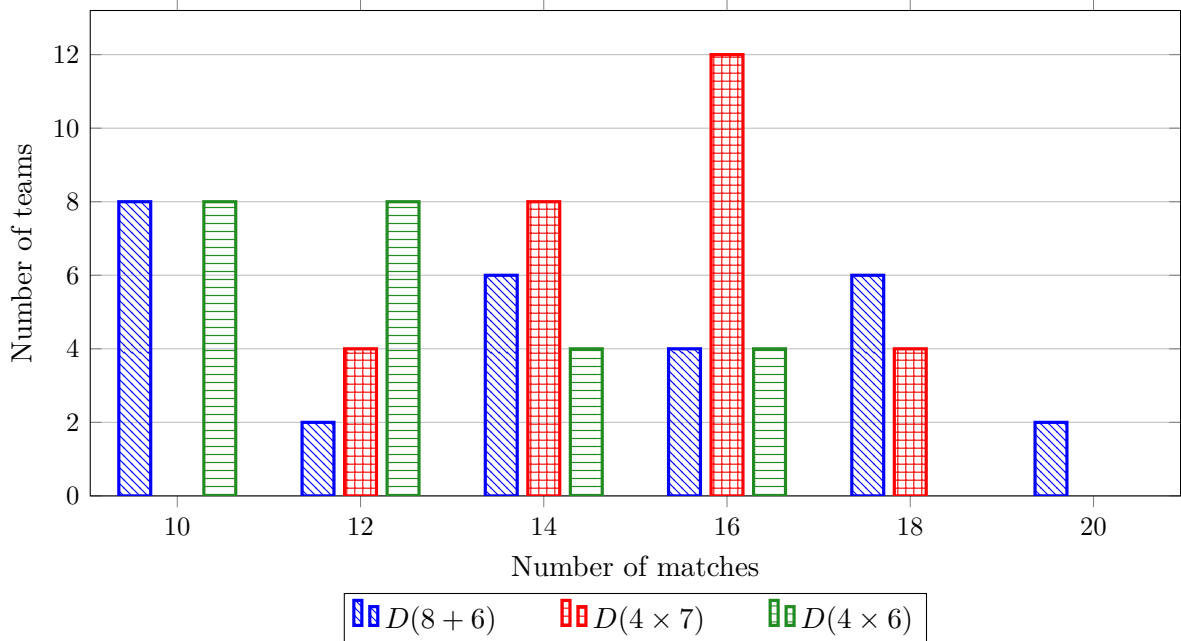
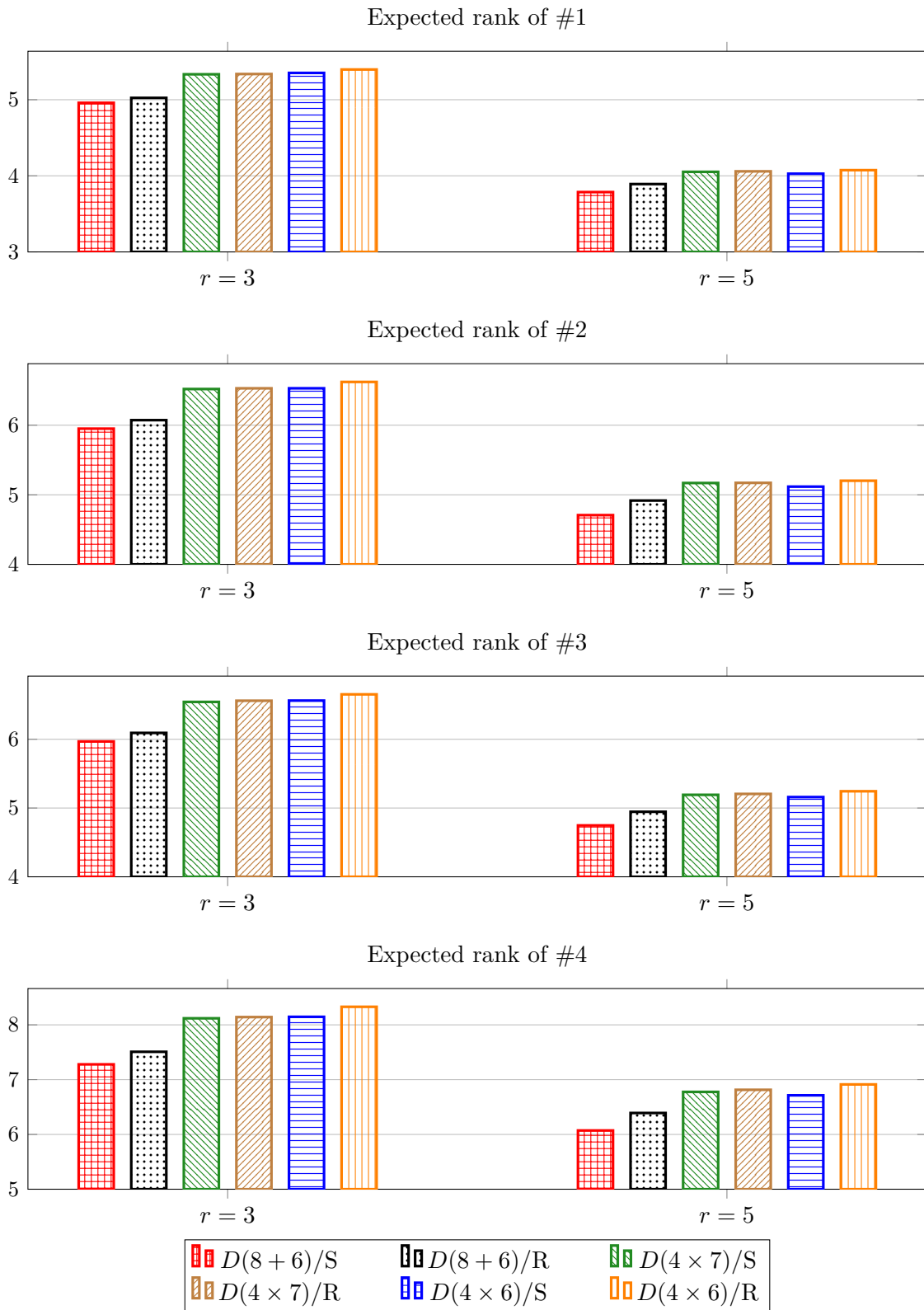


Figure 4 presents the distribution of the matches, the number of teams playing a given number of matches supposed that a stronger team always beats a weaker one (otherwise, the distribution may change). For example, in the design $D(4 \times 7)$, the four teams eliminated after the group stage play 12 matches each. The variance of the number of matches is the greatest in format $D(8 + 6)$.

Our first contest metric is the average pre-tournament ranks of the participants in the Final Four. It is revealed by Figure 5 that the seeded design $D(8 + 6)/S$ is superior in the ability to select the strongest teams, followed by its random variant $D(8 + 6)/R$, while the two versions of the competition formats $D(4 \times 7)$ and $D(4 \times 6)$ are close to each other, the seeded being somewhat more efficacious. This finding is robust across the four places and does not depend on the variation in the strengths of the clubs as the diagrams are

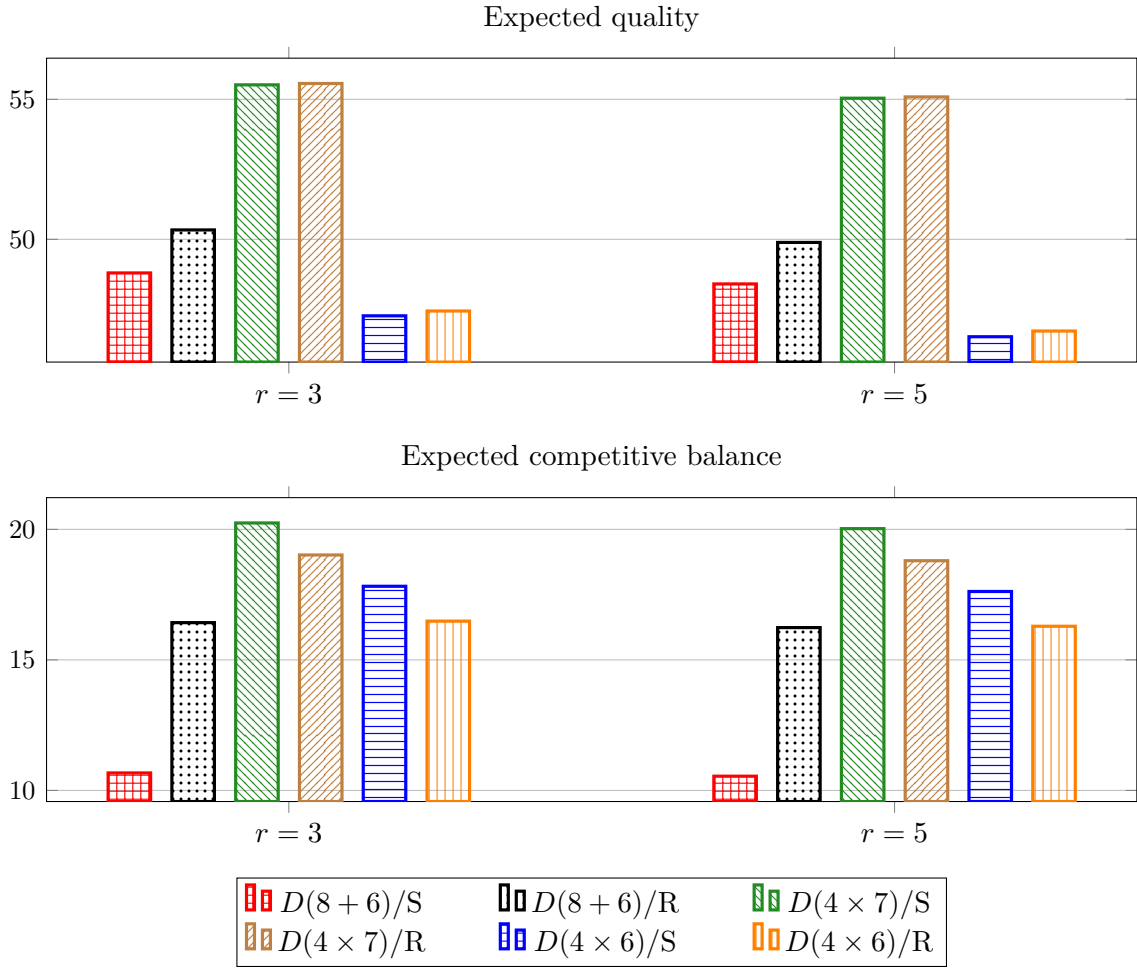
Figure 5: The average pre-tournament ranks of the teams in the Final Four



similar for all values of the competitiveness parameter r .

The picture is somewhat different for the other two tournament success measures, the expected quality and competitive balance of the matches (Figure 6). Remember that a lower value of quality means that the matches are played by stronger teams on average.

Figure 6: Characteristics of all matches played



Therefore, since the format $D(4 \times 6)$ does not contain the four weakest teams, it proves to be the best for this metric. On the other hand, design $D(8+6)$ outperforms $D(4 \times 7)$, even if the teams are not allocated perfectly into the top and bottom groups.

Analogously, a smaller competitive balance corresponds to more uncertainty in the outcome of a match. Now the random variant of formats $D(4 \times 7)$ and $D(4 \times 6)$ has a more favourable value compared to the seeded versions because the optimal seeding guarantees that the strong teams play against underdogs in the groups. While competitive balance substantially deteriorates for design $D(8+6)$ when the clubs cannot be allocated perfectly into the two types of groups, $D(8+6)/R$ is still consistently better than the alternative design $D(4 \times 7)$, and not worse than $D(4 \times 6)$, which is organised with four teams less.

It is worth recalling here that unbalanced groups are recommended to raise the expected quality and competitive balance of all matches played, thus the more visible and robust advantage of the format $D(8+6)$ over $D(4 \times 7)$ in Figure 6 compared to Figure 5 reinforces our main message. In fact, a substantial increase in match quality and competitive balance can be probably tolerated even if the selection ability of the championship, that is, the average pre-tournament ranks of the contestants in the Final Four would somewhat worsen.

The structure of format $D(8+6)$ opens the possibility that a team might benefit from being in the bottom groups C or D instead of the top groups A and B. This is not only an academic problem, for instance, the [2017/18 EHF Champions League](#) was won by [Montpellier Handball](#), a club which started from group C after being the third in the

previous season of the LNH (*Ligue Nationale de Handball*) Division 1, the French premier handball league.

If a strong team gains from masking itself as a weaker one, this can be called unfair because it gives incentives to tank. For example, the qualification for the UEFA European Championship 2020 is demonstrated to exhibit such a shortcoming (Csató, 2020a).

Hence an alternative scenario called *erroneous team identification* is considered when:

- the k th pre-tournament ranked team is correctly identified as the k th strongest one by the seeding procedure for all $1 \leq k \leq 8$ and $18 \leq k \leq 28$;
- the 9th pre-tournament ranked team is identified as the 17th strongest one by the seeding procedure;
- the ℓ th pre-tournament ranked team is identified as the $(\ell - 1)$ th strongest one by the seeding procedure for all $10 \leq \ell \leq 17$.

To be short, contrary to the original case of *correct team identification*, the 9th best club now seems to be only the 17th before the tournament, therefore it might obtain a less difficult path into the Final Four. Note that, while the current quality of the teams is obviously unknown in practice, the random variants of the tournament formats also address this kind of uncertainty.

Figure 7 shows the average number of matches played and the average winning performance for each team under both scenarios. There is a break in both measures between the 16th and 17th strongest teams if the seeded $D(8 + 6)/S$ design is used, but it is smoothed out by the random variant $D(8 + 6)/R$. Format $D(8 + 6)/R$ involves more matches between the leading teams and reduces the variance in winning probabilities compared to the traditional design of $D(4 \times 7)$. Differences in the winning percentage are greater for the random versions of designs $D(4 \times 7)$ and $D(4 \times 6)$ than for the seeded versions because weak teams are forced to play against strong teams by the latter regime.

According to Figure 7.b, the 9th team, which is erroneously identified as an underdog, plays fewer matches with a higher winning percentage as it becomes the best team of the bottom groups in the competition design $D(8 + 6)/S$. However, the random variant $D(8 + 6)/R$ substantially weakens the effect of this error.

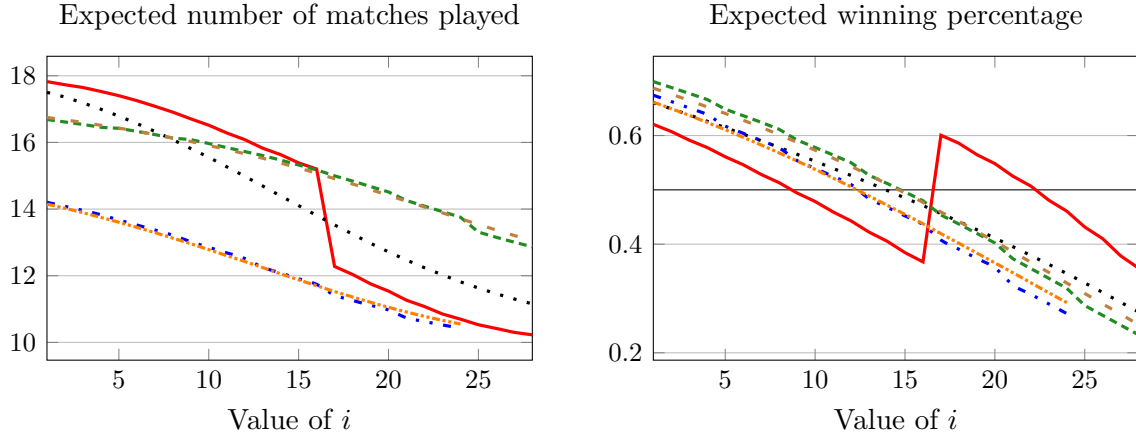
Fairness is highlighted in Figure 8 by plotting the ratio of expected prizes between a club and the best club among the clubs that are weaker than it. The expected prize is defined here by giving five points to the tournament winner, three points to the second-, two points to the third-, and one point to the fourth-placed team. The ratio is not calculated for the last four teams because the expected prize of low-quality clubs becomes volatile.

In the case of correct team identification, these ratios are consistently over one for low values of the competitiveness parameter r , that is, in a more competitive tournament (Figure 8.a). On the other hand, the 17th club has a higher chance to achieve a good position than the 16th club in the format $D(8 + 6)/S$ if the teams' a priori strength differ significantly ($r = 5$). However, this seems to be only a marginal violation of fairness as the expected prize of the 17th team is still lower than the expected prize of the 15th, and introducing randomness into the identification of the teams immediately solves the problem.

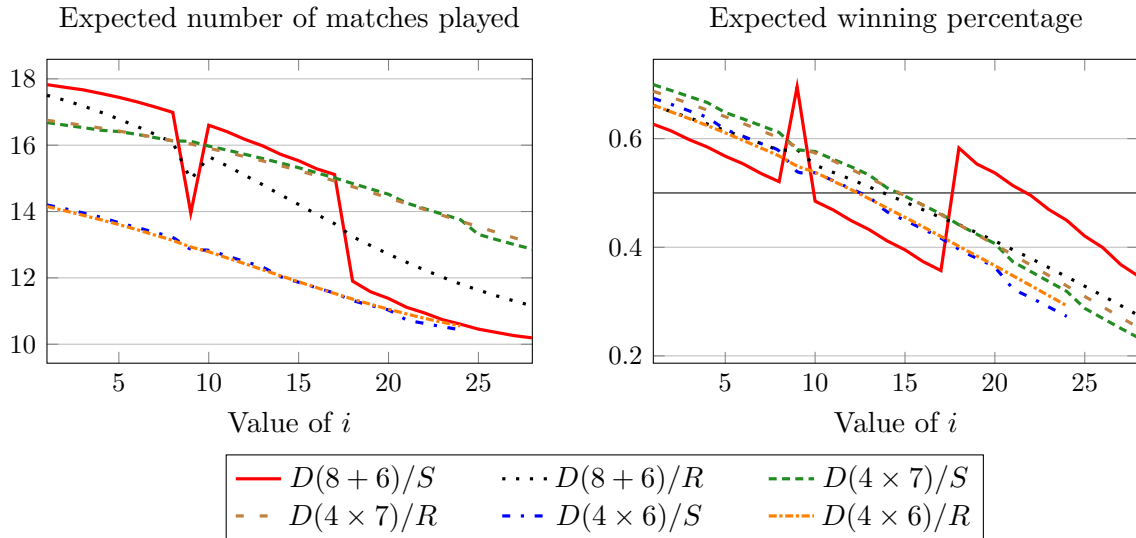
The potential unfairness caused by erroneous team identification is even more mitigated (Figure 8.b). In particular, the 9th club loses by being listed among the underdogs as it obtains a lower expected prize than the 10th club in the seeded design $D(8 + 6)/S$ if $r = 3$.

Figure 7: Individual team statistics ($r = 3$)

(a) Correct team identification



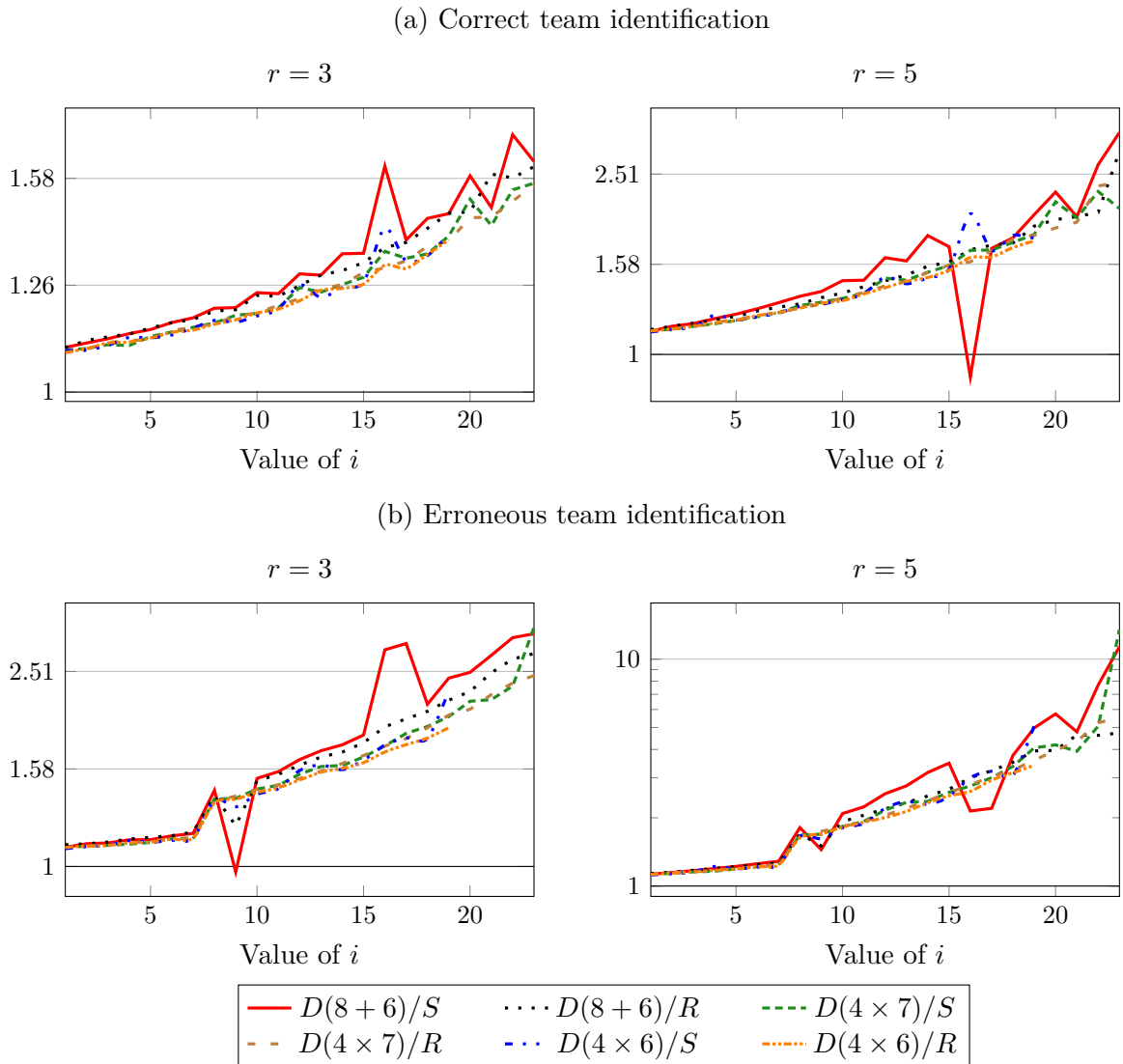
(b) Erroneous team identification



Furthermore, random seeding ($D(8+6)/R$) or more diverse pre-tournament strength of the teams ($r = 5$) eliminates the possibility that a weaker team goes into the Final Four with a higher probability than a stronger one. The existence of top and bottom groups in design $D(8+6)$ did not contribute to the unexpected victory of Montpellier.

According to Table A.1 in the Appendix, erroneous team identification does not influence our tournaments metrics, the efficacy of the average ranks of the first four-placed teams, the expected quality of all matches, and outcome uncertainty, in a meaningful way. It also reinforces that the choice of the tournament format is more important than the correct identification of the teams before the seeding. Furthermore, format $D(8+6)$ remains undoubtedly superior to $D(4 \times 7)$ despite that fewer matches are played. On the other hand, this implication would not be true concerning $D(4 \times 6)$ as selecting the best teams for the Final Four is simpler on the basis of more matches.

Figure 8: The ratio between the expected prizes of team i and team $i + 1$



5 Conclusions

The design of a hybrid tournament consisting of round-robin groups followed by a knock-out phase raises several interesting theoretical questions on various fields such as fairness (Guyon, 2018; Cea et al., 2020), or strategy-proofness (Pauly, 2014; Vong, 2017; Dagaev and Sonin, 2018; Csató, 2019b, 2020b). The current paper has attempted to explore the potential effects of creating groups with different quality through the example of the EHF Champions League, the most prestigious men’s handball club competition in Europe. Its non-traditional design for 28 teams, applied since the 2015/16 season, has been compared to the classical format of four balanced, equally strong groups with seven teams each, and to the previous design of the tournament.

Compared to the formats containing balanced groups, the innovative competition design is able to considerably increase the proportion of high quality and even matches, with positive effects on demand. In addition, it turns out to be more *efficacious* as the average pre-tournament ranks of the teams finishing at any positions in the Final Four are smaller. Our numerical results have revealed the new competition design is in line with the intentions of the EHF (EHF, 2014), too. Nonetheless, both the men’s and the

women’s EHF Champions League competitions will be played with two balanced groups of eight teams from the 2020/21 season (EHF, 2018).

It should be recognised that there are always some unavoidable trade-offs between the goals of a tournament designer, and they cannot be achieved simultaneously as recent theoretical results show (Krumer et al., 2017b). For example, while format $D(8 + 6)$ is better at selecting stronger teams to be the winner, decreasing the chances that a weaker team plays against a stronger one can be detrimental for gradually improving the squad of the former. Furthermore, bottom groups without leading clubs can generate limited media attention.

The idea of unbalanced groups may be used in several other settings. UEFA seems to follow this principle by the structure of the new biennial international football competition [UEFA Nations League](#), which started in 2018: the 55 UEFA national teams are divided into four divisions called leagues such that in each of them, four groups are formed with teams of similar quality, achieved by promotion and relegation between the leagues over time. An alternative design can also be considered for the UEFA Champions League with the first three teams from four top groups and the group winners from four bottom groups qualifying for the Round of 16. This solution could create more matchups between the richest clubs without imposing severe barriers to entry for teams outside the current elite.

To summarise, we have successfully challenged the traditional seeding system used in hybrid tournaments. It is worth considering an alternative format composed of unbalanced groups with more even matchings within groups, which are treated unequally to guarantee an easier path for teams from the top groups to qualify for the next stage.

Acknowledgements

My father (also called *László Csató*) has made a substantial contribution to the paper by helping to code the simulations in Python.

We are grateful to *Tamás Halm* for reading the manuscript.

Five anonymous reviewers provided valuable comments and suggestions on earlier drafts. We are indebted to the [Wikipedia community](#) for collecting and structuring valuable information used in our research.

The research was supported by the MTA Premium Postdoctoral Research Program grant PPD2019-9/2019.

References

- Appleton, D. R. (1995). May the best man win? *Journal of the Royal Statistical Society: Series D (The Statistician)*, 44(4):529–538.
- Boczoń, M. and Wilson, A. J. (2018). Goals, constraints, and public assignment: A field study of the UEFA Champions League. Technical Report 18/016, University of Pittsburgh, Kenneth P. Dietrich School of Arts and Sciences, Department of Economics. https://www.econ.pitt.edu/sites/default/files/working_papers/Working%20Paper.18.16.pdf.
- Borland, J. and MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy*, 19(4):478–502.

- Brams, S. J. and Ismail, M. S. (2018). Making the rules of sports fairer. *SIAM Review*, 60(1):181–202.
- Cea, S., Durán, G., Guajardo, M., Sauré, D., Siebert, J., and Zamorano, G. (2020). An analytics approach to the FIFA ranking procedure and the World Cup final draw. *Annals of Operations Research*, 286(1-2):119–146.
- Corona, F., Forrest, D., Tena, J. D., and Wiper, M. (2019). Bayesian forecasting of UEFA Champions League under alternative seeding regimes. *International Journal of Forecasting*, 35(2):722–732.
- Csató, L. (2019a). A simulation comparison of tournament designs for the World Men’s Handball Championships. *International Transactions in Operational Research*, in press. DOI: [10.1111/itor.12691](https://doi.org/10.1111/itor.12691).
- Csató, L. (2019b). UEFA Champions League entry has not satisfied strategyproofness in three seasons. *Journal of Sports Economics*, 20(7):975–981.
- Csató, L. (2020a). Fair tournament design: A flaw of the UEFA Euro 2020 qualification. Manuscript. arXiv: [1905.03325](https://arxiv.org/abs/1905.03325).
- Csató, L. (2020b). The UEFA Champions League seeding is not strategy-proof since the 2015/16 season. *Annals of Operations Research*, in press. DOI: [10.1007/s10479-020-03637-1](https://doi.org/10.1007/s10479-020-03637-1).
- Dagaev, D. and Rudyak, V. (2019). Seeding the UEFA Champions League participants: Evaluation of the reform. *Journal of Quantitative Analysis in Sports*, 15(2):129–140.
- Dagaev, D. and Sonin, K. (2018). Winning by losing: Incentive incompatibility in multiple qualifiers. *Journal of Sports Economics*, 19(8):1122–1146.
- Dagaev, D. and Suzdaltsev, A. (2018). Competitive intensity and quality maximizing seedings in knock-out tournaments. *Journal of Combinatorial Optimization*, 35(1):170–188.
- Dixon, M. J. and Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280.
- Dumangane, M., Rosati, N., and Volossovitch, A. (2009). Departure from independence and stationarity in a handball match. *Journal of Applied Statistics*, 36(7):723–741.
- EHF (2014). The 119th meeting of the EHF Executive Committee. 21 March. <http://www.eurohandball.com/article/018817/EHF+Executive+meets+in+Vienna>.
- EHF (2015a). Defending champions lurk in the fourth pot before the last 16 draw. 23 February. <https://web.archive.org/web/20160304044027/http://www.ehfcl.com/men/2014-15/article/21689/Defending+champions+lurk+in+the+fourth+pot+before+the+Last+16+draw>.
- EHF (2015b). Last eighth await Tuesday’s quarter-final fate. 23 March. <https://web.archive.org/web/20170705093614/http://www.ehfcl.com/men/2014-15/article/21877/Last+eight+await+Tuesday%27s+quarter%2Dfinal+fate>.

- EHF (2018). All set for 2020 as club competition revamp agreed. 15 December. <http://www.eurohandball.com/article/031868/All+set+for+2020+as+club+competition+revamp+agreed>.
- EHF (2019a). Seeding for the group phase draw released. 24 June. <http://www.eurohandball.com/article/032951/Seeding+for+the+group+phase+draw+released>.
- EHF (2019b). Starting grid for the 2019/20 season confirmed. 21 June. <http://www.eurohandball.com/article/032935/Starting+grid+for+the+2019%2F20+season+confirmed>.
- EHF (2019c). Tickets for VELUX EHF FINAL4 go on sale. 30 May. <http://www.eurohandball.com/article/032759/Tickets+for+VELUX+EHF+FINAL4+2020+go+on+sale>.
- Forrest, D. and Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2):229–241.
- Goossens, D. R., Beliën, J., and Spieksma, F. C. R. (2012). Comparing league formats with respect to match importance in Belgian football. *Annals of Operations Research*, 194(1):223–240.
- Groh, C., Moldovanu, B., Sela, A., and Sunde, U. (2012). Optimal seedings in elimination tournaments. *Economic Theory*, 49(1):59–80.
- Groll, A., Heiner, J., Schauburger, G., and Uhrmeister, J. (2020). Prediction of the 2019 IHF World Men’s Handball Championship – An underdispersed sparse count data regression model. *Journal of Sports Analytics*, in press. DOI: [10.3233/JSA-200384](https://doi.org/10.3233/JSA-200384).
- Guyon, J. (2015). Rethinking the FIFA World Cup™ final draw. *Journal of Quantitative Analysis in Sports*, 11(3):169–182.
- Guyon, J. (2018). What a fairer 24 team UEFA Euro could look like. *Journal of Sports Analytics*, 4(4):297–317.
- Hwang, F. K. (1982). New concepts in seeding knockout tournaments. *The American Mathematical Monthly*, 89(4):235–239.
- Karpov, A. (2016). A new knockout tournament seeding method and its axiomatic justification. *Operations Research Letters*, 44(6):706–711.
- Karpov, A. (2018). Generalized knockout tournament seedings. *International Journal of Computer Science in Sport*, 17(2):113–127.
- Koning, R. H., Koolhaas, M., Renes, G., and Ridder, G. (2003). A simulation model for football championships. *European Journal of Operational Research*, 148(2):268–276.
- Krumer, A. (2020). Testing the effect of kick-off time in the UEFA Europa League. *European Sport Management Quarterly*, 20(2):225–238.
- Krumer, A. and Lechner, M. (2017). First in first win: Evidence on schedule effects in round-robin tournaments in mega-events. *European Economic Review*, 100:412–427.

- Krumer, A., Megidish, R., and Sela, A. (2017a). First-mover advantage in round-robin tournaments. *Social Choice and Welfare*, 48(3):633–658.
- Krumer, A., Megidish, R., and Sela, A. (2017b). Round-robin tournaments with a dominant player. *The Scandinavian Journal of Economics*, 119(4):1167–1200.
- Krumer, A., Megidish, R., and Sela, A. (2020). The optimal design of round-robin tournaments with three players. *Journal of Scheduling*, 23(3):379–396.
- Lago-Peñas, C., Gómez, A. M., Viaño, J., González-García, I., and Fernández-Villarino, M. de Los á. (2013). Home advantage in elite handball: the impact of the quality of opposition on team performance. *International Journal of Performance Analysis in Sport*, 13(3):724–733.
- Laliena, P. and López, F. J. (2019). Fair draws for group rounds in sport tournaments. *International Transactions in Operational Research*, 26(2):439–457.
- Lasek, J. and Gagolewski, M. (2018). The efficacy of league formats in ranking teams. *Statistical Modelling*, 18(5-6):411–435.
- Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, 36(3):109–118.
- Marchand, É. (2002). On the comparison between standard and random knockout tournaments. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 51(2):169–178.
- McGarry, T. and Schutz, R. W. (1997). Efficacy of traditional sport tournament structures. *Journal of the Operational Research Society*, 48(1):65–74.
- Meletakos, P. and Bayios, I. (2010). General trends in European men’s handball: a longitudinal study. *International Journal of Performance Analysis in Sport*, 10(3):221–228.
- Mendonça, D. and Raghavachari, M. (2000). Comparing the efficacy of ranking methods for multiple round-robin tournaments. *European Journal of Operational Research*, 123(3):593–605.
- Palacios-Huerta, I. (2012). Tournaments, fairness and the Prouhet-Thue-Morse sequence. *Economic Inquiry*, 50(3):848–849.
- Pauly, M. (2014). Can strategizing in round-robin subtournaments be avoided? *Social Choice and Welfare*, 43(1):29–46.
- Ryvkin, D. (2010). The selection efficiency of tournaments. *European Journal of Operational Research*, 206(3):667–675.
- Sahm, M. (2019). Are sequential round-robin tournaments discriminatory? *Journal of Public Economic Theory*, 21(1):44–61.
- Scarf, P., Yusof, M. M., and Bilbao, M. (2009). A numerical study of designs for sporting contests. *European Journal of Operational Research*, 198(1):190–198.

- Scarf, P. A. and 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.
- Schwenk, A. J. (2000). What is the correct way to seed a knockout tournament? *The American Mathematical Monthly*, 107(2):140–150.
- Szymanski, S. (2003). The economic design of sporting contests. *Journal of Economic Literature*, 41(4):1137–1187.
- Tutz, G. and Schauburger, G. (2015). Extended ordered paired comparison models with application to football data from German Bundesliga. *AStA Advances in Statistical Analysis*, 99(2):209–227.
- Vong, A. I. K. (2017). Strategic manipulation in tournament games. *Games and Economic Behavior*, 102:562–567.

Appendix

Table A.1: Estimates of tournament metrics for all designs
D86, *D77*, and *D66* stand for competition designs $D(8+6)$, $D(4 \times 7)$, and $D(4 \times 6)$, respectively

(a) $r = 3$ (more competitive), correct team identification

	D86/S	D86/R	D77/S	D77/R	D66/S	D66/R
Average rank of #1	4.959	5.025	5.334	5.338	5.352	5.397
Average rank of #2	5.951	6.073	6.520	6.530	6.531	6.622
Average rank of #3	5.968	6.092	6.543	6.561	6.563	6.653
Average rank of #4	7.279	7.508	8.119	8.142	8.147	8.329
Expected quality of all matches	48.80	50.34	55.52	55.57	47.27	47.44
Expected competitive balance of all matches	10.67	16.43	20.24	19.02	17.82	16.48

(b) $r = 3$ (more competitive), erroneous team identification

	D86/S	D86/R	D77/S	D77/R	D66/S	D66/R
Average rank of #1	4.951	5.021	5.338	5.340	5.349	5.394
Average rank of #2	5.946	6.088	6.528	6.531	6.540	6.631
Average rank of #3	5.963	6.106	6.551	6.562	6.577	6.666
Average rank of #4	7.277	7.536	8.136	8.140	8.184	8.328
Expected quality of all matches	48.75	50.40	55.57	55.61	47.37	47.53
Expected competitive balance of all matches	10.67	16.46	20.27	19.05	17.88	16.53

(c) $r = 5$ (less competitive), correct team identification

	D86/S	D86/R	D77/S	D77/R	D66/S	D66/R
Average rank of #1	3.786	3.891	4.052	4.059	4.028	4.075
Average rank of #2	4.707	4.917	5.169	5.172	5.117	5.202
Average rank of #3	4.747	4.946	5.192	5.205	5.159	5.244
Average rank of #4	6.072	6.391	6.776	6.815	6.714	6.913
Expected quality of all matches	48.41	49.89	55.04	55.09	46.52	46.72
Expected competitive balance of all matches	10.54	16.24	20.03	18.80	17.62	16.29

(d) $r = 5$ (less competitive), erroneous team identification

	D86/S	D86/R	D77/S	D77/R	D66/S	D66/R
Average rank of #1	3.801	3.885	4.055	4.052	4.018	4.072
Average rank of #2	4.736	4.912	5.164	5.165	5.108	5.201
Average rank of #3	4.765	4.942	5.203	5.210	5.152	5.253
Average rank of #4	6.098	6.410	6.781	6.815	6.727	6.910
Expected quality of all matches	48.39	49.97	55.10	55.14	46.67	46.84
Expected competitive balance of all matches	10.56	16.29	20.06	18.84	17.72	16.36

Figure A.1: Design $D(8 + 6)$, which has been used in the men's handball [EHF Champions League](#) since the [2015/16 season](#)

(a) Group stage

Group A	Group B	Group C	Group D
A1	B1	C1	D1
A2	B2	C2	D2
A3	B3	C3	D3
A4	B4	C4	D4
A5	B5	C5	D5
A6	B6	C6	D6
A7	B7		
A8	B8		

(b) Play-off

K1	C2	K2	C1
	D1		D2

(c) Knockout stage

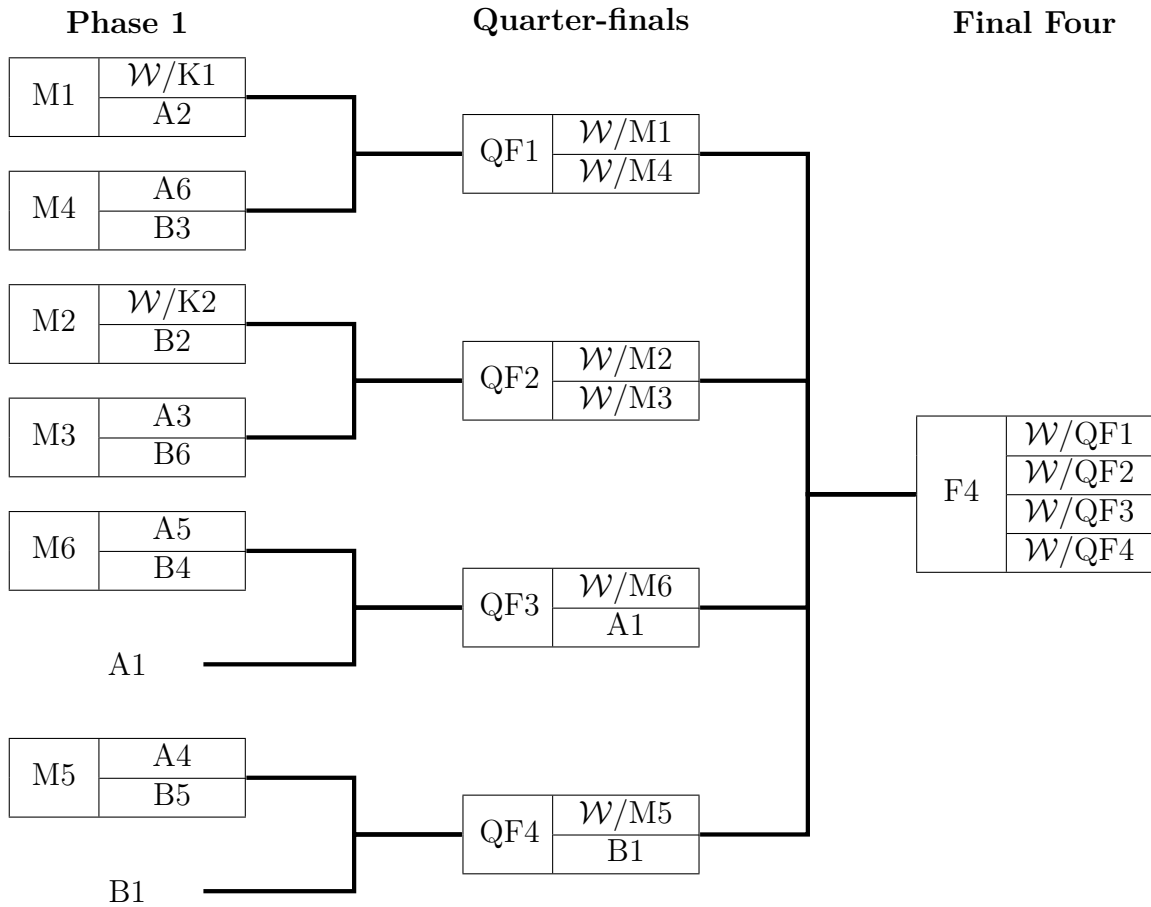


Figure A.2: Design $D(4 \times 7)$, a traditional alternative format with 28 teams

(a) Group stage

Group A	Group B	Group C	Group D
A1	B1	C1	D1
A2	B2	C2	D2
A3	B3	C3	D3
A4	B4	C4	D4
A5	B5	C5	D5
A6	B6	C6	D6
A7	B7	C7	D7

(b) Knockout stage

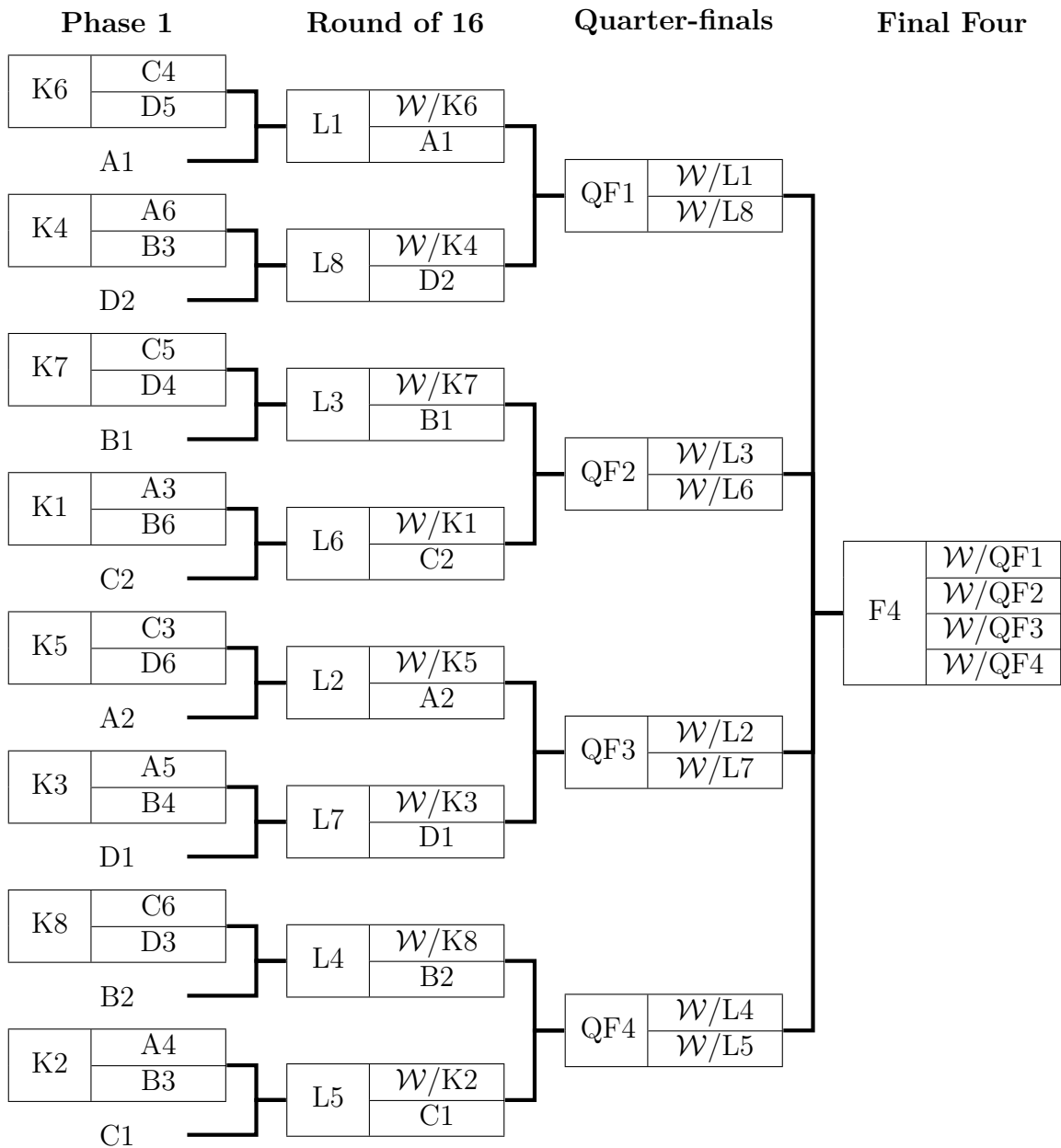


Figure A.3: Design $D(4 \times 6)$, the previous format of the men's handball [EHF Champions League](#)

(a) Group stage

Group A	Group B	Group C	Group D
A1	B1	C1	D1
A2	B2	C2	D2
A3	B3	C3	D3
A4	B4	C4	D4
A5	B5	C5	D5
A6	B6	C6	D6

(b) Knockout stage

