Journal of the Operational Research Society (1997) 48, 65–74 0160-5682/97 $12.00

Efficacy of traditional sport tournament structures

T McGarry and RW Schutz
University of British Columbia

The ranking abilities of some traditional sport tournaments under a variety of initial conditions were analyzed using Monte Carlo procedures. A range of outcome measures were used since a tournament's efficacy will likely depend upon both its objectives and the playing abilities of its contestants. The traditional knockout (KO) is a weak tournament in its ability to rank all players although it requires fewer games than the round robin (RR). The KO tournament's effifacy is notably enhanced, however, in some cases beyond that of the RR tournament if double elimination procedures are used and the seeding is reasonably accurate. Under these conditions, we consider the KO structure to be the best available structure for most tournament purposes. A secondary recommendation of this study is that the fourth and fifth placings be reversed in the traditional KO structures for ranking all players in the eight player situation.

Keywords: probability; simulation; sports; tournaments

Efficacy of traditional sport tournament structures

Researchers in sport and exercise science have a history of interest in sport scoring systems, most notably in their attempts to derive fair and equitable systems for determining contest outcomes. A stochastic processes model for evaluating paired contests was proposed by Schutz[1] and applied to tennis scoring, a methodology that has subsequently been used to investigate the utility of alternative scoring systems in tennis[2], squash[3,4] and volleyball[5,6]. Somewhat surprisingly, virtually all of this work has concentrated on scoring systems for specific sports, and there has been little work in the more general area of sport tournaments. Thus, while we may have an optimal scoring system to determine a game or match outcome between two players in a particular sport, we do not know the most efficient method to select, for example; the best player in the minimum amount of time (contests), the maximum probability that the two best players will meet in a final match, the three medal winners or the best four teams to advance to the next level of competition. The general purpose of this paper then is to extend the work on sport scoring systems and to examine the utility of various tournament structures. To do this, we make use of the extensive work in biometrics, mathematics and statistics in the area of paired comparisons, where paired comparisons refers to a set of binary contests or comparisons with the purpose of rank ordering the top n contests (n can vary from 1 to p, where p is the number of players). We restrict our focus to variations of the two most common tournaments, the Elimination or Knockout (KO) and the Round Robin (RR) and assume that the context in which these tournaments are applied is a sport situation for either individual or team competitions. We acknowledge, but do not address, such interesting and potentially useful systems such as random KO tournaments[7], curtailed RR tournaments and adaptive sequential tournaments[8].

Analytic procedures (mathematical equations yielding exact values) have previously been developed for determining the probability p_{ij} that player i will finish in place j (usually only for the value $j = 1$) for the single KO and RR structures. David[8,9] provided closed equations, applicable for small numbers of players ($t \leqslant 8$), for calculating the probability that player P_1 would win a KO tournament, given a particular probability matrix of game outcomes. Additionally, he used generating functions to produce a full listing of all possible outcomes (but not probabilities) for RR tournaments of size $t = 3$ to $t = 8$[8]. Narayana[7] gives equations for the direct calculation of tournament outcomes for a number of different KO tournaments, but only with respect to the probability of placing first. Searls[10] used a computer to generate all possible draws (315) and all possible outcomes ($2^7 = 128$) for a given draw for a KO tournament of size $t = 8$, and then calculated the exact probabilities for specific probability matrices for each of the 40 320 (315 × 128) possible tournament outcomes. Any of these procedures can be used to compute the probability of a specific player winning a traditional KO tournament, or a slight variation thereof, but none are appropriate for determining the probabilities of finishing second, third, etc., nor do they apply to some of our tournament variations. Furthermore, to attempt to compute the probabilities associated with all possible outcomes in these tournaments would be extremely laborious. Glenn[11] has done it for the 64 (2^6) possible outcomes in a four-player RR, but in our

Correspondence: T McGarry, School of Human Kinetics, University of British Columbia, Vancouver, B.C., Canada V6T 1Z1.

case there are over 268 million (2^{28}) possible outcomes in the eight-player RR. Consequently, we estimated all outcome probabilities and placings through Monte Carlo computer simulation procedures (as others have done for non-traditional tournament structures, for example, David and Andrews[12]). This enabled approximations of the values p_{ij} for all tournament types, thus permitting a tournament organizer to compute the probabilities of primary interest for any specific situation, for example, the probability that the best player will finish in one of the top three positions, the probability that the two best players will meet in the final (not valid in RR), or the probability that a player of rank five or greater will finish in the top three places.

The primary purpose of this study is to ascertain the ranking efficacy of some traditional sport tournament structures for the eight-player situation. A secondary purpose is to develop a more comprehensive, and more valid, consolation procedure for the KO tournament. The consolation component of the KO tournaments developed and presented here is an extension and revision of the standard consolation method in order to rank all eight players. The ranking of all contestants is not an objective of most major championship tournaments, but it is a desirable outcome in qualifying and regional playoffs. It is also of considerable import in those tournaments which relegate to, or exclude from, future sport competition. Thus, if the system proposed here can be shown to be an effective method of accurately ranking all eight players, it will be an attractive alternative to the RR structure which requires a much greater number of matches.

Method

Tournament structures

Five sport tournament structures were analyzed with a draw of eight teams or players (namely $t = 8$; P_1 through P_8). This number of competitors was chosen for reasons of simplicity of analyses and interpretation. The choice of t is not in itself expected to affect a tournament's rank ordering of teams or players (R_1 through R_8) within any particular structure, nor is it expected to affect a tournament's rating efficiency relative to the other tournaments. Increasing the number of competitors in a tournament, however, does have an effect in that the probability of all outcomes are reduced (for example the probability of the best team finishing first, the top two meeting in the final, or the worst team finishing last, all become smaller with increased t, given a similar matrix of playing probabilities). Thus, the findings of this study with $t = 8$ can likely be generalized to larger or smaller tournaments, the differences being only a matter of absolute (and not relative) magnitudes of the probabilities.

Traditional unseeded knockout with consolation (TUKOC). The TUKOC comprises a traditional KO structure which proceeds on an elimination basis and rewards only the contest winner(s) with continued tournament progression until a tournament winner is finally declared. The TUKOC in this study includes consolation contests for unsuccessful players at each stage of the tournament which enables the structure to rank all eight players and allows direct comparison with the rankings of the other tournament structures. The TUKOC tournament is commonly used in situations where there is little or no a priori knowledge about the relative abilities of the competitors, or where random chance is accepted as a component of the tournament. Domestic British and UEFA club soccer tournaments, for example, employ a random draw in the latter stages of KO competititon.

Because the KO is unseeded, P_1 through P_8 are randomly assigned to each run of the simulation. The tournament simulation then progresses according to the assigned playing abilities of each competitor. Player rankings are awarded according to the final placings within the tournament structure. We reversed the fourth and fifth rank placings in the results (see Figure 1) on the basis that the fourth placed player must win twice within the tournament consolation structures whereas the fifth placed player must only win once. We refer to this recommended KO structure as the RUKOC.

Seeded knockout with consolation (SKOC). The SKOC differs from the TUKOC in that the allocation of players to the tournament structure are assigned a priori according to some criterion, usually assumed playing ability. The common seeding allocation is: P_1–P_8, P_5–P_4, P_3–P_6, P_7–P_2, and this seeding protocol was used in this study. An alternate ('optimal') seeding structure is reported by Searls[10] as P_1–P_8, P_6–P_7, P_5–P_4, P_3–P_2, and is included in this study as the SOKOC tournament. The SKOC, except for the consolation component, is the most common tournament

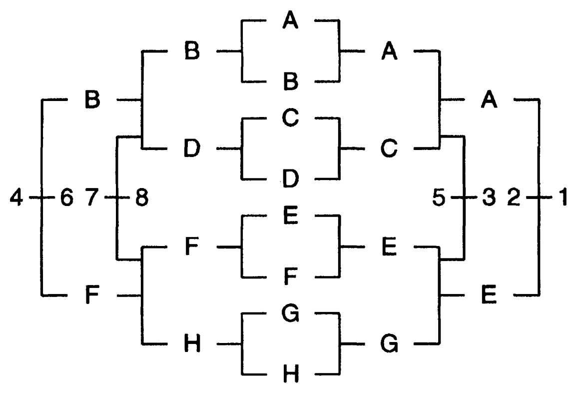

Figure 1 Example of single knockout tournament structure with consolation at each stage of competition (RUKOC).

structure for many championship competitions, for example, tennis, NFL and NHL playoffs.

Investigating the effect of seeding in KO tournaments assumes that the seeding of players is accurate. This study also analyzed the effect of erroneous seeding. The seeding manipulations we undertook were SKOC(A), P_2–P_8, P_5–P_4, P_3–P_6, P_7–P_1 (the best player was seeded second and the second best player seeded first); SKOC(B), P_3–P_8, P_5–P_4, P_2–P_6, P_7–P_1 (the best player was seeded second, the second best player seeded third and the third best player seeded first); and SKOC(C), P_1–P_2, P_3–P_4, P_5–P_6, P_7–P_8 (the worst case seeding, where the best player is matched against the next best player and so on).

Double unseeded knockout with consolation (DUKOC). The DUKOC differs from the TUKOC in that each player must lose twice before elimination from the tournament which offers the chance of recovery from an aberrant result. The loss of one game in the tournament does not therefore preclude that player from winning that tournament, provide he(she) wins all future contests. This study extends the double elimination structure to allow consolation play-offs at each stage of the tournament in order to place all eight players (see Figure 2).

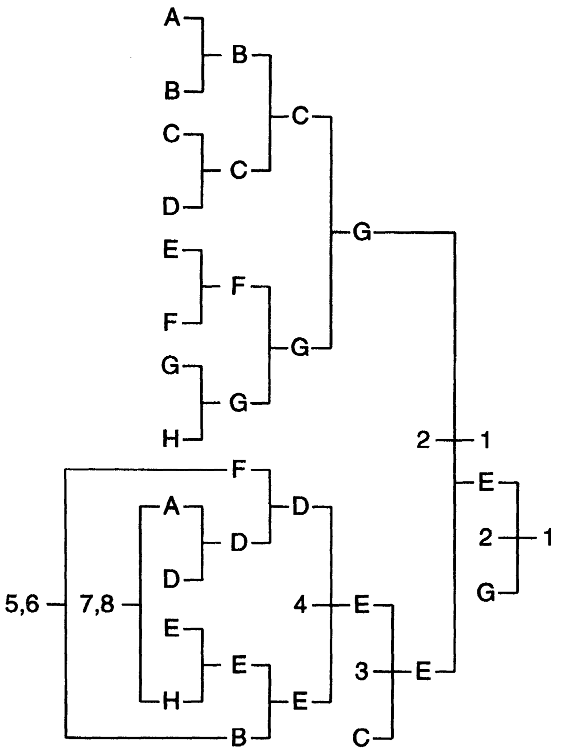

Figure 2 Example of double knockout tournament structure with consolation at each stage of competition (DUKOC).

Double seeded knockout with consolation (DSKOC). The DSKOC uses the same tournament structure as the DUKOC, including the consolation play-offs, and the same a priori seedings as those of the SKOC tournament (namely P_1–P_8, P_5–P_4, P_3–P_6, P_7–P_2).

Round robin (RR_8). The RR_8 structure comprises one group of eight players and each player competes against every other player once. The players are randomly assigned to the RR_8 and sorted after the simulation according to the number of games won to determine the final rankings. Tied outcomes (equal numbers of games won) in athletic tournaments are often resolved through some measure such as goals for and against, but we have no data in this regard. In paired comparison experiments, analytic procedures such as the Bradley–Terry model[13] are common but sport officials have never been willing to utilize mathematical procedures in the context of competition. We therefore resolved game-ties in our simulation procedures with an equal-odds ($p = 0.5$) 'coin toss' for all RR tournaments. The RR tournament structure is representative of all league based sport competitions.

Combination round robin with sift (RR_{4S}). The combination RR_{4S} consists of two RRs each comprising four players. The best two ranked players from each RR_4 form a further RR_4 (championship RR) and likewise for the third and fourth place ranks (consolation RR) (see Figure 3). The subsequent rankings of the championship and consolation RRs determine the final player rankings. The championship RR_4 ranks the players first through fourth and the consolation RR_4 ranks the players fifth through eighth. This

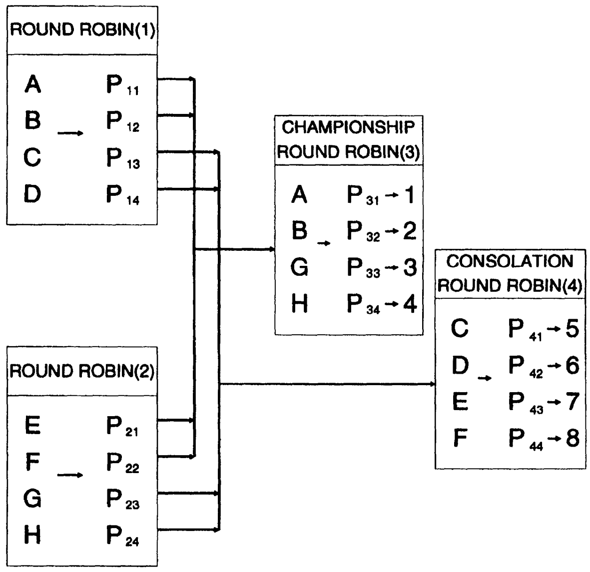

Figure 3 Example of combination round robin with sift (RR_{4S}).

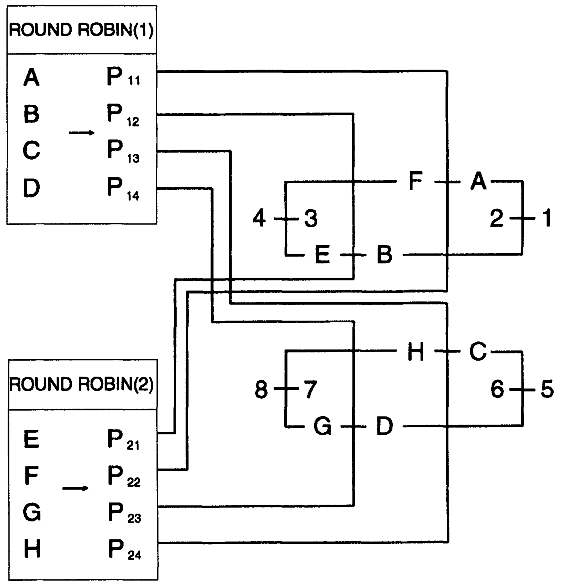

Figure 4 Example of combination round robin and seeded knockout (RR_{4KO}).

tournament structure is not commonly used in sport, but is presented here as a possible alternative to the RR_8 because it is more efficient in that it requires fewer games.

Combination round robin with seeded knockout (RR_{4KO}). The combination RR_{4KO} also consists of two RR_4s which rank the players before advancing to a seeded knockout (see Figure 4). The first ranked player from the first RR_4 meets the second ranked player from the second RR_4, and the first ranked player from the second RR_4 meets the second ranked player from the first RR_4. A similar seeding exists for the respective third and fourth RR_4 placings. The tournament then advances in a KO structure and players are then further ranked in the traditional manner. The winning and losing finalists are ranked first and second; the winning and losing consolation finalists ranked third and fourth and so on through eighth placing. The World Cup Soccer competition, for example, employs a similar tournament structure.

Simulation procedures

The probability with which any player would beat any other player was fixed a priori (see Table 1) to represent the following situations: A linear ordering of playing strengths such that adjacent seeds are closely matched and the probabilities of defeating weaker players increases linearly (0.05 per seeded placing) from best to worst (Matrix I); a situation in which P_1 (the best player) is slightly better than P_2 (the next best player), and both P_1 and P_2 are considerably stronger than P_3 through P_8, who are all of equal ability (Matrix II); a truncated linear ordering similar to Matrix I except that the bottom three players are very weak relative to the top three players (Matrix III); and a linear ordering identical to (Matrix I) except that P_3 is stronger than P_1, thus violating the principle of transitivity (Matrix IV). The playing matrices read row by column to represent the a priori probability that any player will beat any other player. the matrices are complementary of course and the a priori probability of losing to any player is thus represented column by row. For example, the probability that P_1 beats P_2 from Matrix I is 0.55 (Table 1). Note that the TUKOC data show that violating transitivity (Matrix IV) did not adversely affect the a priori player rankings, P_1 through P_8, as indicated from the average of each row.

Each simulation proceeds on a probabilistic basis according to the relative playing abilities and the particular tournament. Ties between players were not permitted for any contest outcome. This is not to be confused with game-ties which commonly arise in the RR procedures. (Table 2 shows the number of times that game-ties were resolved in the RR simulations.) A simulation was run ten thousand times ($N = 10000$) using different starting seeds from the random number stream and the frequency for each player finishing first through last place (R_1 through R_8) was recorded. The validity of the simulation procedures was tested in two ways. First, a matrix representing equality among all players ($p = 0.50$ for all combinations) was used for all the tournament structures. The outcomes indicated that the Monte Carlo procedures were unbiased in that, as expected, all players won an approximately equal number of times and placed first through eighth equally often. Second, analytic procedures were used to calculate from Matrix II the probability of P_1 and P_2 placing first, placing second, and meeting in the final, for both the TUKOC and SKOC tournaments. The analytic results are in close agreement (0.004) with the simulated results for both the TUKOC and SKOC tournaments (Table 2) and lie within the standard error ($\leqslant 0.005$) expected from ten thousand simulation runs. It was concluded from these data that the simulation procedures were valid.

Assumption of initial a priori playing abilities

We assume stationarity and independence of the playing matrices such that the a priori playing abilities for player P_i to beat player P_j does not change throughout the tournament and is independent of previous contests. In practice of course these probabilities are unknown and dynamic. The existence, or otherwise, of transient probabilities and their degree of divergence from the assumed conditions of stationarity and independence remains an unexplored area of enquiry, although Pollard[2] did conclude, from an analysis of tennis scoring, that the assumptions could not be rejecteed on statistical grounds and thus are appropriate for modelling the practical situation. Thus, while changing probabilities would be expected to alter the outcome of

Table 1 Playing Matrix I, Matrix II, Matrix III and Matrix IV denoting differential playing abilities between players P_1 through P_8[a]

	Matrix I									Matrix II								
Player	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	*Avg*	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	*Avg*
P_1	—	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.70	—	0.60	0.80	0.80	0.80	0.80	0.80	0.80	0.77
P_2	0.45	—	0.55	0.60	0.65	0.70	0.75	0.80	0.64	0.40	—	0.75	0.75	0.75	0.75	0.75	0.75	0.70
P_3	0.40	0.45	—	0.55	0.60	0.65	0.70	0.75	0.59	0.20	0.25	—	0.50	0.50	0.50	0.50	0.50	0.42
P_4	0.35	0.40	0.45	—	0.55	0.60	0.65	0.70	0.53	0.20	0.25	0.50	—	0.50	0.50	0.50	0.50	0.42
P_5	0.30	0.35	0.40	0.45	—	0.55	0.60	0.65	0.47	0.20	0.25	0.50	0.50	—	0.50	0.50	0.50	0.42
P_6	0.25	0.30	0.35	0.40	0.45	—	0.55	0.60	0.41	0.20	0.25	0.50	0.50	0.50	—	0.50	0.50	0.42
P_7	0.20	0.25	0.30	0.35	0.40	0.45	—	0.55	0.36	0.20	0.25	0.50	0.50	0.50	0.50	—	0.50	0.42
P_8	0.15	0.20	0.25	0.30	0.35	0.40	0.45	—	0.30	0.20	0.25	0.50	0.50	0.50	0.50	0.50	—	0.42

	Matrix III									Matrix IV								
Player	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	*Avg*	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	*Avg*
P_1	—	0.60	0.70	0.80	0.90	0.95	0.95	0.95	0.84	—	0.55	0.40	0.65	0.70	0.75	0.80	0.85	0.67
P_2	0.40	—	0.60	0.70	0.80	0.90	0.95	0.95	0.76	0.45	—	0.55	0.60	0.65	0.70	0.75	0.80	0.64
P_3	0.30	0.40	—	0.60	0.70	0.80	0.90	0.95	0.66	0.60	0.45	—	0.55	0.60	0.65	0.70	0.75	0.61
P_4	0.20	0.30	0.40	—	0.60	0.70	0.80	0.90	0.56	0.35	0.40	0.45	—	0.55	0.60	0.65	0.70	0.53
P_5	0.10	0.20	0.30	0.40	—	0.60	0.70	0.80	0.44	0.30	0.35	0.40	0.45	—	0.55	0.60	0.65	0.47
P_6	0.05	0.10	0.20	0.30	0.40	—	0.50	0.50	0.29	0.25	0.30	0.35	0.40	0.45	—	0.55	0.60	0.41
P_7	0.05	0.05	0.10	0.20	0.30	0.50	—	0.50	0.24	0.20	0.25	0.30	0.35	0.40	0.45	—	0.55	0.36
P_8	0.05	0.05	0.05	0.10	0.20	0.50	0.50	—	0.21	0.15	0.20	0.25	0.30	0.35	0.40	0.45	—	0.30

[a] Win probability reads row by column, for example, the probability that P_1 beats P_2 from Matrix I is 0.55.

the tournament on any single occasion, it is probably not unreasonable to expect a good approximation of tournament function from stationary probabilities over the long run average.

Results and discussion

Accuracy of placing player i in position i

We suggest that the most valid tournament is that structure which best ranks the players according to their a priori standings. While the a priori rankings are unknown in practice, they are known in the analysis since they are used to determine the initial modelling conditions. The kappa coefficient (κ) is a suitable measure of validity and reflects the degree of agreement between two observations. In this instance the observations are the a priori player rankings (based on the playing matrices) and the subsequent tournament placings (produced by simulation procedure). κ is expresseed as the ratio of observed agreement to maximum possible agreement after accounting for chance in both the numerator and the denominator. Thus, it is provided in the formula $\kappa = (p_0 - p_c)/(1 - p_c)$, where p_0 is the proportion of observed agreement and p_c is the proportion of chance agreement, and reflects the parity of agreement in the cross diagonal cells of a two-way (for example ranking by placing) observation matrix[14]. In fact, we used a weighted kappa (κ_w) which accounts for the degree to which the two observations differ, that is, the disagreements in the off diagonal cells that locate further from the cross diagonal are weighted disproportionately more than those that locate closer to it. A κ_w of 1.0 would indicate that the best player finished first in all ten thousand simulations, the next best player second in all ten thousand simulations, and so on. Since κ_w in this analysis is dependent upon the playing matrix, however, it can only be used as a comparative measure of tournament validity. The κ_w values reported in Table 2 are low, relative to values obtained in most observational research studies, although it was not expected that they would be much above 0.10 due to the low accuracy of the contests. For example, the probability that the best player beats the second best is only 0.55.

κ_w is applicable only to that situation which comprises an ordinal progression of playing ability, as evidenced in Matrix I (Table 1) and is inappropriate for those situations which contain two or more players of equal ability. The validity of a tournament, as inferred from κ_w is highest for the DSKOC, the RR_8 and the SKOC, assuming accurate seeding (Table 2). It is not surprising to find that the RUKOC and the SOKOC most poorly rank all eight players. Indeed, promotion of the best player through favourable seeding (SOKOC) is at the expense of the other player rankings otherwise accorded by initial chance allocation (RUKOC, Table 2).

An alternative analysis of the validity of a tournament might be the criterion of maximizing the expected value of the quality of the winning team. Monahan and Berger[15], for example, used the points awarded over a season of league based hockey competition as a rating index for each team's entry into the Stanley Cup play-offs (eight team knockout tournament where each knockout stage comprises the first

Table 2 Tournament summary statistics for Matrix I, Matrix II, Matrix III and Matrix IV. Theoretical values are reported in parentheses

	Matrix I							
	RUKOC	SKOC	SOKOC	DUKOC	DSKOC	RR_{4KO}	RR_{4S}	RR_8
No. Games	12	12	12	16	16	20	24	28
kappa (κ_w)	0.078	0.113	0.075	0.096	0.121	0.089	0.095	0.119
W_{MAX}	0.597	0.611	0.606	0.615	0.625	0.602	0.604	0.613
P_1–R_1	0.313	0.352	0.403	0.366	0.398	0.318	0.331	0.356
P_1–$R_{1,2}$	0.477	0.575	0.663	0.566	0.611	0.506	0.558	0.586
P_1–$R_{1,2,3}$	0.633	0.756	0.782	0.688	0.743	0.679	0.699	0.751
P_2–R_1	0.221	0.242	0.189	0.241	0.252	0.237	0.231	0.238
P_2–$R_{1,2}$	0.385	0.453	0.341	0.443	0.476	0.424	0.441	0.468
P_2–$R_{3,2,3}$	0.552	0.636	0.486	0.595	0.634	0.596	0.592	0.642
$Final_{P1-P2}$	0.128	0.258	0.222	0.289	0.369	0.158	—	—
Ties (10 000)						7866	9568	9981

	Matrix II							
	RUKOC	SKOC	SOKOC	DUKOC	DSKOC	RR_{4KO}	RR_{4S}	RR_8
W_{MAX}	0.663	0.673	0.673	0.702	0.709	0.678	0.673	0.701
P_1–R_1	0.428 (0.425)	0.441 (0.440)	0.441	0.527	0.537	0.464	0.450	0.508
P_1–$R_{1,2}$	0.579 (0.583)	0.644 (0.640)	0.644	0.720	0.737	0.645	0.694	0.765
P_1–$R_{1,2,3}$	0.722	0.762	0.762	0.807	0.827	0.799	0.803	0.873
P_2–R_1	0.279 (0.276)	0.298 (0.295)	0.298	0.292	0.302	0.287	0.286	0.309
P_2–$R_{1,2}$	0.467 (0.464)	0.565 (0.561)	0.565	0.564	0.613	0.519	0.555	0.608
P_2–$R_{3,2,3}$	0.642	0.700	0.700	0.695	0.745	0.699	0.689	0.776
$Final_{P1-P2}$	0.206 (0.206)	0.364 (0.360)	0.364	0.417	0.485	0.285	- -	—
Ties (10 000)						7743	9547	9987

	Matrix III							
	RUKOC	SKOC	SOKOC	DUKOC	DSKOC	RR_{4KO}	RR_{4S}	RR_8
W_{MAX}	0.759	0.777	0.780	0.782	0.790	0.768	0.764	0.776
P_1–R_1	0.471	0.521	0.630	0.557	0.587	0.486	0.477	0.517
P_1–$R_{1,2}$	0.650	0.807	0.905	0.779	0.826	0.712	0.748	0.792
P_1–$R_{1,2,3}$	0.805	0.914	0.940	0.882	0.927	0.873	0.871	0.926
P_2–R_1	0.271	0.289	0.203	0.266	0.277	0.281	0.283	0.284
P_2–$R_{1,2}$	0.498	0.626	0.446	0.581	0.635	0.557	0.577	0.613
P_2–$R_{3,2,3}$	0.704	0.851	0.586	0.760	0.831	0.768	0.754	0.830
$Final_{P1-P2}$	0.253	0.503	0.402	0.542	0.650	0.342	—	—
Ties (10 000)						5672	8603	9728

	Matrix IV							
	RUKOC	SKOC	SOKOC	DUKOC	DSKOC	RR_{4KO}	RR_{4S}	RR_8
W_{MAX}	0.590	0.603	0.596	0.603	0.613	0.594	0.594	0.605
P_1–R_1	0.265	0.310	0.370	0.291	0.326	0.267	0.261	0.297
P_1–$R_{1,2}$	0.434	0.575	0.663	0.497	0.569	0.462	0.478	0.521
P_1–$R_{1,2,3}$	0.597	0.738	0.774	0.635	0.719	0.640	0.627	0.700
P_2–R_1	0.222	0.242	0.189	0.245	0.255	0.234	0.239	0.251
P_2–$R_{1,2}$	0.385	0.453	0.341	0.444	0.477	0.418	0.442	0.474
P_2–$R_{3,2,3}$	0.553	0.636	0.486	0.594	0.635	0.589	0.591	0.648
$Final_{P1-P2}$	0.113	0.258	0.222	0.261	0.349	0.135	—	—
Ties (10 000)						7871	9583	9980

team to win four games). The criterion of maximum expected value as an indicant of the best tournament in this case is to maximize the expected number of points of the cup winner by maximizing $W_{MAX} = \sum P$ $\{i \text{ wins cup}\}\{i \text{ point's rating}\}$ where $i = 1$–8. We extend the same analysis to this study but use a different index, namely the a priori probability of beating any other player (averaged across row). The optimal tournament according to this criterion is thus the maximization of the cross product of the a priori and winning probabilities. Table 2 shows the KO tournaments to generally outperform the RR tournaments. The efficacy of the KO tournaments is improved with both seeding and the introduction of the double elimination procedure although the respective impact of each, as measured by W_{MAX}, is dependent on the initial playing matrix. Taking both κ_w and the W_{MAX} into account, we suggest that the DSKOC tournament is the best tournament to accurately discriminate between players of different playing abilities. The SKOC well approximates the RR_8 tournament and would seem an expedient alternative given the much fewer number of games required. The extra four (or five) games required for the double KO tournament (DUKOC, DSKOC), however, would seem to yield a good dividend as compared to its single KO counterparts (TUKOC, SKOC) and is probably a worthwhile addition to the tournament structure in most cases. It is noteworthy that the DUKOC is better than the combination RRs for each initial playing condition while requiring fewer games and no prior information and better than the SKOC for some measures of tournament performance for some playing matrices.

Placing the top players in the top positions

The DUKOC and DSKOC typically outperform the RR_8 in ranking the best player in first place (P_1–R_1) but perform less well in ranking that same player in the top two (P_1–$R_{1,2}$) and three (P_1–$R_{1,2,3}$) places. In contrast the SOKOC better promotes the stronger player for Matrix I, III and IV but not for Matrix II (P_1–R_1; P_1–$R_{1,2}$; P_1–$R_{1,2,3}$). This finding is expected since the SOKOC is favoured when the playing abilities of the entrants are more discriminatory. With a more homogeneous group of contestants (Matrix II) the extra number of games in the RR_8 is required to accurately ascertain the final placings. Interestingly, the SKOC and SOKOC are less sensitive to the particular violation of transitivity than the other tournament structures, as indicated by the respective P_1–R_1 placings for Matrix I and Matrix IV. This is because P_1 will most likely contest P_3 in the RR and double KO structures, but this is less likely the case for the SKOC and SOKOC tournaments which seed P_1 and P_3 in different halves of the draw. The combination RRs appear to have little to recommend them if the selection criterion is determination of the best player. These findings are reasonably stable for placing the best player in the top one (P_1–R_1), two (P_1–R_2) or three (P_1–R_3) places. Note that for an heterogeneous set of contestants (Matrix III) virtually any tournament structure provides a high probability ($0.805 \leqslant p < 0.940$) that the best player will finish in the top three (P_1–$R_{1,2,3}$), whereas for more equally matched contestants (Matrix I) the probability that the best player finishes in the medals is, of course, never as high ($0.633 \leqslant p \leqslant 0.782$).

The placing of P_2 in the tournament is also contingent upon the playing matrix. The DSKOC produces the highest probability of placing P_2 first, not a wholly desirable feature, for Matrix I and IV, the RR_8 for Matrix II and the SKOC for Matrix III. The SOKOC produces the lowest probability of P_2 placing first for Matrix I and IV, given its bias to the best player. The abilities of tournaments to place P_2 in the top one (P_2–R_1), two (P_2–$R_{1,2}$), or three (P_2–$R_{1,2,3}$) places generally rank RR_8, DSKOC, SKOC and then the combination of RRs.

A tournament may legitimately have a different purpose to that of correctly ranking the entire list of entrants. A frequent aim is to produce an exciting final contest between the two best players. Optimizing the probability of meeting in the final stage of competition (Final_{P1-P2}) is thus an important consideration of many tournaments. While the relative merits of various structures depend upon the particular playing matrix, the double KO structures produce notably higher probabilities. Of particular note is that the DUKOC is better than seeded KO structures. Thus, if the tournament organizer wishes to maximize the likelihood of the two best players meeting in a final contest then a double elimination KO structure is recommended. Seeding according to prior information would further enhance the prospects of producing such a contest. The TUKOC and RR_{4KO}, which incidentally constitute many soccer tournaments, perform much worse in all circumstances and should, in our view, only be used if chance is recognized as an acceptable facet in determining the tournament outcome.

Given these findings, coupled with the measures of tournament validity and the number of contests required, it is concluded tthat the double KO and or seeded KO tournaments offer the best structures for most competition purposes. Note though that this recommendation assumes reasonably accurate seeding of the players and inaccurate seeding would perform much worse, as indicated by the RUKOC. An additional advantage of the KO structures is that the consolation contests can be discarded if the tournament victor(s) are the sole criterion of interest, thereby shortening the tournament length. A noteworthy feature of this study is that the competition objectives and the expected playing abilities of the contestants are important considerations in selecting a particular tournament.

Ranking all players with the knockout tournament

Traditionally, the KO claims only to identify the best player and placing the remaining players is of little or no consequence. The SKOC can reliably rank the best two players, and, placings two through five can additionally be identified by adding a consolation component. The usual consolation round, however, consists only of a loser's bracket in which the first round losers play a KO to determine a consolation winner, and sometimes a consolation final contested by the losers of the two semi-finals. These procedures do not permit the ranking of all players which may be desirable in some contests. The consolation structure incorporated into the TUKOC tournament is unique, to our knowledge, and results in the placement of all eight players (see Figure 1).

The results from Matrix I (see Figure 5) show that the derived ranking procedure is a valid method. The RR_8 is presented for comparative purposes, as it is generally accepted as the gold standard for ranking all contestants in a tournament. Figure 5 shows that the RUKOC is a better model of the RR_8 than the TUKOC (which assigns fifth place to the winner of the consolation and fourth to the loser of the winner's side), because the TUKOC often results in P_4 finishing in fifth place, or worse. For the TUKOC the probability that player i finishes in position i or higher is greater than that for any lower positioned player. For example, exact values for the RUKOC (Matrix I), showed that P_4 has a higher probability of placing first (0.122), second (0.138), third (0.143) or fourth (0.146) than do P_5–P_8. P_5 showed a higher probability of placing first (0.076), second (0.129), third (0.127) or fourth (0.127) than P_6–P_8, although P_6–P_8 have a greater chance of placing fifth (0.159, 0.162, 0.149 respectively *vs.* 0.141), P_5 has a higher cumulative probability of placing fifth or higher (0.600) than P_6 (0.522), P_7 (0.451) or P_8 (0.362). These findings validate our ranking system (Figure 1) which places the winner of the consolation in fourth place and the loser of the winner's bracket in fifth position, the reverse of what is often done. Thus we conclude that the derived ranking system is valid for the eight- player KO, and, requiring only twelve games in comparison to the

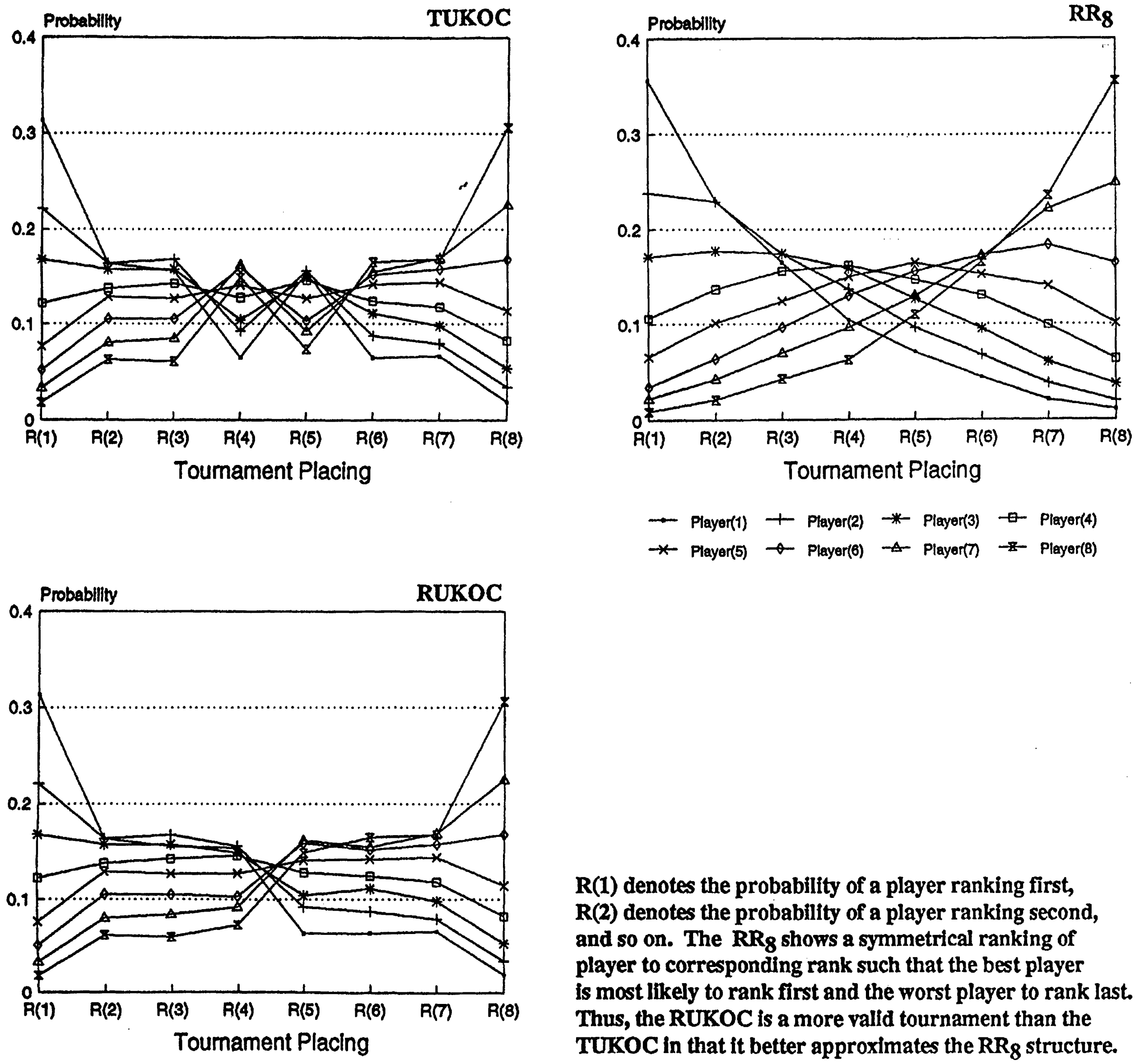

Figure 5 Ranking efficacy of the TUKOC and RUKOC tournaments.

twenty eight for the RR_8, is a much more efficient method for placing all eight players.

An overview of sport tournament structures

We have detailed an analysis of the KO (both single and double elimination) and the RR structures on the basis that they are often used in some form, or some combination or variation thereof, in many of the major sports events. Notwithstanding, there are many other tournament formats that are used by a variety of sports organizations that extend beyond the scope of the present analysis. We now discuss some alternative tournament structures while acknowledging at the outset that our selected classification system is not definitive, or even the only classification system, and that our consideration of some of the different tournament structures is by no means exhaustive.

There are many individual performance sports in which the contestants do not compete directly against each other but rather against some external performance criterion. Points, or some other equally suitable measure, are usually assigned in grading performance according to some pre-established criterion and the individual with the most (or sometimes the least) points is declared the winner. Golf, archery and the decathlon are common examples of this type of tournament. Likewise, gymnastics, figure skating, diving and what have come to be referred to as 'extreme' sports assign some measure to sports performance though the scoring procedures in these cases are primarily subjective. Scording adjustments, such as the discarding of the highest and lowest judges' ratings, are often invoked in an attempt to improve the reliability of these scoring systems. This type of tournament classification also extends to sports that involve some form of time trial where performance is measured by the timepiece. Here, the fastest (namely shortest) time yields the best performance, as observed in competitive sports like Alpine (downhill) skiing and also some stages of the Tour de France bicycle race. An advantage of the type of tournament structure used in individual performance sports is that it can accommodate a large number of competitors in a single event, each of whom can be rank ordered in a relatively short period of time. This is because the entrants compete solo rather than head to head which frequently affords more efficient scheduling of sports events and, also, often reduces the number of performances required to establish a winner. Only one (multiple) comparison, or contest, is minimally required to determine a winner in individual performance sports whereas a minimum $N-1$ and $\log N/\log 2$ (paired) comparisons are required in a RR and a single elimination KO respectively. Individual performance sports are not limited to single contests, however, and multiple contests, or stages, are often employed in order to derive a composite performance score, thereby increasing both the reliability and validity of the final result. A limitation of this type of tournament structure is that it necessarily lends itself to those individual sports in which performance standards are comparable across contestants and is thus not sensitive to varying external conditions.

The above comments with regard to individual performance sports do not exclude the possibility of an elimination structure, nor do they preclude the deployment of head to head confrontations. Olympic track sports events, for example, use some kind of elimination procedure in head to head competition, with a second change given to the some of the eliminated athletes. In these tournaments, the contestants form heats of typically four or eight and the top placed individuals, or teams, (usually the top two or three) from each heat automatically advance to the next heat. (Note that this structure can be likened to the RR_{4S} tournament except that each entrant is now ranked as a result of one comparison per heat rather than $N-1$ comparisons per RR.) The remaining available places are then selectively assigned to the near qualifying finishers from the best performance scores from all the heats combined. In sports such as rowing, where performance comparisons may not be suitable because of changing conditions, the near qualifying finishers participate in a repechage, or loser's heat, with the winner of that heat advancing to the next heat to join the earlier placed qualifiers. The repechage increases the validity of the tournament structure in determining the final placings since it compensates for the possibility that the better contestants are matched against each other by random draw in an earlier heat. While the repechage is not necessary in seeded tournament structures (providing that the seeding is accurate), it nevertheless insures somewhat against an early upset, for whatever reason, and is therefore an attractive feature for tournament organizers.

The Swiss tournament is a tournament structure that is commonly used in sports like chess and bridge. Successive rounds are played without elimination but the pairings in each round are determined by the outcomes of the previous rounds. Thus, the better players tend to compete against each other as the tournament progresses. The final scoring is based on the total number of wins that a player has attained but, importantly, the value of a win is a function of the total number of wins by one's opponent. In other words, a win over a better player is more highly valued than a win over a lesser player. Thus, the tournament structure is expected to be highly valid since it not only accounts for the present comparison (contest) but it also accounts for the expected value of that comparison. Weak tournament rankings, which accommodate for ties, and other ranking procedures that value wins over a strong player more highly than wins over a weak player have been proposed[16] and are also expected to yield a more valid composite score. The complexity of the analysis, however, can be a drawback of such tournament designs since the outcomes are often not well understood by the competitors and spectators alike.

It is evident that many tournament structures exist to determine sports champions and that different structures have different ranking capacities depending on, amongst other factors, the initial starting conditions. First, the nature of the sport is a limiting factor with regard to whether paired or multiple comparisons can be used. Second, the tournament's objectives, be it to identify the best player, to produce an exciting final contest or to maximize spectator appeal is a worthy consideration, as are the number of games, or measures, that a tournament requires. Third, the equity of a tournament structure should be such that comparable teams have a comparable chance of success. The 1986, 1990 and 1994 World Cup Soccer Finals provide an interesting commentary in this regard since it does not equally promote the chances of comparable teams, at least under certain initial conditions. This is because six RRs, of four teams each, compete before the two best placed teams from each RR, as well as the four (from six) best third placed teams, advance to a seeded KO. If we assume that second placed teams are generally stronger than the third placed teams, then the tournament structure must favour some teams at the expense of other comparable teams depending on their a priori allocation to a RR. The recent change in the 1998 World Cup Soccer Finals tournament to eight RRs of four teams each with the best two placed teams from each RR advancing to a seeded KO promises to be a more equitable tournament structure.

Conclusions

The results of our analyses clearly show that no single best tournament structure exists. In general, it is that structure which best fits the competition's objectives with regard to the initial playing abilities of the players. The RR_8 is the most accurate tournament in ranking all the competitors, but it does so at a high cost in that it requires over twice as many games as a single KO tournament. The introduction of both a double elimination to the KO structure and a priori seeding, assuming the seeding is reasonably accurate, notably enhances tournament performance, in some cases beyond that of the RR_8. For example, the RR_8 is not as accurate as the SKOC in optimizing the best player's chance of finishing first, second, or third for some combinations of relative playing abilities. We conclude that the KO structure is probably the most suitable tournament structure in most cases, given its ranking ability of all players, its promotion of the stronger players and the relatively few games required. This finding, however, is conditional on the use of a double elimination structure and or reasonably accurate seeding of all eight players. A random KO tournament (for example RUKOC) by contrast performs much worse in ranking players of different abilities, although this result does not extend to the DUKOC. The inclusion of a double elimination structure would thus seem to be a worthwhile addition in any KO tournament.

References

1 Schutz RW (1970). A mathematical model for evaluating scoring systems with specific reference to tennis *Res Q* **41**: 552–561.
2 Pollard GH (1987). A new tennis scoring system *Res Q Exerc Sport* **58**: 229–233.
3 Schutz RW and Kinsey WJ (1979). Comparison of North American and International squash scoring systems—a computer simulation *Res Q* **48**: 248–251.
4 Clarke SR and Norman JM (1979). Comparison of North American and International squash scoring systems—analytic results *Res Q* **50**: 723–728.
5 Pfeifer PE and Deutsch SJ (1981). A probabilistic model for evaluation of volleyball scoring systems *Res Q Exerc Sport* **52**: 330–338.
6 Fellingham GW, Collings BJ and McGown CM (1994). Developing an optimal scoring system with a special emphasis on volleyball *Res Q Exerc Sport* **65**: 237–243..
7 Narayana TV (1979). *Lattice Path Combinatorics with Statistical Applications*. University Toronto Press: Toronto.
8 David HA (1988). *The Method of Paired Comparisons*. Oxford Press: New York.
9 David HA (1959). Tournaments and paired comparisons *BIOKA* **46**: 139–149.
10 Searls DT (1963). On the probability of winning with different tournament procedures *J Amer Stat Assoc* **58**: 1064–1081.
11 Glenn WA (1960). A comparison of the effectiveness of tournaments *BIOKA* **47**: 253–262.
12 David HA and Andrews DM (1987). Closed adaptive sequential paired-comparison selection procedures *J Statist Comp and Simulation* **27**: 127–141.
13 Bradley RA and Terry ME (1952). The rank analysis of incomplete block designs. I. The method of paired comparisons *BIOKA* **39**: 324–345.
14 Fleiss JL (1975). Measuring agreement between two judges on the presence or absence of a trait *BIOMA* **31**: 651–659.
15 Monahan JP and Berger PD (1977). Playoff structures in the national hockey league. In: Ladany SP and Machol RE (eds). *Optimal Strategies in Sports 1977*. North Holland, Amsterdam, pp 123–128.
16 Cook WD and Kress M (1990). An *m*th generation model for weak ranging of players in a tournament *J Opl Res Soc* **41**: 1111–1119.

Received August 1995
accepted July 1996 after two revisions.