European Journal of Operational Research 198 (2009) 190–198

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

A numerical study of designs for sporting contests

Philip Scarf *, Muhammad Mat Yusof, Mark Bilbao

Centre for Operational Research and Applied Statistics, Salford Business School, University of Salford, Salford, Manchester M5 4WT, UK

ARTICLE INFO

Article history:
Received 14 September 2006
Accepted 26 July 2008
Available online 16 September 2008

Keywords:
OR in sport
Sports tournaments
Simulation
Competitive balance

ABSTRACT

Operational Research may be used to compare different designs for a sporting contest or tournament. This paper considers a methodology for this purpose. We propose a number of tournament metrics that can be used to measure the success of a sporting contest or tournament, and describe how these metrics may be evaluated for a particular tournament design. Knowledge of these measures can then be used to compare competing designs, such as round-robin, pure knockout and hybrids of these designs. We show, for example, how the design of the tournament influences the outcome uncertainty of the tournament and the number of unimportant matches within the tournament. In this way, where new designs are proposed, the implications of these designs may be explored within a modelling paradigm. In football (soccer), the UEFA Champions League has adopted a number of designs over its 50 year history; the design of the tournament has been modified principally in response to the changing demands of national league football and television – the paper uses this particular tournament to illustrate the methodology.

1. Introduction

The success of a major tournament can have a significant influence on the development of a sport. Szymanski (2003) suggests "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". Often, major sporting events receive enhanced national and global coverage that is in contrast to the regular season. Therefore, the success of a global tournament, for example, can boost the interest in a particular sport. The Rugby Football Union of England estimates some 51,000 new players took up the sport in the 2 years following the success of the England team in the 2003 World Cup, representing a 25% growth in the playing base. At the time of this tournament, England were considered to be the number 1 ranked team (Official IRB World Rankings 6 October 2003: First England, second New Zealand, third Ireland). Thus, for this tournament the "best" team won. However, there are contrasting viewpoints with regard to the desirability of this outcome. Owen and Weatherston (2004) suggest that the consequence of low uncertainty of outcome is that fan, broadcaster and sponsor interest will not be maximised. In contrast Szymanski (2003) states that in both Rugby Union and Cricket, "competitive balance plays no obvious role in the popularity of these sports". The following question then arises: can a tournament be considered successful if the "best" team wins? Or more generally: what defines a successful tournament? Does the format of a tournament influence its success? This paper seeks to address these questions using a modelling approach.

A number of high profile events have received regular modification to their structure, for example the cricket and football (soccer) World Cup tournaments. Also in football, the UEFA Champions League has seen perhaps the most significant change since its inception in the form of the old European Cup. Originally, the competition was a 2 leg (home and away) knockout or straight elimination competition with no seeding or late round entry; the competition took place between the winners of each of the national football leagues in Europe. Later, the design was adapted to include a group-round, and later still two group-rounds, and the possibility of more than one team from each European nation entering the competition. In the group-round, teams in each group play a round-robin to determine two qualifiers from the four in the group. Thus, the extension to group-rounds increased the number of games for successful teams. The second group-round was then replaced by a knockout round in 2003 to make the tournament more exciting. Recently, UEFA has been coming under pressure from the G14, a group of the wealthiest clubs in Europe, to again increase the size of the tournament and there is speculation about a tournament re-design (Sinnott, 2006). As Owen and Weatherston (2004) point out, "spectator interest is likely to be maintained the longer more teams remain in contention". However, there may be a limit to the number of games that a (television) audience may want to appreciate. Furthermore, with the nature of these round-robin group-rounds, the later matches in the group often have no bearing on the progression of the teams involved. For example, in the recent 2006 FIFA (football) World Cup, the final two ties of the first round group (of four teams) appear to have been

* Corresponding author. Tel.: +44 161 295 3817.
E-mail address: p.a.scarf@salford.ac.uk (P. Scarf).

0377-2217/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2008.07.029

scheduled between the two seeded teams in the group and between the unseeded teams in the group. The qualifiers from a number of the groups had been determined prior to these two final games, however, and so these matches in the group-round between the best teams in the group were often of little importance. These points suggest three further factors in tournament design that are relevant in the design of a tournament: the minimum number of games each competitor plays; the importance of individual matches; and the order of matches in group-rounds.

In 2007, the Cricket World Cup changed its format for a fourth straight tournament. Historically for a sport that finds national coverage often difficult to achieve, the World Cup and the success of the national team is crucial to the survival of the sport. In 2003 when political turmoil engulfed the proposed match between England and Zimbabwe, the refusal of England to travel to Zimbabwe had huge implications on the tournament outcome. This single game resulted in England failing to qualify for the second round; it also facilitated the progression of a second class team to the final stages, which in turn impacted upon the tournament as a whole. Interest in the tournament dropped not only with the exit of England but also because a significant number of matches in the final stages were grossly one-sided. A more robust tournament design might have coped with this situation; in future modelling could be used to find such a design.

Thus, we argue that Operational Research can assist in analysing the effect that certain tournament formats have on particular aspects of the event. In addition, given a particular metric as the tournament design criterion, and using other metrics in design constraints, it will be possible to choose the best design (from a set of designs) for a future tournament. This paper considers these issues. In the next section, we describe a number of tournament designs. Metrics that may be used to assess tournament designs are then proposed and described. We then discuss a model that can be used in order to calculate tournament metrics and so explore tournament designs. We conclude with an example relating to the UEFA Champions League.

The approach described could also be used to consider rule changes within a sport. For example, the effect of changes to point scoring systems on the outcome of individual matches could be explored. An example in football would be changes to the rules for determining the result when a match is level at full-time. Various schemes relating to goals scored in extra time have been used in the recent past such as the "golden goal" (first to score in extra time wins) or the silver goal (the second period of extra time is only played if the scores are still level at the end of the first period of extra time). More radical ideas could also be explored through modelling, such as determining match results, when goals scored are tied at the end of extra time, using shots (on target, say).

Although the scheduling of matches within a given tournament format is strictly a tournament design problem, we do not consider this here. Match scheduling is however firmly an OR problem and has been extensively studied in the literature (e.g. Anderson, 1997; Wright, 2006).

2. Tournament designs

Two tournament designs may be regarded as fundamental, with all other designs considered as variations and hybrids. The first of these is the *round-robin* in which: every competitor plays every other; competitors earn points for wins (and draws as appropriate); and the winning competitor is defined as that with largest points score at the end of the tournament. In the round-robin, the collection of competitors is called the league. The second fundamental design is the (pure) *knockout* tournament in which: ties are played in rounds; competitors progress to round $r+1$ on winning in round r; losers are eliminated; and the tournament winner is the sole, remaining competitor – the winner of the final round.

Hybrids designs can then be constructed from these fundamental designs. Where a subset of competitors play a round-robin, this is referred to as a group, and the collection of all groups from which competitors proceed to a common position, is called a group-round or group-stage. Thus, a tournament with an initial group-round and then a knockout tournament between competitors qualifying from the group-round is a hybrid tournament. Appleton (1995) calls this an *American blocks with play-off*. Group-round qualifiers may again play in groups in a second round – in extremis, all rounds can be played in groups in an *all group-rounds* tournament – here all ties are played in group-rounds, with competitors proceeding to group-round $r+1$ on qualification from group-round r. Non-qualifiers are eliminated and the tournament winner is the winner of the final group-round – bizarre as this may sound, the 1950 football World Cup used just this format: 13 teams played in four first round groups, one team qualified from each of these groups to form a final group which played a round-robin – Uruguay were winners. Knockout rounds may precede group-rounds as in the UEFA Champions league, although not all teams play in the initial knockout rounds which form a qualifying stage of the tournament.

Variations on the fundamental designs and thus on the hybrid designs are many. Both home and away ties between the same pair of competitors may be played or ties may be repeated – the English Football Association Premier League is a *home and away round-robin*. In an *incomplete round-robin* (Cochran, 1971), not all competitors play each other but all play an equal number of ties – this format was used in the England and Wales Cricket Board county cricket championship in the recent past. Teams may be assigned to groups and ties in a knockout on the basis strength – this is called seeding and is used to maximise the probability that the best competitors meet in the later stages of the tournament. Losers in round r may be given the opportunity to re-qualify for round $r+1$ in a so-called *repechage* which is played between losers of round r. This repechage may itself be a group-round or a knockout. In Olympic judo, a variation of the repechage gives those who lose early to athletes that perform well until late in the tournament an opportunity to win a bronze medal. Teams may enter the tournament at different rounds. In English football, the FA Cup is an example of a knockout tournament with multiple entry; teams from the top two league divisions enter the competition in the third round. Groups may be formed on a regional basis as in the FIFA (football) World Cup qualifying stage. Designs can be made more complex still with qualification rounds. Qualification rules for group-rounds can be varied. The simplest is when k from each group of size l qualify, although variations may be used with best m third place finishers to qualify and unequal numbers of qualifiers if group sizes vary.

Tournament designs have largely evolved to accommodate the characteristics of the sport concerned, the competitive balance of the competing teams or individuals, and the size of the tournament. With a large number of teams, the round-robin is a large tournament. The balance of competitors may be such that novel designs are developed to maintain interest in a tournament. The ICC (cricket) 2003 World Cup used a second round, single group – the super-six – since cricketing nations are few, with teams falling into two classes that are well-separated in terms of competitive strength; ties across these two classes are very predictable. In the ICC 2007 World Cup, the second round was a single group of size eight playing an incomplete round-robin (six matches each), with the results of the other "missing ties" in the second round carried over from the first round. Two characteristics of the sport concerned that have a large influence on the design of a tournament are: (a) the possibility for a drawn tie (no winner); (b) the scoring system. Appleton (1995) discusses the sport of croquet at length. In this sport, draws cannot occur and the scoring in an individual match is largely irrelevant. Therefore, a round-robin is not guaranteed to produce a winner, and group-rounds with a small number of competitors will often not distinguish qualifiers for later rounds. A knockout tournament, on the other hand, may be unsatisfactory for championship events since 50% of competitors will play only one tie (even if ties have multiple games). A number of remarkable designs have been developed to cope with these circumstances. For example, in the *draw and process*, two knockout sub-tournaments are played between the same set of competitors – the draw for the first knockout sub-tournament, called the *draw*, is random and the draw for the second knockout, called the *process*, is arranged such that if two players meet in an early round in the first, they will not meet until a late round in the second. The two knockout sub-tournaments are then played independently. Two possibilities then arise: (a) a particular player wins both sub-tournaments, in which case this player is the winner of the tournament; or (b) different players win the two sub-tournaments, in which case they play-off in a super-final. Under this design, players are only eliminated if they lose two matches. By contrast, in association football, draws are possible and scores are relevant – then group-rounds nearly always distinguish qualifiers when taking account of points and scores (goals for, etc.). Where qualifiers are not distinguished, count backs can be used: for example, the actual tie between tied competitors may be used to determine the winner or qualifier; or the number of ties won by the competitors' defeated opponents may be counted. In one-day cricket, draws are rare and scores cannot be compared in a straightforward manner. Where teams are tied at the end of group-rounds, obscure run-rate calculations are performed to separate them. Teams can then appear to proceed to later rounds on a technicality. However, it is unlikely that a tournament design like the draw and process would be satisfactory here since a super-final is not guaranteed. In the Swiss system, competitors are not eliminated at each round, but after a random or seeded first round, competitors with similar records in the tournament are paired in subsequent rounds; thus first round winners will play each other in the second round, and likewise for losers; in round 3, competitors with two wins will be paired, competitors with a win and a loss will be paired, and competitors with two losses will be paired. The winner is the competitor with most wins after a pre-determined number of rounds. Such a design leads to fewer one-sided matches than in a round-robin tournament (Appleton, 1995).

Ties are not necessarily restricted to pairs of competitors. Group-rounds then make little sense. Knockout designs with k from n to qualify at each round are then typically used. The initial number of competitors and number that can be accommodated in a final then determines k, n, and the number of rounds (e.g., track athletics). To guarantee every competitor at least two ties or races, repechages may be used, with the repechage designed so that first round losers who qualify for a final will have raced more to get there (e.g., rowing).

3. Tournament metrics

Glenn (1960), Seals (1963), Appleton (1995) and Marchand (2002) focus on the probability that the best competitor wins the tournament. Appleton concludes that in order for a tournament to select "as winner the best of entrants", the seeded draw and process is a strong contender for the "optimal" design when it is necessary to play relatively few games, and the round-robin played twice is most effective when players are closely matched and there can be more games. Thus, the probability that the best team wins is a natural tournament metric. This metric is a measure of outcome uncertainty of the tournament. Other possible measures of outcome uncertainty have been proposed. We believe that it is helpful to distinguish between outcome uncertainty and competitive balance, with competitive balance describing the relative strengths of teams in the tournament, and outcome uncertainty describing variability in outcome – thus, competitive balance measures the systematic variability in strength of teams; outcome uncertainty measures the systematic variability plus the unexplained or random variability. Further, we can consider the competitive balance and outcome uncertainty of individual matches and of the tournament as a whole. The tournament format can influence both competitive balance of individual matches and the outcome uncertainty of the tournament – for example, seeding leads to early matches in a tournament of low competitive balance, and a knockout design will lead to high tournament outcome uncertainty even when the competitive balance of the tournament is low. For a fuller discussion of competitive balance and outcome uncertainty, see Koning (2000) who takes an historical look at competitive balance in Dutch football, or Szymanski (2003) who reviews these concepts within the context of the organisational structure and regulation of sports, and considers whether restrictive practices should be introduced in an attempt to maximise revenue. Other sports economists have considered the drivers of demand for sport – for example, it has been proposed that demand for match tickets peaks when the home side are twice as likely to win as the away (Knowles et al., 1992; Peel and Thomas, 1997).

For simplicity, we assume that the competitive strengths of teams entering the tournament are fixed, but that the tournament design can influence the pattern of competitive balance of individual ties – through seeding for example as is typical in tennis tournaments, or through the number of teams taking part in particular rounds (e.g., super-six in cricket) – and the uncertainty of tournament outcome. We propose the following as a measure of tournament outcome uncertainty: the probability $P_{q,R}$ that a team in the top $100q$ pre-tournament rank percentile progresses forward from round R. The collection of curves $P_{q,R}$ versus q for all rounds would represent the tournament outcome characteristics. 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. For example, in a 32 team tournament there are five rounds, so that $P_{1/32,5}$ is the probability that the best team wins, and $P_{1/2,1}$ is the probability that a team in the top 50% pre-tournament ranking progresses beyond round 1. In a tournament without elimination, such as a round-robin, $P_{q,R}$ is the probability that a team in top $100q$ pre-tournament rank percentile finishes the tournament in the top 100×2^{-R}%.

Some shocks may heighten level of excitement in a tournament, but may also have a serious downside – for example, the ICC Cricket World Cup suffered when one or two results that were affected by rain and political turmoil forced the exit of both England and the host nation, South Africa. Tournament designers may want to minimise the impact of shock results and maximise the probability that the better teams reach the later and more lucrative rounds; $P_{q,R}$ provides a useful metric for considering these issues. However, the outcome characteristics of a tournament will depend on its competitive balance, since a tournament with perfect competitive balance will have $P_{q,1} = 0.5$, $P_{q,2} = 0.25$, etc. Thus, one cannot determine general outcome characteristics for a particular tournament design – one can only compare outcome (uncertainty) characteristics of competing designs given a particular competitive balance structure. It will then be important to use a reasonable estimate of competitive balance structure when comparing designs. Estimates of competitive balance would ideally then be obtained based on data of previous occurrences of the tournament in question.

When comparing designs, some sensitivity analysis to the competitive balance of the participants would be recommended.

Summary metrics similar to $P_{q,R}$ can be calculated such as the (rank) correlation between the pre-tournament rank and post-tournament or exit rank (finish position) of competitors. McGarry and Schutz (1997) consider a number of other similar metrics. If the rank correlation metric is close to 1 then there is relatively little movement by sides above or below their expected performance, based on their pre-tournament rank, and tournament outcome uncertainty is low. A tournament with a correlation of 0.3 has a more competitive nature than one with a value of 0.6. The method of calculation of the correlation will need to be able to handle tied values since many competitors will have tied exit ranks – in a pure knockout with 2^k teams, 2^{k-1} teams will have equal exit rank $2^{k-1} + 1$, for example.

Other simple tournament metrics that may be considered are as follows. The number of ties in the tournament, and the maximum and minimum number of ties for an individual team or competitor are simple and important metrics. These are fundamental in relation to tournament logistics and for revenue, both from the point of view of the tournament organisers and individual competitors. For a pure knockout tournament with $n = 2^k$(k integer) teams, $n - 1$ ties are required, taking $k = \log_2 n$ rounds, competitors playing a maximum of k ties and minimum of 1. In a round-robin with n competitors, $n(n - 1)/2$ ties are required with each competitor playing $n - 1$ ties. For the all group-rounds design with $n = 2^k$ teams playing in groups of size four with two to qualify, $6 \times \sum_{i=0}^{k-2} 2^i$ ties are required with competitors playing a minimum of three ties and a maximum of $3(k - 1)$. Further, the distribution of the number of ties played by each team would be an interesting metric.

Finally, the percentage of unimportant ties in a tournament may be defined. A tie may be defined to be unimportant if it has no effect on the tournament outcome. In a knockout tournament, all ties are important. In a round-robin or a round-robin, knockout hybrid, ties may be unimportant depending on the definition of tournament outcome. If the tournament outcome is simply defined as the label of the tournament winner, then many ties in the later stages of a round-robin or group-rounds may be unimportant. If the tournament outcome is defined as the exit-ranks of all competitors, then some group-round ties in a round-robin, knockout hybrid may be unimportant – this is because the results of certain group-round ties may have no influence on the group-round finishing positions. Scarf and Shi (2008) consider this metric in some detail. These metrics are not an exhaustive list of measures, but form a useful set. Other metrics have been considered by other authors. Russell (1980) considers carry-over effects in sequences of matches in order to design fair schedules for round-robins. Lundh (2006) develops a tournament outcome uncertainty measure for individual matches in particular tournament and uses this to compare sports rather than to design tournaments.

Table 1
Attacking and defensive parameters used in the Poisson model for individual ties (UEFA Champions League simulation). Home advantage parameter, $\gamma = 1.39$

Team	Attack strength (α)	Defence strength (β)	Team	Attack strength (α)	Defence strength (β)
1	1.235	0.527	17	1.446	0.772
2	1.278	0.649	18	1.622	0.775
3	1.730	0.534	19	1.192	0.649
4	1.183	0.760	20	1.281	0.732
5	1.238	0.658	21	0.957	0.667
6	1.115	0.699	22	0.990	0.821
7	1.177	0.667	23	1.341	0.676
8	1.510	0.583	24	0.917	0.961
9	1.448	0.561	25	0.959	0.832
10	1.232	0.728	26	0.864	0.884
11	1.869	0.402	27	0.912	0.884
12	1.244	0.750	28	1.170	0.737
13	1.659	0.578	29	1.122	0.710
14	1.460	0.658	30	0.955	0.921
15	1.497	0.717	31	1.448	0.509
16	1.387	0.698	32	0.964	0.949

4. Tournament model

For a given tournament design, to calculate the majority of the metrics proposed it is necessary to use simulation. While limited results can be obtained from historical tournament data, these will not be useful for considering new designs or tournaments that occur infrequently. A prediction model for individual ties will form the basis of the tournament simulation model—this prediction model can be used to simulate the results of individual ties. Given the design, we can then simulate a complete tournament repeatedly and obtain estimates of the metrics of interest.

4.1. Prediction model for individual ties

The prediction model will depend on the sport considered. Logistic or probit regression is a natural candidate for modelling results of ties between two competitors. Such models are used in

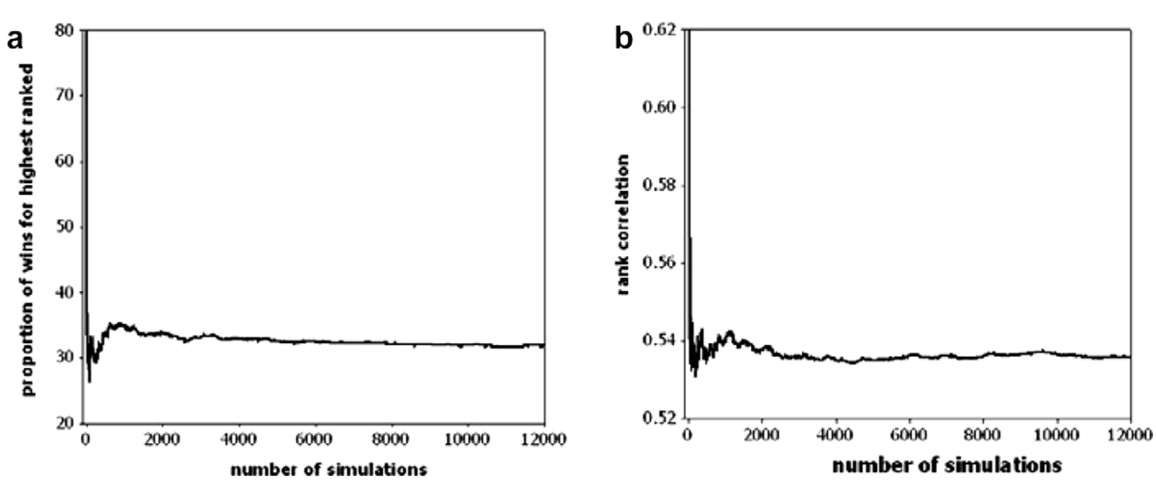

Fig. 1. Plots of: (a) the proportion of tournament wins for the highest pre-tournament ranked team, (b) the average pre-tournament to exit rank correlation, as a function of the number of tournament simulations (UEFA Champions League simulation).

association football, with simple models quantifying the strengths of the home (or reference) and away teams (Koning, 2000). In Koning's model, the outcome of the game is determined by the random variable $D^*_{ij} = \phi_i - \phi_j + \varepsilon_{ij}$, where ϕ_i and ϕ_j represent the measured strengths of team i and j respectively, and $\varepsilon_{ij} \sim N(0,\sigma^2)$ represents unexplained variation. The result is given by D_{ij}, where

$$D_{ij} = \begin{cases} 1 & \text{if } D^*_{ij} > c_2, \\ 0 & \text{if } c_1 < D^*_{ij} < c_2, \\ -1 & \text{if } D^*_{ij} < c_1. \end{cases}$$

with a win coded as 1, a draw as 0 and a loss as −1. Logistic regression is similar and has been used in cricket (Scarf and Shi, 2005). Alternatively, for football, goals for and goals against (the reference side) may be modelled using bivariate Poisson models (Maher, 1982; Lee, 1997; Dixon and Coles, 1997; Rue and Salvesen, 2000; Karlis and Ntzoufras, 2003; McHale and Scarf, 2007). For simplicity, for the numerical investigation in the following section we use Maher's model which assumes independent Poisson distributions for each of the number of goals for and against (the reference team) with means specific to the past performances of each team. Letting X_{ij} and Y_{ij} be the number of goals scored in a tie between team i and team j, then $X_{ij} \sim \text{Poisson}(\alpha_i\beta_j\gamma)$ and $Y_{ij} \sim \text{Poisson}(\alpha_j\beta_i)$, where α measures the attacking strength, β the defensive strength and γ is a parameter which allows for the effect of playing at home. For the purpose of the simulation, values for model parameters must be available. For ties with many competitors (>2), multinomial logistic or probit regression could be used. More simply, we could subjectively propose win, draw, and loss probabilities for all potential ties i versus j (all i,j).

Competitor strength may vary with the design of a tournament. For example, weaker competitors may put more effort into a tie in a knockout tournament than in a round-robin tournament as they have a better chance of success in the former. We might attempt to model this effect, and given match results for competitors in tournaments with different designs, attempt to estimate model parameters. For new designs, this presents a serious difficulty, although strengths might again be assigned subjectively. Strengths may also evolve over time. This effect could again be modelled along the line of that proposed by Crowder et al. (2002). The sensitivity of tournament metrics to competitor strengths could also be explored.

4.2. Technical aspects of tournament simulation

For a given tournament design, there will be a number of technical issues that will need to be addressed. These technical issues will arise from features of the design, such as how competitors are assigned to ties (the draw) and how draws (no result) in knockout ties are resolved. We do not go into detail here, but merely highlight issues that may arise. For example, determining whether a game is unimportant is not necessarily straightforward, and likewise for the separation of teams on equal points in group-rounds.

4.3. Choosing between tournament designs

With many metrics to consider, it may be sensible to use Multiple Decision Criteria Analysis (MDCA) (e.g. Belton, 1999). Numerous techniques are available, but a standard feature is the performance matrix with element *ij*, representing the value of metric *j* for design *i*. This performance matrix can be analysed using one of a number of techniques. With direct inspection, it may be possible to identify dominance (when one design performs at least as well as another on all criteria and strictly better than the other on at least one criterion) and trade-offs between different criteria that are acceptable, so that good performance on one criterion can in principal compensate for weaker performances in another. If it can be reasonably assumed that the criteria are preferentially independent of each other and if uncertainty is not formally built into the model, then a simple linear additive evaluation model may be used. Here, the value of a design is calculated by multiplying the value of each metric by the weight attached to that metric, and adding the weighted scores. The Analytical Hierarchy Process also develops a linear additive model, but, in its standard format, uses procedures for deriving the weights and the scores achieved by alternatives which are based, respectively, on pairwise comparisons between criteria and between designs (options). Outranking methods seek to eliminate options that are, in a sense, 'dominated'. Unlike the dominance method, dominance within the outranking

Table 2
UEFA Champions League simulation: estimates of the tournament metrics for various tournament designs (5000 simulations for each tournament design)

	Min # Games	Max # Games	Total # Games	Ave. % Unimportant games	Ave. rank of winners	Proportion of wins for highest ranked	Pre- to exit rank correlation	Ave. rank of teams in round 3/ best 8	Ave. rank of teams in semi-final/best 4
Home and away round-robin	62	62	992	18.0	1.47	0.79	0.85	6.34	4.04
Seeded, all group-rounds	6	24	180	4.0	3.89	0.41	0.51	9.60	7.16
Seeded, 1 group-round (2 leg seeded knockout in later rounds)	6	13	125	2.7	5.60	0.28	0.49	10.20	8.59
Seeded, 1 group-round (2 leg unseeded knockout in later rounds)	6	13	125	2.9	5.85	0.29	0.42	10.83	8.95
Seeded, 1 group-round (1 leg seeded knockout in later rounds)	6	10	111	2.6	6.96	0.22	0.48	10.48	9.43
Seeded, 2 group-rounds (2 leg seeded knockout in later rounds)	6	17	157	3.2	5.51	0.30	0.51	9.60	7.97
Seeded, 2 group-rounds (2 leg unseeded knockout in later rounds)	6	17	157	3.5	5.57	0.30	0.43	10.16	8.24
Seeded, 2 group-rounds (1 leg seeded knockout in later rounds)	6	15	151	3.4	5.98	0.27	0.50	9.60	8.21
Seeded, 2 leg knockout	2	9	61	0.0	6.69	0.25	0.37	11.51	9.69
Unseeded, 2 leg knockout	2	9	61	0.0	6.72	0.25	0.30	12.02	10.05
Seeded, 1 leg knockout	1	5	31	0.0	7.50	0.19	0.45	11.15	9.85

frame of reference uses weights to give more influence to some criteria than others.

5. Numerical example – simulation of the UEFA Champions League

At its conception the UEFA Champions League or European Cup was a knockout tournament for the winners of all the major domestic football leagues in Europe. This tournament was a pure (unseeded) knockout where rounds consisted of teams playing against each other home and away. By many this was considered the purest form of tournament football available. This format was used until the 1990s when the tournament was expanded to include a first round group-stage – a hybrid round-robin, knockout tournament. This change increased the minimum number of games each qualifying club played from two to six games (since groups comprised four teams playing each other home and away). From UEFA's perspective, this modification increased the number of matches that it could sell to the European broadcasters.

The inclusion of the group-round also generated unimportant games, since final group positions are often decided prior to the final game of the group. In spite of this, UEFA expanded the tournament yet further to include a further group-stage. This experiment lasted only 2 years however, when in 2003 UEFA reverted back a single, first round group, followed by knockout rounds. Throughout these changes, UEFA also experimented with seeding policy, finally settling on a seeded first and second round.

Since this tournament has undergone many changes and further change is anticipated, it is interesting to apply the ideas presented in the paper to it. The designs described may be analysed using the simulation techniques developed and designs compared. Taking the fundamental designs discussed above (round-robin, pure knockout) and the all group-rounds design as reference points, other designs can be thought of as contained within these formats. It is therefore relevant to compare working designs with the three fundamental designs: 32 team round-robin; 32 team pure knockout; 32 team all group-rounds (8 Groups of 4, then 4 Groups of 4, then 2 Groups of 4 and then a final group of 4 teams).

The bivariate Poisson model due to Maher (1982) was used to simulate the results of individual ties. This meant that for group-round ties and home and away knockout ties, the number of goals for each team could be calculated. Then, group tables could be formed using points, goals-for and goals-against. The technicality of away goals counting extra in the event of drawn matches over

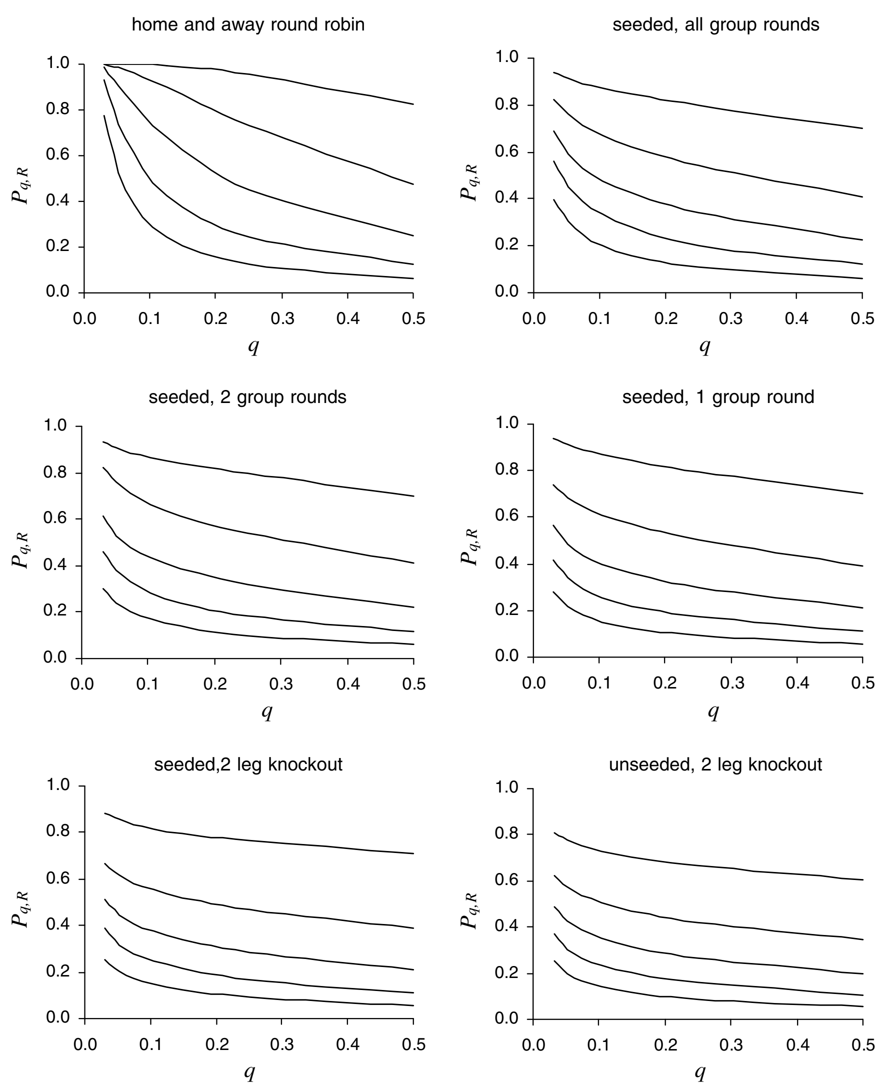

Fig. 2. UEFA Champions League simulation: tournament outcome characteristics $P_{q,R}$ against q for various rounds, R, for six tournament designs. Knockout rounds in hybrid tournaments here are 2 leg (home and away). Top curve $R = 1$ (progression from round 1), next curve $R = 2$ (progression from round 2),. . ., bottom curve $R = 5$ (winning the tournament).

two ties could also be resolved. Where ties remained unresolved, the winner was chosen purely randomly (win prob = 0.5). Estimates for the parameters of the Maher model were obtained from that paper – these were deemed to show variation representative of the 32 teams that comprise the later stages of the Champions League (Table 1). Pre-tournament ranks were based on these parameters. Note we have ignored the qualification stages in our consideration of the design (although in practice, the simulation could easily be extended). There is an argument that qualification starts in the national leagues themselves, but this is really another matter.

The tournament was simulated using Mathcad to construct the seeded draws, to obtain results of individual ties, and to form group-round tables and to determine qualifying and eliminated competitors at each round, and allowed repeat simulations for the calculation of tournament metrics.

5.1. Results

In order to obtain reasonable estimates of tournament metrics, it was determined that 5000 simulations were required (Fig. 1). Table 2 shows the values of various metrics for the tournament designs considered. Fig. 2 shows the tournament outcome characteristics for six tournament designs.

From Table 2, we can see that as expected the round-robin is the design that maximises the probability that the best team wins. Furthermore, this design also maximises the correlation of the pre-tournament to exit ranks – thus this design is in a sense the "fairest". Tournaments with group-rounds are then "fairer" in this sense than knockout tournaments with an all group-rounds tournament more likely to result in the winner being the highest ranked team than the 1 or 2 group-rounds tournament. "Fairness" in this sense is to be traded off against tournament size, obviously, and also the proportion of unimportant matches (18% for the round-robin). The round-robin and all group-rounds formats also mean that a final match to select the winner does not take place.

The tournament outcome characteristic curves, $P_{q,R}$, for the round-robin provide a reference for other designs. For the round-robin we can see that a team with high pre-tournament ranking is much more likely to finish in the top 50% than for other designs (upper curve), and that it is most "discriminating" – the slope of the lower curves is maximum among the designs. The curves for the unseeded 2 leg knockout design are flattest indicating that this design is least discriminating among the teams – here tournament outcome uncertainty is greatest.

The impact of seeding and of 2 leg rounds can be observed in Table 2. Both of these factors reduce tournament outcome uncertainty. One might say they increase fairness in the sense that the pre- to post-tournament rank correlation is higher with seeding and with 2 leg rounds. One could argue however that seeding introduces an element of unfairness since then the draw is not random, and seeded competitors have a higher probability of progressing to later rounds. The seeded team must be chosen by the tournament organisers although an element of objectivity could be introduced here by using official rankings if they exist. These are, however, often modified by the local tournament organisers, for example as in tennis where seeding reflects international rankings and local playing conditions.

5.2. Sensitivity analysis

In order to consider the sensitivity of the designs to the values of competitor strengths, the tournament metrics are calculated for more and less competitive tournaments than the original. We first fit a bivariate normal distribution to the competitor strengths of Table 1. Then, to simulate a more competitive tournament, we generate 32 realisations from a bivariate normal distribution with the same means and correlation but with the variances halved – we

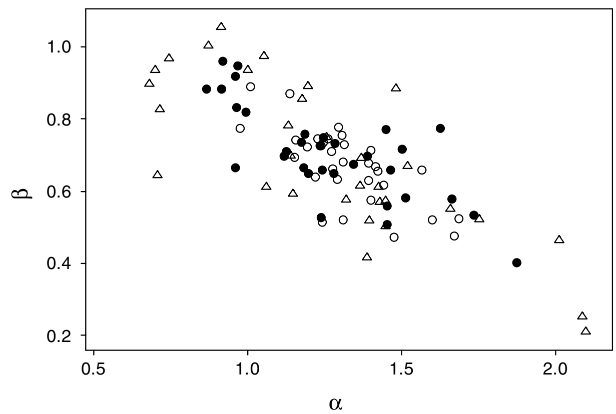

Fig. 3. Values of competitor strengths for the sensitivity analysis: •, original values; ○, more competitive; △, less competitive.

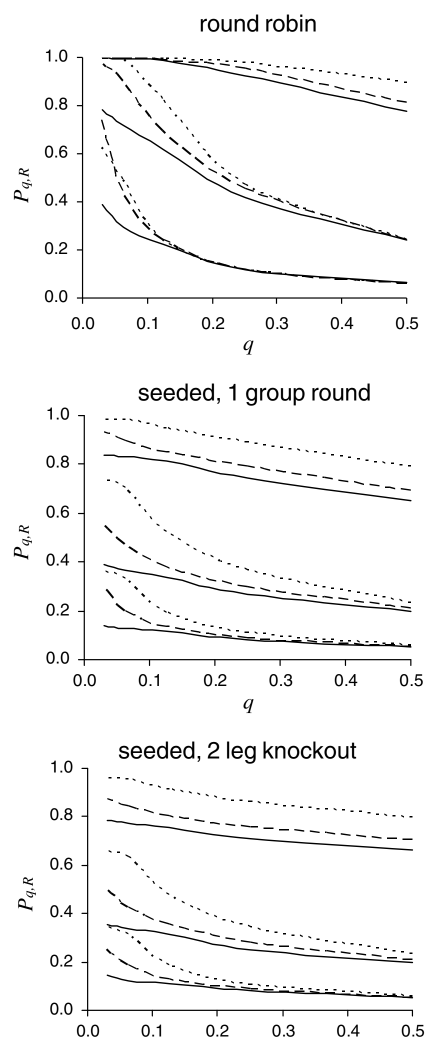

Fig. 4. Sensitivity analysis of tournament outcome characteristics $P_{q,R}$ against q for various rounds, R, for three tournament designs. Top curves $R = 1$ (progression from round 1), next curves $R = 3$ (progression from round 3), bottom curves $R = 5$ (winning tournament). (– – –) original values of competitor strengths; (—) more competitive values of strengths; (- - -) less competitive values of strengths.

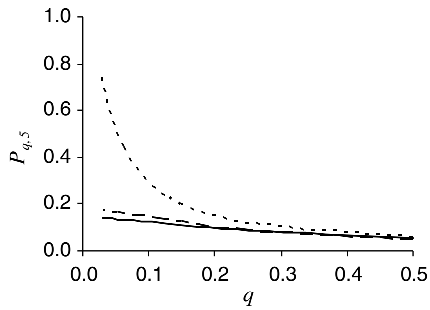

Fig. 5. Tournament outcome characteristics for $R = 5$ (winning). (---) original home and away round-robin; (—) more competitive seeded, 2 leg knockout; (– – –) more competitive seeded, 1 group-round (both with later knockout rounds over 2 legs).

then have a set of strengths (α, β) that are more tightly clustered than the original values. The less competitive tournament is simulated by doubling the variances. This then gives three sets of competitor strengths (α, β) for comparison (Fig. 3). In principle, we might use such a mechanism to produce Monte Carlo estimates of the variance of the tournament outcome characteristic, $P_{q,R}$. This can be done by first sampling a set of strengths, (α, β), simulating the tournament of interest with these strengths, calculating $P_{q,R}$, (for all q and R), and then repeating to find a Monte Carlo confidence region for the curve $P_{q,R}$, vs q (for each R). We do not pursue this further here.

The tournament outcome characteristics for the original and the more and less competitive tournaments are plotted in Fig. 4. These behave as expected. We can see that the effect of increasing competitiveness is greater on the knockout tournaments than on the round-robin – across the competitiveness range $P_{q,R}$ is less widely spread for the round robin so that even if teams are closely competitive the round-robin, as expected, is more discriminating. Using these we can also compare, for example, the original round-robin tournament with a more competitive 1 group-round tournament and a knockout tournament (Fig. 5). Table 3 presents the values of the correlation of the pre-tournament to exit rank and the proportion of wins for the highest (pre-tournament) ranked team ($P_{1/32,5}$). The effect of varying competitive strength can be seen to be quite large, particularly at the top end, that is, on $P_{1/32,5}$.

6. Discussion

This paper has been written with practitioners in mind. We feel that the content is accessible to a wide audience and has the scope to impact on practice. The key elements are a discussion of tournament formats, the definitions of a number of tournament metrics, not necessarily exhaustive but extensive, and a simulation approach that includes a prediction model for individual matches or ties based on input measures that represent the strengths of competitors (and hence competitive balance) in the tournament. Simple techniques can be used to analyse the simulation output. European football is used as the main example. On the basis of this, it is reasonable to conclude that pure knockout tournaments maximise tournament outcome uncertainty and match importance, while round-robins and tournaments with a significant element of group-rounds are the reverse. While knockout tournaments may maximise excitement, they may be unacceptable because half the competitors play only one match (or two if matches are played over two legs). This trade-off is well known and major football tournaments (UEFA Champions League, FIFA World Cup) have settled on the one group-round design.

Fairness in a tournament is an interesting concept and could be defined in a number of ways: a tournament might be considered "fairest" if it maximises the association between competitor strength and finish position; alternatively, a tournament might be considered "fair" if the competition is the same for all teams. On this latter definition we would argue that tournaments with random draws are fair. Under the former definition, seeding which implies non-random draws produces a fairer tournament! By draw here, we mean the allocation of a competitor to a match. Furthermore, seeding methods might also be explored (Schwenk, 2000).

The methodology in this paper offers benefits for tournaments for which there is no settled design. There is scope for real application and an impact on practice in this context. The ICC Cricket World Cup is a case in point, and the design of this tournament presents particular difficulties, not least the significant gap in strength between competing teams. As well as considering standard tournament metrics described, such as the probability that a team in the top q quartile progresses beyond round R, the simulation model might also be used to consider the effect of the weather and scoring systems for separating teams on equal points in group stages.

A limitation of the methodology is the requirement for the quantification of team strengths and for a model of individual match outcomes. These may be difficult to specify and any conclusions based on the simulation will be conditional on such a model. Different sports have different levels of individual match outcome uncertainty and the match outcome model would need to reflect this. However, for many sports such match outcome models are well established. Even then, there may be difficulties if competitor strengths vary with the design. A sensitivity analysis may help. In future work, there is scope for modelling the effect of a design on competitor strengths. Where match outcome models are not available, probabilities may be determined subjectively. Thus the value of the simulation model will be constrained by the quality of the inputs, and some particular circumstances for a tournament may not be envisaged in advance, such as matches that are forfeit.

We have written principally about European sports but the methodology would also be appropriate for North American sports, although the geography of the nations concerned presents tournament organisers with particular issues such as travel. For instance, College baseball in the US uses an interesting tournament design in order that overall team travel is not excessive. Changes to rules and point-scoring systems may also be considered using the methodology described.

Table 3
Sensitivity analysis of tournament metrics (correlation between pre-tournament and exit ranks and proportion of wins for the highest (pre-tournament) ranked team for the original competitor strengths, and more and less competitive values of the strengths)

	Pre-tournament to exit rank correlation			Proportion of wins for highest ranked, $P_{1/32,5}$		
	More comp.	Original	Less comp.	More comp.	Original	Less comp.
Home and away round-robin	0.75	0.85	0.93	0.38	0.79	0.64
Seeded, all group-rounds	0.39	0.51	0.68	0.20	0.41	0.50
Seeded, 2 group-rounds	0.39	0.51	0.68	0.16	0.30	0.39
Seeded, 1 group-round	0.37	0.49	0.67	0.15	0.28	0.37
Seeded, 2 leg knockout	0.26	0.37	0.54	0.13	0.25	0.35

Acknowledgement

The authors would like to thank an anonymous referee whose comments helped to significantly improve the paper.

References

Anderson, I., 1997. Combinatorial Designs and Tournaments. Oxford University Press, UK.
Appleton, D.R., 1995. May the best man win? The Statistician 44, 529–538.
Belton, V., 1999. Multi-criteria decision analysis (MCDA): A structured approach to making choices. Journal of Defence Science 4, 47–55.
Cochran, R.S., 1971. Designs with redeeming social aspects for evenings of social bridge. The American Statistician 25, 12–15.
Crowder, M., Dixon, M., Ledford, A., Robinson, M., 2002. Dynamic modelling and prediction of English Football League matches for betting. The Statistician 52, 157–168.
Dixon, M.J., Coles, S.G., 1997. Modelling association football scores and inefficiencies in the football betting market. Applied Statistics 46, 265–280.
Glenn, W.A., 1960. A comparison of the effectiveness of tournaments. Biometrika 47, 253–262.
Karlis, D., Ntzoufras, I., 2003. Analysis of sports data using bivariate poisson models. The Statistician 52, 381–393.
Knowles, G., Sherony, K., Haupert, M., 1992. The demand for major league baseball: a test of uncertainty of outcome hypothesis. American Economist 36, 72–80.
Koning, R., 2000. Balance in competition in Dutch soccer. Journal of the Royal Statistical Society D 49, 419–431.
Lee, A.J., 1997. Modelling scores in the premier league. Chance 10, 15–19.
Lundh, T. 2006. Which ball is the roundest? – a suggested tournament stability index. Journal of Quantitative Analysis in Sports 2(3), http://www.bepress.com/jqas/vol2/iss3/1.
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., Schutz, R.W., 1997. Efficacy of traditional sport tournament structures. Journal of the Operational Research Society 48, 65–74.
McHale, I., Scarf, P.A., 2007. Modelling soccer matches using bivariate discrete distributions with general dependence structure. Statistica Neerlandica 61, 432–445.
Owen, P., Weatherston, C.R., 2004. Uncertainty of outcome, player quality and attendence at NPC rugby union matches. Journal of Sports Economics 5, 347–370.
Peel, D., Thomas, D., 1997. Handicaps, outcome uncertainty and attendance demand. Applied Economic Letters 4, 567–570.
Rue, H., Salvesen, O., 2000. Prediction and retrospective analysis of soccer matches. The Statistician 49, 399–418.
Russell, K.G., 1980. Balancing carry-over effects in round robin tournaments. Biometrika 67, 127–131.
Scarf, P.A., Shi, X., 2005. Modelling match outcomes and decision support for setting a final innings target in test cricket. IMA Journal Management Mathematics 16, 161–178.
Scarf, P.A., Shi, X., 2008. Measuring the importance of a match in a tournament. Computers and Operations Research 35, 2406–2418.
Schwenk, A.J., 2000. What is the correct way to seed a knockout tournament? The American Mathematical Monthly 107, 140–150.
Seals, D.T., 1963. On the probability of winning with different tournament procedures. Journal of the American Statistical Association 58, 1064–1081.
Sinnott, J. 2006. G14 eyes Champions League changes. BBC Sport, Football. <http://news.bbc.co.uk/sport1/hi/football/europe/4822224.stm> (accessed 9.9.06).
Szymanski, S., 2003. The economic design of sporting contests. Journal of Economic Literature 41, 1137–1187.
Wright, M.B., 2006. Scheduling fixtures for Basketball New Zealand. Computers and OR 33, 1875–1893.