A numerical study of tournament structure and seeding policy for the soccer World Cup Finals

P. A. Scarf* and M. M. Yusof

Centre for Operations Management, Management Science and Statistics, Salford Business School, University of Salford, M5 4WT, United Kingdom

Tournament outcome uncertainty depends on: the design of the tournament; and the relative strengths of the competitors – the competitive balance. A tournament design comprises the arrangement of the individual matches, which we call the tournament structure, the seeding policy and the progression rules. In this paper, we investigate the effect of seeding policy for various tournament structures, while taking account of competitive balance. Our methodology uses tournament outcome uncertainty to consider the effect of seeding policy and other design changes. The tournament outcome uncertainty is measured using the tournament outcome characteristic which is the probability $P_{q,R}$ that a team in the top $100q$ pre-tournament rank percentile progresses forward from round R, for all q and R. We use Monte Carlo simulation to calculate the values of this metric. We find that, in general, seeding favours stronger competitors, but that the degree of favouritism varies with the type of seeding. Reseeding after each round favours the strong to the greatest extent. The ideas in the paper are illustrated using the soccer World Cup Finals tournament.

Keywords and Phrases: competitive balance, soccer, tournament design, Monte Carlo simulation.

1. Introduction

When designing a tournament, it is useful to understand the characteristics of differing tournament designs. The characteristics of a tournament design are many and include the number of teams or competitors, the progression rules, the schedule of matches, the seeding policy, and the competitive balance of the teams. The progression rules, the schedule of matches, and the seeding policy collectively we consider to be the tournament design. The progression rules and the schedule of matches, but without the allocation of particular competitors to the schedule, we call the tournament structure. In this paper, we will be concerned with how the tournament structure influences the outcome of the tournament, and in particular, taking a statistical approach, how the tournament structure influences the tournament

*P.A. Scarf@salford.ac.uk

outcome uncertainty. We measure the tournament outcome uncertainty using the tournament outcome characteristic which we define as the probability $P_{q,R}$ that the team in the top $100q$ pre-tournament rank percentile progresses forward from round R, for all q and R. This metric was proposed by SCARF, BILBAO and YUSOF (2009) for comparing tournament designs. The metric is calculated using simulation in a manner similar to KONING *et al.* (2003).

We distinguish tournament outcome uncertainty from the competitive balance of the tournament. We consider the competitive balance to be the relative strength of the competitors (KONING, 2000). We make this distinction for the following reason. When a tournament design is fixed, e.g. FA Premier League, competitive balance and tournament outcome uncertainty are directly related – the one implies the other. The terms are therefore frequently given the same interpretation. However, when the tournament design is not fixed, they are subtly different measures of a tournament. For example, if we take a set S of 2^n teams say with defined relative strengths, that is prob(i beats j) is known for all $i, j \in S$, then the probability that the best team wins, for example, will depend on not just these relative strengths but also on the design of the tournament. The probability that the best team wins will be smaller for a knockout (KO) or a single elimination tournament than for a round-robin (RR). Thus there is greater uncertainty of outcome in a KO tournament than in a RR even though the relative strengths of teams playing are the same. Furthermore, if one considers the hypothetical repeated RR tournament design in which every team plays every other k times, as $k \to \infty$ the tournament outcome will be certain unless the tournament is perfectly balanced (prob(i beats j) $= 1/2$ for all $i \neq j$). It is our idea that when comparing tournament designs, one should consider the tournament outcome uncertainty given the competitive balance of the teams.

MCGARRY and SCHUTZ (1997) stated that a tournament can be considered as 'fair' if it ranks competitors in accordance with their ability. Note, by the rank of competitors here, we mean their finishing position in the tournament. At the outcome of a tournament, a competitor of stronger (higher) ability should rank above a competitor of weaker (lower) ability. Of course, the outcome of a tournament is uncertain – this is, part of the attraction for viewers – and so one can only consider the tendency of a tournament to rank competitors according to ability. In this sense, given this uncertainty of outcome, a RR is 'fairer' than a KO tournament because the RR has a greater tendency to rank entrants on the basis of ability. That is, for a given competitive balance of the entrants, the RR exhibits less outcome uncertainty than a KO.

The philosophical issue of fairness in a tournament is interesting. Tournament uncertainty and fairness using McGarry's definition [that a tournament is fair if it ranks competitors on the basis of ability (MCGARRY and SCHUTZ, 1997)] are to an extent conflicting criteria. One cannot typically maximise both tournament outcome uncertainty and fairness. Fans of weak teams and neutral observers would be expected to favour a tournament in which uncertainty is high, while fans of strong teams will want to see systems in place that ensure their team progresses to the later stages of a tournament. Other stakeholders will want to ensure that a tournament

is an economic and political success, and the early exit of strong teams will undermine this. A partial solution is to match weaker competitors in an earlier qualifying tournament. The UEFA Champions League is a case in point. Alternatively, KO tournaments can be extended so that early rounds only feature weaker competitors. Junior tennis tournaments in the UK are organised along these lines. The FA Cup in England and Wales also operates such a system. Unbalanced designs such as the Swiss system used in chess and bridge might also be employed. Here successive rounds are played without elimination but with pairings for matches in later rounds determined by performance in earlier rounds. Final rankings are based on both the number of wins and the strengths of opponents played.

If the tournament structure is preset, one way to influence tournament outcome is through seeding (Schwenk, 2000). Seeding is the ordering of entrants prior to the tournament on the basis of playing history and/or the judgement of experts. The accuracy of the seeding and the effect of inaccurate seeding on the tournament outcome are then pertinent questions. McGarry (1998) highlights these points. However, in our paper, we are immediately concerned with comparing differing seeding policies. To do so, we consider the effect of seeding policy on tournament outcome uncertainty, while taking account of competitive balance. It extends a preliminary analysis carried out in Yusof and Scarf (2009); that paper only looked at standard seeding. Knowledge of the effect of seeding policy can be used to consider the implications of changing a tournament structure, and to assess the 'fairness' of a tournament. In section 3, we describe the seeding policies that we consider in detail, having first discussed the two principal factors that influence tournament outcome uncertainty: the design and the competitive balance. In section 4, we discuss in detail the tournament structures that we study, the model of competitive balance that we use, and the calculation of the tournament outcome characteristic. Section 5 presents the results of our analysis. We conclude the paper with a discussion and make recommendations for further work.

2. Tournament design and competitive balance

Many different tournament structures are used in practice. Two tournament structures may be regarded as fundamental, with all other designs considered as variations or hybrids. These are the RR and the KO.

In the RR, every competitor or team plays every other competitor an equal number of times. Typically, points are awarded for wins, and draws or ties (if applicable). The winner of the tournament (the 'champion') is the competitor with the largest number of points at the end of the tournament. If two or more teams finish with the same number of points, tiebreakers are used, e.g. goal difference in soccer. A RR tournament is often called a league. In a RR with n teams playing each other on l occasions, there are of $ln(n-1)/2$ matches with $l(n-1)$ matches for each team. Often $l = 2$ and teams play home and away matches, e.g. the Premier League in England,

the Bundesliga in Germany, Seria A in Italy and the Primera division in Spain. The Danish Superliga in Denmark uses a triple RR in which each team plays every other team once at home, once away, plus one more time home or away dependent on the team's placement in the previous season ($l = 3$). In the Austrian Bundesliga, 10 teams play a double, home and away RR. Each team plays every other four times, twice at home and twice away ($l = 4$). In an incomplete RR, each team does not play every other team an equal number of times even though the number of ties each team plays is usually the same. The Belgian football league have considered introducing more exotic designs along these lines. These are discussed in GOOSSENS, BELIEN and SPIEKSMA, (2010).

In the KO design or elimination tournament, teams compete in head-to-head matches. In every match, there are only two outcomes: win or lose. The winning team progresses to the next round and the losing team is eliminated from the tournament. The sole unbeaten team remaining after the final round is the 'champion'. In this design, the matches are played in rounds. With n teams and n an integer power of 2, there are $\log_2 n$ rounds. In each round, half of the teams are eliminated. If $n \neq 2^k$ for integer k, byes occur when teams do not play in a round and progress by default.

The fundamental RR and KO structures can be combined to create hybrid structures. In the one group RR, KO structure (1GRR-KO or 1G-KO for short), disjoint subsets of competitors play a RR, in groups, and those that progress then play in KO rounds until the end of the tournament. This structure has been used for the FIFA World Cup Finals since 1998. Previously, a two group RR, KO (2GRR-KO or 2G-KO for short) was used in 1974, 1978 and 1982. Structures in which every round is a group round may be constructed. Such a design was used in 1950. A full history of World Cup Finals design follows in section 4.1.

Besides the structure of a tournament, the equalisation of competitive strengths is an important factor in characterising a tournament design and hence the tournament outcome uncertainty. As the degree of competitiveness increases (i.e. the variation in the relative strengths of competitors becomes smaller), the uncertainty of the tournament outcome will also increase. The excitement generated because of the uncertainty of the tournament outcome has a significant impact on the success of a tournament. Therefore, in order to ensure a tournament will be exciting and competition is balanced, especially in later stages of a tournament, seeding is used to control the pattern of competitive balance through the competition. When seeding, one assigns competitors to a particular position in the tournament schedule based on certain criteria. By introducing seeding, one can avoid matching up the best teams in the early rounds, increasing the probability that all of the best teams progress from early rounds, and thus balancing the competition in later rounds. Thus, the chances of the tournament being won by one of the best teams is ultimately increased (using pre-tournament ranking to define 'best' here). There are a number of ways for carrying out seeding and these seeding policies are described next. Various studies have been carried out in the past to explore seeding, but this has been done only in the context of a KO tournament, see for example the work of HWANG (1982), APPLETON

(1995), SCHWENK (2000), MARCHAND (2002) and GLICKMAN (2008). We extend this work to various structures and systematically compare seeding policies.

3. Seeding policies

3.1. Standard seeding

In standard seeding in a KO design, higher ranked teams are matched with the lower ranked teams in the first round, with the first and fourth seeds in the top half of the draw and the second and third seeds in the bottom half, etc. With $n = 2^k$ teams playing k rounds, matches between teams seeded s_i and s_j must satisfy the condition: $s_i + s_j = 2^k + 1$ in round one. In tournaments with group rounds, standard seeding proceeds as in the KO design, but now the first four in the draw are put in the group 1, second four in group 2, etc. Standard seeding is widely used in KO structures (e.g. in NCAA Basketball and tennis Grand Slam events).

3.2. Cohort seeding

In cohort seeding, n teams are divided into m cohorts, where the teams in cohort C_v seeded $2^{v-1} + 1, \ldots, 2^v$ for $v = 2, \ldots, m$ are allocated randomly in the tournament draw. C_1 always comprises teams 1 and 2. For example, in a KO design, each team in the cohort of the first and second seeds has an equal chance of being selected either in top half or bottom half of the draw. In the group rounds design, the first four teams in the draw are put into group 1, second four in group 2, etc. This seeding was proposed by SCHWENK (2000). He generalised cohort seeding with varying numbers of $n = 2^k$ teams.

3.3. Reseeding at each round

In this type of seeding, teams are seeded after each round and the seeding draw is continuously updated, with standard seeding used at each round. Thus in the first round of a KO, higher ranked teams are matched with the lower ranked teams – as in standard seeding. However, in the second round, the highest surviving seed plays with the lowest surviving seed; the second-highest surviving seed plays the second-lowest surviving seed, etc. HWANG (1982) proposed this seeding policy to ensure monotonicity in seeding in each round. It favours the highest seeded teams. In practice, no sport tournament uses this design. It has been used only in comparative studies in DAVID and SAMUEL (2006) and GLICKMAN (2008).

3.4. m Pot seeding

Here 2^k teams are divided into m pots. Then teams are chosen at random from the pots. With $m = 2$, we divide 2^k teams into two pots. The first pot is reserved for

the top half and the second pot is allocated to the bottom half of pre-tournament ranked teams. In a KO, one team is chosen randomly from each pot and these are paired or matched in the first round. In a tournament with group-rounds two teams from each pot are chosen at random and put it in group 1, etc. For $m = 4$, 2^k teams are divided into four pots with the first pot for the highest ranked teams, second pot for the second highest, etc. In a tournament with group rounds, one team from each pot is selected at random and put into group 1, etc. If $2^k = m$, then we have standard seeding. Two pot seeding is generally used in soccer tournaments (e.g. FIFA World Cup Finals and UEFA Champions League), although other complicating factors, such as a geographical constraint, can come into play.

3.5. Unseeded, random allocation

Here, the placement of the teams across a tournament draw is chosen at random. In a KO, each team has an equal chance to play with every other team in the first round. In a tournament with group rounds, n teams are randomly assigned into r groups of size s in the first round. The subsequent round pairings are based on the progression from the earlier round. The FA Cup in England is an example of sports tournament that uses random allocation.

4. Analysis

4.1. Tournament designs

The tournament structure is the part of tournament design which specifies how the matches are arranged without specifying which particular teams will be paired. In this paper, we limit our discussion to structures as follows; RR, all group rounds (4G), two group rounds and KO in later rounds (2G-KO), one group round and KO in the later rounds (1G-KO) and the KO.

The FIFA World Cup Finals have used a number of such tournament structures since the first tournament in 1930. The first finals used a 1G-KO structure comprising 13 teams, although 16 had been expected. The following two tournaments were KO (1934, 1938). The next tournament in 1950 used a 16, 2G structure, although only 13 teams participated. This was the only occasion at which there was no 'World Cup final' – the final round was a group round and the winner was determined 'on points'. In 1954, a 16, 1G-KO structure was used but the first round groups were incomplete RRs. The top two seeds played the bottom two seeds, and then if second and third place teams in the group were tied, these teams played-off for the second qualification spot. This structure was never used again. A complete 16, 1G-KO structure was subsequently used through to 1970. A 16, 2G-KO tournament was used for the first time in 1974 and again in 1978. In 1982, the tournament was expanded to 24 teams, and a 2G-KO format was used with the second round

comprising groups of three. Then in 1986 the tournament reverted to a single group round with six groups of four. The top two from each group progressed to a last 16 KO, along with the four best third placed teams (as measured by points and goal difference). This structure persisted until 1998 when the tournament was expanded to 32 teams using a simple 1G-KO structure. This structure has persisted until the present. So in their simple forms, the structures we study cover the World Cup formats. The RR is included for comparison purposes – it is not a practical structure for a 3-week tournament as too many matches are required. In our study we restrict ourselves to 32 teams.

In the RR, each team plays each other team once. For the hybrid structures such as 4G, 2G-KO and 1G-KO, we divide the 32 teams into eight groups of size four. Each group comprises six RR matches, with each team playing three matches. The top two teams from each group advance to play in the second round. In the 1G-KO structure, the second round and beyond is settled by KO rounds. In the 2G-KO structure, the qualifiers from round one play again in a group in the second round; rounds three, four and five (the final) are KO rounds. For the 4G structure, all rounds are played in groups; the teams that qualify from group-round three proceed to play again in the final group of size four. The winner of the final round (league) is the champion of the tournament.

The seeding of the World Cup Finals tournaments has generally followed a 2-pot policy, one pot comprising seeded teams, the other unseeded teams. The draw then proceeds randomly but with geographical constraints with, for example, at most two European teams per first round group. These geographical constraints mean that the number of feasible seeding policies is limited. In our study, in order to explore the effects of seeding policy, we do not use additional constraints. Standard seeding would be likely to violate a geographical constraint although the ranking of teams might be done in such a way as to avoid this problem. Then, however, the ranking of teams may not be according to merit. Arguably, the ratings of teams (that is their measured strengths) may be very close for the majority of teams and so a number of rankings may be equally valid.

4.2. *Modelling match outcome*

In this paper, tournament structures are compared using the results of tournament simulations. To simulate a tournament, we require a model of match outcome. Match outcomes in soccer can be defined in terms of the scores and it is useful to do so here in order to resolve progression from groups where teams are tied on points. In predicting the number of goals scored, note that the probability that ball possession will result in a goal is small but the number of ball possession events is large. Thus, if we assume a constant ability over time of a team to score goals, the Poisson distribution seems a plausible model (Maher, 1982; Lee, 1997; Karlis and Ntzoufras, 2003). Letting X_{ij} and Y_{ij} be the number of goals scored in the match between teams indexed i and j, then a model for match scores is given by

$(X_{i,j}, Y_{i,j}) \sim \text{Poisson}(\alpha_i \beta_j \gamma, \alpha_j \beta_i)$ for team i playing at home

$\qquad\qquad \sim \text{Poisson}(\alpha_i \beta_j \kappa, \alpha_j \beta_i \kappa)$ for both teams playing at neutral ground

$\qquad\qquad \sim \text{Poisson}(\alpha_i \beta_j, \alpha_j \beta_i \gamma)$ for team j playing at home

where $X_{i,j}$ and $Y_{i,j}$ are independent variables, and α_i, $\beta_i > 0 \forall i$. The α_i measure the attacking strength of the teams, the β_j measure the defensive weakness and γ and κ are parameters that allow for home and neutral effects, respectively.

To predict match scores, other more sophisticated models are available, e.g. Dixon and Coles (1997), Rue and Salvesen (2000), Koning *et al.* (2003), McHale and Scarf (2010). Dixon and Coles (1997) developed a model of two interacting birth processes, and allow the attacking and defensive parameters for all teams to vary over time. Rue and Salvesen (2000) also consider time varying parameters in a random walk model. McHale and Scarf (2010) predict match outcomes using a bivariate negative binomial distribution with general dependence structure. We use the simpler Maher model here as the focus is on tournament simulation and not on the models themselves.

We use data on the results of 10,151 international soccer matches for 175 national teams from 1 January 1996 until 8 June 2010. National teams that played less than 10 matches within the period were excluded. These match outcomes were collected from two main sources, the official FIFA website (http://www.fifa.com) for match outcomes from 2002 to 2010 and RSSSF website (http://www.rsssf.com) for the match outcomes from 1996 to 2001.

To estimate the model parameters, we use maximum likelihood estimation. The log-likelihood is maximised using the *R* package (R Development Core Team, 2009), and we estimate a total of 352 parameters. This includes 175 attacking strength parameters, 175 defensive strength parameters, one home advantage parameter and one neutral advantage parameter. In order for the maximisation problem to be of full rank, one parameter is set equal to one, here South Africa, and all other attacking strengths and defensive strengths are relative to this value.

To demonstrate our methodology to compare different types of tournament design, we choose the 32 national teams that qualified for the FIFA World Cup Finals in 2010. The strength parameters for these teams are shown in Table 1. Note that the pre-tournament rank in this paper is determined based on the final ranking after simulating a very large, repeated RR tournament. The teams with the highest total number of points is given top pre-tournament rank, etc. The actual FIFA ranking at the time of the 2010 tournament might have been used here, although generally World Cup tournament seeding is based on both the FIFA ranking and performance in previous World Cup Finals and the actual ranking used was unknown to us.

4.3. Tournament metrics

In order to measure tournament outcome uncertainty, we use the following: the probability $P_{q,R}$ that the team in the top $100q$ pre-tournament rank percentile

Table 1. Maximum likelihood estimates for 32 teams that qualified for FIFA 2010 World Cup Finals

Pre-rank	Teams	α	β	Pre-rank	Teams	α	β
1	Brazil	2.81	0.54	17	South Korea	1.39	0.66
2	France	2.31	0.40	18	Japan	1.45	0.71
3	Spain	2.27	0.44	19	Chile	1.54	0.74
4	Netherlands	2.32	0.45	20	Switzerland	1.61	0.84
5	Portugal	2.28	0.52	21	Slovakia	1.31	0.76
6	Argentina	2.26	0.53	22	Ivory Coast	1.69	0.98
7	England	2.11	0.49	23	Australia	1.45	0.85
8	Italy	1.88	0.43	24	Cameroon	1.22	0.72
9	Germany	2.45	0.67	25	Nigeria	1.30	0.76
10	United States	1.62	0.57	26	Honduras	1.59	1.01
11	Denmark	1.84	0.65	27	Slovenia	1.29	0.95
12	Mexico	1.84	0.66	28	South Africa	1.00	0.92
13	Serbia	1.26	0.45	29	Algeria	0.96	1.09
14	Greece	1.47	0.63	30	Ghana	0.99	1.12
15	Paraguay	1.51	0.68	31	New Zealand	0.86	1.19
16	Uruguay	1.64	0.76	32	North Korea	0.76	1.17

Note: Home advantage parameter, $\gamma = 1.20$ and neutral advantage parameter, $\kappa = 1.13$

progresses forward from round R. The collection $P_{q,R}$ versus q for all R is the tournament outcome characteristic. In a tournament with 2^k teams playing k rounds in which half the teams are eliminated at each round, it follows that $P_{q,R}$ with $R=k$ and $q=2^{-k}$ is the probability that the best team wins; that is, the probability that the team with the highest pre-tournament rank wins. This study considers 32 teams playing in five rounds, so that $P_{1/32,5}$ is the probability that the best team wins, $P_{1/16,4}$ is the probability that a team which is in the top 2 (based on pre-tournament ranking) progresses beyond round 4, $P_{1/8,3}$ is the probability that a team in the best four progresses beyond round 3, etc. In a tournament without elimination, such as a RR, $P_{q,R}$ is the probability that a team in top $100q$ pre-tournament rank percentile finishes the tournament in the top $100 \times 2^{-R}\%$. In a tournament with perfect competitive balance, we will have $P_{q,1}=2^{-1}$, $P_{q,2}=2^{-2}$, $P_{q,3}=2^{-3}$, etc., for all q.

Besides $P_{q,R}$, we use Spearman's correlation coefficient between the pre-tournament rank and the exit rank (finishing position) of each team. If the Spearman's correlation coefficient is close to 1, it means that there is relatively little movement by competitors above or below their expected performance and therefore minimal tournament outcome uncertainty. This may be due either to low competitive balance or due to a tournament design that is highly discriminating. A limitation of this metric is that, apart from in the RR structure, it will need to be able to handle tied values since many competitors have tied exit ranks – in a single elimination tournament (KO) with 2^k teams, 2^{k-1} teams will have equal exit rank $2^{k-1}+1$, for example.

In a third (set of) tournament metric(s), we also calculate the average rank of competitors progressing beyond each round. A number of other tournament metrics that we do not use here are discussed by MCGARRY (1998).

5. Results

Our results were calculated using Monte Carlo simulation. Match outcomes were simulated using the double Poisson model of section 4.2 with the estimates as in Table 1. Each tournament design, corresponding to a row in Table 2 was simulated 5000 times ($n = 5000$). The values of a number of tournament metrics are shown in Table 2. From these values, we can see that the seeding policy has a significant impact on the progression of the best teams. In each class of design, the average rank of progressing teams is lower in a seeded structure than the unseeded (randomly seeded) structure. For example in the 2G-KO structure, the average rank of the team progressing beyond round one is 11.60 (standard), 11.57 (2 pot), 11.49 (4 pot), 11.54 (cohort), 11.60 (reseeded) and 12.57 (random, unseeded).

The RR design is the design that maximises the correlation of the pre-tournament to exit ranks. The correlation measure also shows that random seeding designs appear to maximise the outcome uncertainty. In each class of structure, the correlation of pre-tournament rank to exit rank is lower for the unseeded, random scheme than for the seeded scheme. However, the rank movement of the teams in each class of seeding method is similar. Again in the 2G-KO design, the correlations are 0.59 (standard), 0.59 (cohort), 0.59 (2 pot), 0.60 (4 pot) and 0.60 (reseeded).

Figures 1 and 2 show the tournament outcome characteristic curves, $P_{q,R}$ for the 23 different designs. We can see that the unseeded KO design has the flattest curves

Table 2. Estimates of tournament metrics for various tournament designs (32 teams in a five round tournament; groups of four where applicable)

| Design	Pre-to-exit rank correlation	Average rank R1	Average rank R2	Average rank R3	Average rank R4	Average rank of winner	Probability wins highest rank
RR	0.93	9.17	4.92	3.58	3.07	2.90	0.20
4G (Random)	0.50	12.57	9.13	6.72	5.43	4.85	0.15
4G (Standard)	0.59	11.59	8.48	6.34	5.12	4.66	0.16
4G (2 pot)	0.60	11.57	8.53	6.43	5.30	4.86	0.15
4G (4 pot)	0.60	11.49	8.40	6.33	5.23	4.80	0.14
4G (Cohort)	0.60	11.54	8.45	6.36	5.23	4.70	0.16
2G-KO (Random)	0.50	12.57	9.13	7.49	6.26	5.41	0.14
2G-KO (Standard)	0.59	11.60	8.48	7.05	5.93	5.12	0.15
2G-KO (2 pot)	0.59	11.57	8.53	7.10	5.93	5.34	0.13
2G-KO (4 pot)	0.60	11.49	8.40	6.98	5.90	5.28	0.13
2G-KO(Cohort)	0.59	11.54	8.45	7.10	5.86	5.21	0.13
2G-KO (Reseeded)	0.60	11.60	7.83	7.22	6.09	5.13	0.16
1G-KO (Random)	0.47	12.57	10.15	8.22	6.86	5.78	0.13
1G-KO (Standard)	0.55	11.59	10.21	8.43	6.82	5.83	0.13
1G-KO (2 pot)	0.57	11.57	9.39	7.82	6.63	5.70	0.13
1G-KO (4 pot)	0.58	11.49	9.19	7.58	6.30	5.40	0.14
1G-KO (Cohort)	0.57	11.54	9.24	7.58	6.27	5.44	0.14
1G-KO (Reseeded)	0.54	11.60	10.09	9.42	8.73	6.84	0.16
KO (Random)	0.35	13.81	11.26	9.15	7.47	6.11	0.12
KO (Standard)	0.48	12.37	10.30	8.77	7.38	6.33	0.10
KO (2 pot)	0.42	13.12	10.74	8.79	7.20	6.00	0.13
KO (Cohort)	0.42	13.06	10.91	9.18	7.84	6.65	0.10
KO (Reseeded)	0.50	12.37	9.67	7.94	6.98	6.85	0.14
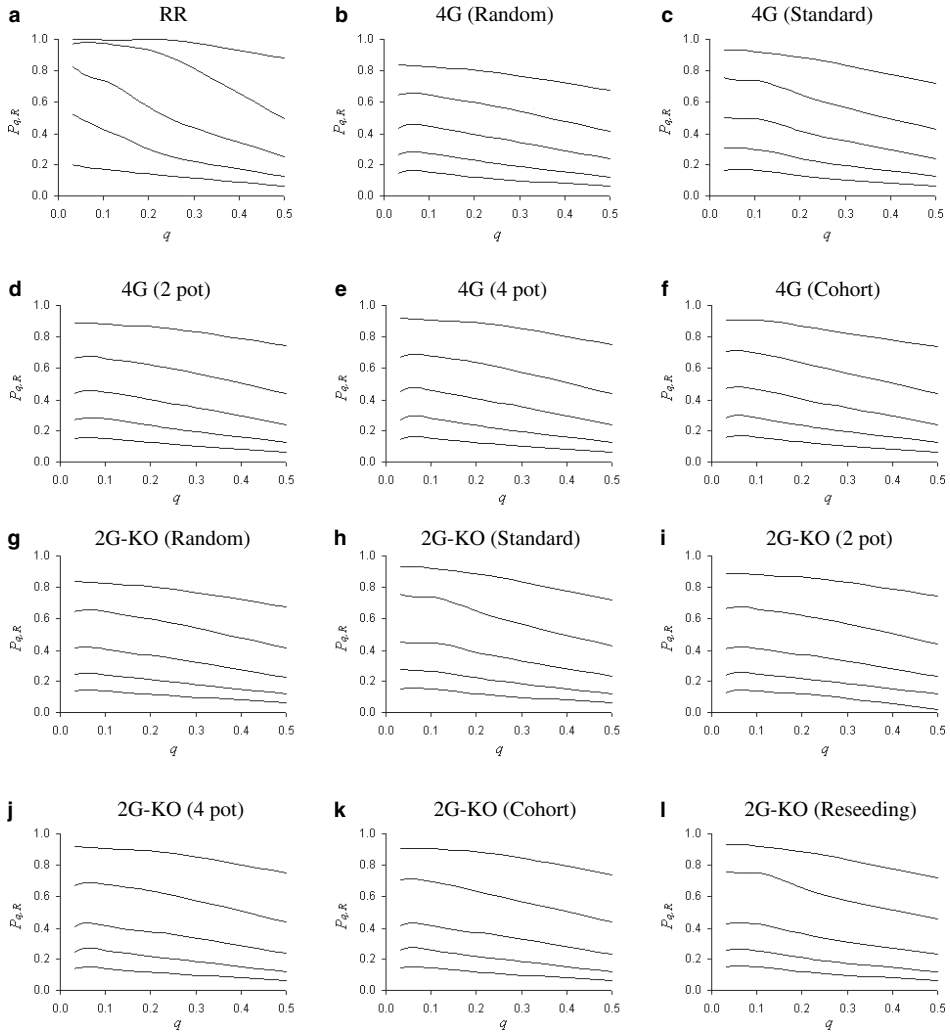
Fig. 1. The tournament outcome characteristic curves $P_{q,R}$ against q for various rounds, R, for RR, 4G and 2G-KO tournament structures for various seeding policies. Top curve $R=1$ (progression from round 1), next curve $R=2$ (progression from round 2),..., bottom curve $R=5$ (winning the tournament)

when compared with other designs, and again with all designs, seeding appears to increase the tournament outcome characteristic and hence to decrease tournament outcome uncertainty. Therefore, based on the fixed competitive balance we have chosen (that reflects the typical competitive balance of a FIFA World Cup Finals tournament), in the unseeded designs the outcome is more uncertain. Also, the greater the KO element in the tournament, the greater the outcome uncertainty. It appears that seeding has less of an effect when the number of group stages is larger.
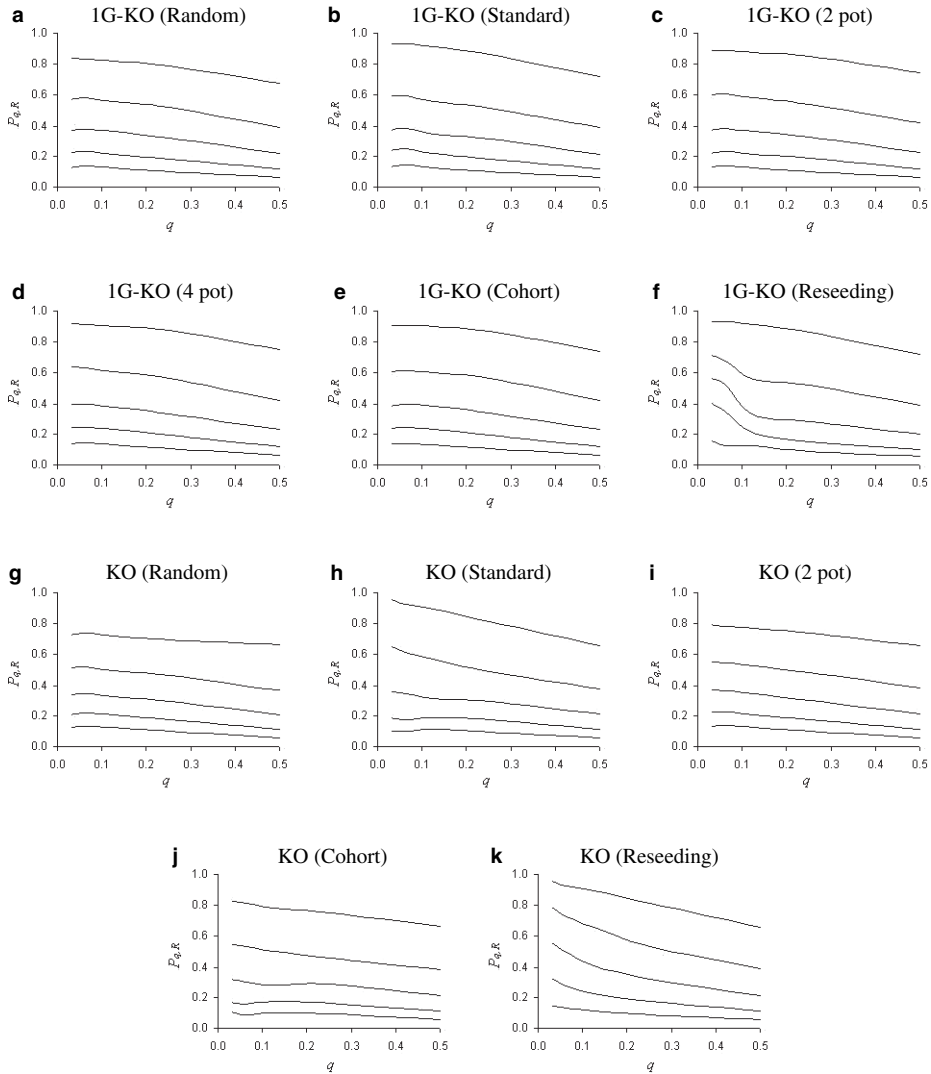
Fig. 2.   The tournament outcome characteristic curves $P_{q,R}$ against q for various rounds, R, for
1G-KO and KO tournament structures for various seeding policies. Top curve $R=1$ (pro-
gression from round 1), next curve $R=2$ (progression from round 2),..., bottom curve $R=5$
(winning the tournament)

Reseeding can be seen to have quite a dramatic effect on the tournament outcome
uncertainty, and appears to favour the higher ranked teams to quite a degree. The
favouring of the highest ranked teams then appears to be next greatest for standard
seeding. Cohort and pot seeding have a more rank neutral effect. Seeding aside,
2G-KO and 1G-KO are two types of tournament structure that have been used for
the FIFA World Cup Finals in the recent past.

6. Discussion

We describe how the structure of a tournament and the relative strengths of the competitors, the competitive balance, influence the tournament outcome uncertainty. We explore the implications of different types of structure, and study the effect of seeding on the progression of competitors given the competitive balance of a typical tournament. Knowledge of the implications of seeding and structural changes can then be used to support the work of a tournament designer, in order to create a fair and efficient tournament. Based on the results of our analysis, we show that seeding increases the probability that the best teams qualify beyond early rounds, but that seeding policy on the whole reduces tournament outcome uncertainty.

Based on the six types of seeding method we study, we find that reseeding at each round favours the top teams to the greatest extent, followed by standard seeding. These effects persist over a range of tournament structures. Within specific tournament structures, there is little to choose between *m*-pot seeding and cohort seeding.

Whether seeding is fair is another matter. Seeding is designed to increase competitive balance in the later stages of a tournament. Therefore, by design it favours the top teams over the weaker teams. To us, giving one team an advantage over another seems to contradict normal notions of fairness. Furthermore, seeding and the implicit pre-tournament ranking of competitors often involves a subjective element. We therefore conclude that seeding is not 'fair'; seeding is designed to increase the discriminatory power of a tournament.

Our results depend to a small extent on how competitive balance is modelled. Essentially, we model the competitive balance using Maher's model (MAHER, 1982), and estimate parameters in this model using 14 years of data with matches equally weighted. This approach therefore has the limitations that the competitive balance is fixed, and that it may not be representative of the team strengths in the tournament of interest. The first of these can be overcome by repeating the calculation of tournament metrics for various sets of competitors. One could for example, in a base set select the best 32 teams; and in a less competitive tournament select 32 teams from the best $n > 32$. To allow for the development of team strengths over time, we might give more weight to recent matches in the estimation procedure or use a more sophisticated model. To model the effect of the tournament itself on team strengths is a much more difficult problem, although estimates of strength for a competitor based on results from previous runs of the tournament in question might be compared with estimates of strength based on results achieved between tournaments. Another difficulty lies with the determination of pre-tournament rankings and the fact that no unique, true ranking exists. The official (FIFA) rankings themselves could be used, although this ranking system involves a subjective element and its reliability has been criticised (MCHALE and DAVIES, 2007). Another solution is to base the rankings on the values of estimated attack strengths and defence weaknesses. We have also only considered six types of seeding policy. Other seeding schemes are possible.

Designing tournaments that are both fair [in the sense of McGarry and Schutz (1997)] and uncertain in outcome is an interesting challenge. While extended tournaments with preliminary rounds provides one possible solution, tournaments that play an incomplete RR, in which teams of similar strengths are matched and novel points systems are used, may provide another. Such a potential solution might be explored.

The methods we describe would also be useful for considering design changes to a tournament of the type considered in Goossens *et al.* (2010). Such design changes have occured for example in the Cricket World Cup. The methods could also be used for studying rule changes in individual matches and changes to scoring systems of the sort studied by Percy (2009) in badminton.

Acknowledgement

References

Appleton, D. R. (1995), May the best man win? *The Statistician* **44**, 529–538.

David, H. A. and S. W. Samuel (2006), A comparison of potential playoff systems for ncaa i-a football, *The American Statistician* **60**, 151–157.

Dixon, M. J. and S. G. Coles (1997), Modelling association football scores and inefficiencies in the football betting market, *Applied Statistics* **46**, 265–280.

Glickman, M. E. (2008), Bayesian locally optimum design of a knockout tournament, *Journal of Statistical Planning and Inference* **21**, 2117–2127.

Goossens, D., J. Belien and F. Spieksma (2010), Comparing league formats with respect to match unimportance in belgian football, *Annals of Operations Research*. DOI: 10.1007/510479-0764-4.

Hwang, F. K. (1982), New concept in seeding knockout tournaments, *American Mathematical Monthly* **89**, 235–239.

Karlis, D. and I. Ntzoufras (2003), Analysis of sports data using bivariate poisson models, *The Statistician* **52**, 381–393.

Koning, R. (2000), Competitive balance in dutch soccer, *Journal of the Royal Statistical Society D* **49**, 419–431.

Koning, R., M. Koolhaus, G. Renes and G. Ridder (2003), A simulation model for football championships, *European Journal of Operational Research* **148**, 268–276.

Lee, A. J. (1997), Modelling scores in the premier league, *Chance* **10**, 15–19.

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

Marchand, E. (2002), On the comparison between standard and random knockout tournaments, *The Statistician* **51**, 169–178.

McGarry, T. (1998), On the design of sport tournaments, in: J. Bennett (ed.), *Statistics in Sport*, Edward Arnold, London, pp. 199–217.

McGarry, T. and R. W. Schutz (1997), Efficacy of traditional sport tournament structures, *Journal of the Operational Research Society* **48**, 65–74.

McHale, I. and S. Davies (2007), Statistical analysis of the effectiveness of the fifa world rankings, in: J. Albert and R. Koning (eds), *Statistical Thinking in Sports*, CRC Press, Florida, pp. 77–89.

McHale, I. and P. A. Scarf (2010), Modelling the dependence of goals scored by opposing teams in international soccer matches, *Statistical Modelling* (forthcoming).

Percy, D. F. (2009), A mathematical analysis of badminton scoring systems, *Journal of the Operational Research Society* **60**, 63–71.

R Development Core Team (2009), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org, ISBN 3-900051-07-0.

Rue, H. and O. Salvesen (2000), Prediction and retrospective analysis of soccer matches, *The Statistician* **49**, 399–418.

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

Schwenk, A. J. (2000), What is the correct way to seed a knockout tournament, *The American Mathematical Monthly* **36**, 72–80.

Yusof, M. M. and P. A. Scarf (2009), Modelling the effect of seeding policy in tournament design, in: R. Koning and P. A. Scarf (eds), *IMA Sport 2009*, The Institute of Mathematics and Its Applications, UK, pp. 191–196.