Journal of Sports Research
2015 Vol.2, No.3, pp. 62-76
ISSN(e): 2410-6534
ISSN(p): 2413-8436
DOI: 10.18488/journal.90/2015.2.3/90.1.62.76

MEASURING THE RELATIVE EFFICIENCY OF TOURNAMENTS

Graham Pollard[1] † --- Denny Meyer [2] --- Geoff Pollard[3]

[1]*Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, Australia*

[2,3]*Faculty of Health, Arts and Design, Swinburne University of Technology, Melbourne, Australia*

ABSTRACT

The efficiency of various tournament structures such as knock-outs and round robins has been considered by several authors, and it has been noted that the round robin structure has a higher probability that the best player wins than do the other structures, but it is at the cost of playing a larger number of matches. On the other hand the knock-out tournament requires relatively few matches, but has a smaller value for the probability that the best player wins. Thus, the compromise between the number of matches played in a tournament and the probability that the best player wins that tournament has remained to some extent an unsolved problem. The Masters Tennis tournament, being a combination of two round robins and a knock-out, is one attempt at such a compromise. In this paper the balance between the number of matches played and the probability that the best player wins is addressed, and the efficiency of several different tournament structures is evaluated. It is noted that although the efficiency has been evaluated for just some of the commonly used tournament structures, the approach outlined in this paper can be used for a wide range of other structures.

Keywords: Knock-out, Round robin, Partial round robin and draw and process tournaments, Efficiency of tournament structures.

Received: 16 December 2014/ **Revised:** 20 January 2015/ **Accepted:** 6 February 2015/ **Published:** 10 February 2015

Contribution/ Originality

This paper formulates a method for finding the relative efficiency of tournaments. The knock-out structure is shown to be a very efficient one, and the draw and process quite efficient.

1. INTRODUCTION

Several authors have considered the efficiency of various tournament structures. For example, McGarry and Schutz (1997) noted in their introduction that '…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

† Corresponding author

in the minimum amount of time (contests), the maximum probability that the best two players will meet in a final match, the three medal winners or the best four teams to advance to the next level of competition.' They concluded that 'no single best tournament structure exists', and that a particular round robin structure they studied '…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 knock-out tournament'. Further they concluded that 'the knock-out 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'. Appleton (1995) reported: 'Simulations of various kinds of sporting tournament have been carried out to assess their relative ability to produce as winner the best of the entrants. A strong contender when it is necessary to play relatively few games is the seeded draw and process.' In this paper a balance between the number of matches played and the probability of a favourable tournament outcome is achieved by using appropriate efficiency measures. Using a previously developed efficiency measures which is explained below, simulation analyses allow the comparison of the efficiency of several different tournament designs. Further, a method for determining the absolute efficiency rather than the relative efficiency of tournament structures is outlined.

Miles (1984) developed a very elegant theory for the relative efficiency of different match scoring systems, assuming that points were independent. He considered 'win-by-n' (Wn) scoring systems in which the winner is the first player to win n more points than his opponent. Assuming points are independent and player A has a constant probability p of winning every point (which Miles called 'unipoints'), it can be shown that the probability P that player A wins by n points (Wn), the expected number of points played μ, and n satisfy the '(P, μ, n) equation'

$$(P-Q)/\mu=(p-q)/n$$

with the ratio P/Q given by

$$(P/Q)=(p/q)^n$$

where $Q = 1 - P$ and $q = 1 - p$.

Pollard (1992) showed that this Wn scoring system has the constant probability ratio property. That is, the ratio of the probability that player A wins in n + 2m points divided by the probability that he loses in n + 2m points (m = 0, 1, 2, …) is constant, and is equal to P/Q. Wald and Wolfowitz (1948) had earlier showed that the *optimally efficient* statistical test for testing whether a Bernoulli probability was greater than 0.5 or less than 0.5 was in fact Wn for the case in which the probability of type 1 error was equal to the probability of type 2 error, and this is required for such a scoring system to be fair. Thus, it was natural for Miles to give this optimally efficient family of scoring systems Wn an efficiency of unity. He showed that the efficiency ρ of a general 'unipoints' scoring system SS with key characteristics P and μ is given by

$$\rho=\frac{(P-Q)\ln(P/Q)}{\mu(p-q)\ln(p/q)} \qquad (1)$$

This efficiency measure of a general 'unipoints' scoring system SS, as described by Miles (1984), is defined as the expected duration of the 'interpolated' Wn system with the same P-value as SS (as derived from the above '(P, μ, n) equation') divided by the expected duration of SS, namely μ. Note that the value of n for this 'interpolated' Wn system is given by

$n = \ln(P/Q)/\ln(p/q)$,

resulting directly from the constant probability ratio property of Wn. (Note that whilst Wn systems exist for integer values of n only, it is possible to find the (possibly non-integer) value of n for any valid P value.) It follows, by ignoring the factors involving p (and q) in (1) above, that the expression

$$((P-Q)/\mu)\ln(P/Q)$$

is the measure for the relative efficiency of a unipoints scoring system given underlying independent points. Thus, based on (1), the efficiency of the unipoints scoring system 1 *relative to* the unipoints scoring system 2 for any probability p, is given by

$$\frac{((P_1-Q_1)/\mu_1)\ln(P_1/Q_1)}{((P_2-Q_2)/\mu_2)\ln(P_2/Q_2)} \qquad (2)$$

where the subscript 1 refers to scoring system 1 and the subscript 2 refers to scoring system 2.

Miles (1984) also considered scoring systems relevant to tennis (and other sports such as volleyball), which he called 'bipoints' scoring systems. He assumed that the probability player A (B) wins a point on service is pa (pb), and that points are independent. Noting the work of Wald (1947) and using Wn (point-pairs) as the standard family of scoring systems with *unit efficiency*, he showed (using the same 'interpolated' approach as above) that the efficiency of a general bipoints scoring system 1 relative to the bipoints scoring system 2 is also given by (2) above. That is, the relative efficiency 'formula' extends to the bipoints situation. Pollard (1992) noted that the most efficient bipoints scoring system (based on the 'play-the-loser' structure in the tennis context) also has the constant probability ratio characteristic.

Pollard and Pollard (2010), (1)) used the above interpolation method to show that (2) is also the measure for the relative efficiency in the independent quad-points case (e.g. tennis doubles with parameters pa1, pa2, pb1 and pb2 for the four players on service). Further, they showed (Pollard and Pollard, 2010), (2)), again using the interpolation approach, that (2) is the relative efficiency expression for scoring systems where unipoints or bipoints become one-step dependent probabilities.

It was possible to derive the relative efficiency for each of these four situations (unipoints, bipoints, quadpoints (e.g. tennis doubles), and 1-step dependent unipoints and bipoints) because in each case the underlying point probability structure of the situation being modeled lead to both a '(P, μ, n) equation' and the constant probability ratio property for the relevant underlying Wn system. This however is not always the case for the situation being modeled, and the 'interpolation' approach is then not possible. For example, supposing player A has a probability p of winning a point when the players are equal, p+ when he is ahead, and p- when he is behind, it can be seen that the Wn system of scoring systems does not have the constant probability ratio property when $n > 2$ and so the 'interpolated' approach is not possible. Nevertheless, when the 'interpolated' approach is not possible, it is possible to use an 'extrapolated' approach (Pollard and Pollard, 2012), which is now described.

Suppose the two scoring systems SS1 and SS2 have identical underlying probabilistic structures (for example they each involve unipoints, or bipoints, or something more complicated such as the example above with parameter values p, p+ and p-), and that SSi has an expected duration of μ_i points and a probability that player A wins of p_i ($i = 1, 2$). Then, considering the nested scoring systems Wn1(SS1) and Wn2(SS2), the probability player A wins Wn_i(SSi), P_i ($i = 1, 2$), can be evaluated using the relationship

$$P_i / Q_i = (p_i / q_i)^{n_i}$$

where $P_i + Q_i = 1$ and $p_i + q_i = 1$, and, using the recurrence methods outlined in Feller (1957), the expected duration of Wn_i(SSi) can be shown to equal

$$((P_i - Q_i)/(p_i - q_i))n_i\mu_i$$

Now suppose n1 and n2 are two (possibly very large) values such that player A has the same probability of winning under either nested system. That is, $P_1 = P_2$, and hence $P_1/Q_1 = P_2/Q_2$ and $P_1 - Q_1 = P_2 - Q_2$.

It follows that

$$n_1 / n_2 = (\ln(p_2 / q_2)) / \ln(p_1 / q_1)$$

Using the underlying concept of efficiency and noting that $P_1 = P_2$ for the two nested systems, the efficiency of the system Wn1(SS1) relative to the system Wn2(SS2) is given by the mean of Wn2(SS2) divided by the mean of Wn1(SS1). That is, it is given by the expression.

$$\frac{((p_1 - q_1) / \mu_1)\ln(p_1 / q_1)}{((p_2 - q_2) / \mu_2)\ln(p_2 / q_2)} \qquad (3)$$

ince the efficiency of Wn(SS) is equal to that of SS (Pollard and Pollard, 2012) as Wn has an efficiency of 1, it follows that the efficiency of the system SS1 relative to SS2 is given by expression (3). Thus, the expression for the relative efficiency for this case (where a '(p, μ, n) equation' does not necessarily exist) is identical to (2) (for the case when the '(p, μ, n) equation' and the constant probability ratio do exist). That is, the measure of relative efficiency is no longer limited to the situation where the underlying probability point structure necessarily allows a Wn system with the constant probability ratio property and a '(p, μ, n) equation' to be established. Thus, it is possible to find the relative efficiency of two scoring systems *in a much broader range of situations.*

Note that in the above, unlike in the work of Miles, the points within the scoring systems can be dependent. For example, in the example a p point can be followed by a p+ point or a p- point depending on the outcome on the p point. That is, independence of points within the scoring systems is not necessary. All that is necessary is to have pi and μi for the 2 systems.

Pollard and Pollard (2012) showed that expression (3) also applied in the situation where the two scoring systems SS1 and SS2 have expected durations of μ1 and μ2 points as above, but could result in a win, a draw, or a loss to player A with probabilities p_i, d_i, q_i respectively ($p_i + d_i + q_i = 1$, i = 1, 2). In this case, the efficiency of W1(SS1) relative to W1(SS2) is still given by (3), even though $p_i + q_i$ is less than 1, as can be the case in the tournament situations considered below.

2. METHOD

The method for evaluating the efficiency of a 4-person tournament is now outlined. We can consider a tournament won by the best player (player 1) as having a favourable outcome, and a tournament won by the weakest player (player 4) as having an unfavourable one. If either of the other players wins, the outcome can be considered neutral. Alternatively, a win by either of the best two players (1 or 2) might be considered favourable, and a win by either of the weakest two players (4 or 3) might be considered unfavourable, there being no neutral outcomes in this case. An analogy between a match involving two players and a tournament involving four players is made. A win by the better player in a match is replaced by a favourable outcome for the tournament and a loss by the better player is replaced by an unfavourable outcome, whilst a draw in a match is replaced by a neutral outcome. Thus, just as we can use (2) to find the relative efficiency of any two match scoring systems, we can use it to find the relative efficiency of any two tournament structures (for each favourable/unfavourable definition above). We will refer to the expression in the numerator of (2) as the *relative efficiency measure for tournament structure 1*, and the expression in the denominator of (2) as the *relative efficiency measure for tournament structure 2*. This relative efficiency can only be interpreted in relation to some other tournament structure.

2.1. Analysis of the Efficiency of 4-Player Tournament Structures

Four tournament structures are considered: knock-out 1 (K-O 1) in which player 1 plays player 4, and 2 plays 3 in the first round, the draw and process, the partial round robin (described below) and the round robin (RR). In the draw and process (D&P) the draw consists of K-O 1, and the process consists of a knock-out in which 1 plays 3 and 2 plays 4 in the first round (called knock-out 2). The winner of the draw plays the winner of the process in the final. If the same player wins both the draw and the process there is no need for a final. A referee has suggested that, as the 'term process applies normally to some sort of industrial application of a sequence of steps', the term 'redraw' might be used instead of 'process', as it 'has an intuitive meaning that should be understood by the average reader'. The authors agree with this assessment, but have continued to use the term 'process' as it has been used in a range of earlier publications and situations. In the RR each player plays each other player, so there are 6 matches in total. We have set up a tournament structure called the partial RR, in which the best two players each play the other two players, so there is a total of 4 matches. There are no ties possible in the K-O or the D&P and any ties within the RR or partial RR are split on an equal probability basis. It is noted in passing that knock-out 1 is more efficient than knock-out 2. Using the notation p_{ij} to represent the probability player i beats player j, we assume for simplicity that the probability of the better playing winning is 0.55 when the difference in ranks is one, 0.60 when the difference in ranks is two and 0.65 when the difference in ranks is three. The situation where the winner has been determined is called a 'reduction from 4 players to 1 player', and the case where 2 players remain is called a 'reduction to 2 players'.

It is noted that these probability values of 0.55, 0.60 and 0.65 above are used merely for illustrative purposes. Each sport has its own rating system, and in tennis for example the rating system is related to the order of finishing in a tournament and the importance of that tournament. The ordering of the players resulting from the application of the rating system can thus be different from the correct ordering. We note here that efficiency is not a criterion that is used by a range of tournament organizers. For example, in world competitions in sports such as football, rugby and basketball, a round robin group-phase format is desirable as players and supporters are rewarded with a minimum number of games, enhancing enjoyment and profitability of the tournament. In such cases a reasonably high probability that higher-rated teams advance in the tournament is more relevant than the tournament's efficiency. In these tournaments the probability that the higher-rated teams are successful is the important consideration, and not the efficiency. Nevertheless, an understanding of the efficiency of various tournament structures is of interest, and has been a topic studied by a range of researchers over the years.

We noted further that, whilst a particular focus has been given to tennis in this paper, the results would appear to be applicable to all sports and tournament considerations.

Table 1 shows the results for the reduction to 1 player. The measures refer to various successful tournament outcomes. For the first tournament outcome measure (1: 4), P is the probability player 1 wins, and Q is the probability player 4 wins. For the second measure (12: 43), P is the probability player 1 or 2 wins, and Q is the probability player 4 or 3 wins. The mean number of matches in each tournament and the tournament's relative efficiency measure is given. It can be seen that it is normally appropriate to focus on values of the relative efficiency measure when P + Q is not too small (not less than say 0.2), and not too large (not greater than say 0.85). For example, when P + Q is close to unity, outcomes that can be considered essentially neutral contribute to both P and Q, thus typically lowering the corresponding relative efficiency measure. In Table 1 the efficiencies of the tournaments are compared to the RR with K-O1 2.53 and 2.54 times more efficient than the RR for outcome measures 1 and 2 respectively. In the case of the first tournament outcome measure the D&P is 83% more efficient than the RR and the Partial RR is 50% more efficient than the RR. However, when the second tournament outcome measure the D&P is 97% more efficient than the RR and the Partial RR is 90% more efficient than the RR.

Table-1. Two relative efficiency measures for the reduction from 4 players to 1 player

Tournament Outcome Measure	Tournament	P	Q	Number matches, μ	Relative efficiency ((P-Q)/ μ)ln(P/Q)	Efficiency Relative to RR
1 : 4	K-O 1	0.3721	0.1479	3	0.0690	2.53
	D & P	0.4045	0.1237	6.7240	0.0495	1.83
	Partial RR	0.3541	0.1559	4	0.0407	1.50
	RR	0.3562	0.1572	6	0.0271	1
1 or 2 : 4 or 3	K-O 1	0.6485	0.3515	3	0.0606	2.54
	D & P	0.6937	0.3063	6.7240	0.0471	1.97
	Partial RR	0.6482	0.3518	4	0.0453	1.90
	RR	0.6322	0.3678	6	0.0239	1

For both measures of the reduction from 4 players to 1 player, the knock-out is the most efficient and the RR the least efficient structure, whilst the D&P is somewhat more efficient than the partial RR.

As an aside we have a very brief look at the relative merits of these 4 systems when the players are 'rated' is an incorrect order. This is merely a very brief introduction to such considerations, as this work is considered part of further studies rather than part of this study. Let us suppose for example that player 1 is seeded 2 and player 2 is seeded 1. Then it can be shown that the results for the D & P, partial RR and RR are unaltered. However, the K-O1 draw becomes [(2,4),(1,3)], and so the probability player 1 wins equals 0.354, the probability player 4 wins equals 0.156, and so the relative efficiency of the first measure (1 : 4) equals 0.0541, and the efficiency of K-O1 relative to the RR equals 2.00 rather than 2.53. Further, the probability player 2 wins equals 0.294 and the probability player 3 wins equals 0.196. Thus, for this second measure (1 or 2 : 3 or 4), P still equals 0.6485, Q still equals 0.3515, the relative efficiency still equals 0.0606, and the efficiency relative to RR still equals 2.54. Thus, it can be seen that the second measure is unchanged in regard to this particular incorrect 'rating'. As a second example let us suppose that player 2 was seeded 3 and player 3 was seeded 2. In this case it can be seen that the results for the K-O1 and RR are unaffected, whereas those for the D&P and the Partial RR are affected. Interestingly, if player 4 was seeded 1, player 3 was seeded 2, player 2 was seeded 3 and player 1 was seeded 4, the results for the 4 systems remain unchanged. There are clearly $4! - 1 = 23$ such incorrect ordering for 4 players and $8! - 1 = 40{,}319$ incorrect ordering for the case of 8 players. Due to the apparent complexity of a study of the effect of incorrect 'ratings', further study of this issue is not considered in this paper.

The reduction from 4 to 2 players is now considered. In the case of the D&P, these 2 players are the winner of the draw and the winner of the process, and unlike the other three structures, it is possible that the winner of the draw is the same player as the winner of the process. The results are given in Table 2 showing that the K-O1 is 2.40 and 2.43 times more efficient than the RR for outcome measures 1 and 2 respectively. In the case of the first tournament outcome measure the D&P and Partial RR are respectively 74% and 43% more efficient than the RR and in the case of the second tournament outcome measure the D&P and Partial RR are respectively 53% and 18% more efficient than the RR.

Table-2. Two relative efficiency measures for the reduction from 4 to 2 players

Tournament Outcome Measure	Tournament	P	Q	Number matches, μ	Relative efficiency ((P-Q)/ μ)ln(P/Q)	Efficiency Relative to RR
12 : 43	K-O 1	0.3575	0.1575	2	0.0820	2.40
11, 12 : 44, 43	D & P	0.3390	0.0838	6	0.0594	1.74
12 : 43	Partial RR	0.3208	0.1208	4	0.0488	1.43
12 : 43	RR	0.2712	0.0882	6	0.0342	1
12, 13 : 43, 42	K-O 1	0.65	0.35	2	0.0929	2.43
11, 12, 22, 13 : 44, 43, 33, 42	D & P	0.5652	0.2103	6	0.0585	1.53
12, 13 : 43, 42	Partial RR	0.4860	0.2360	4	0.0451	1.18
12, 13 : 43, 42	RR	0.4789	0.2064	6	0.0382	1

For both outcome measures for the reduction to 2 players, the most efficient structure is again the knock-out structure, and the least efficient is the RR. As above the D&P is more efficient than the partial RR. *The overall conclusion for the 4 player situations is that the order of efficiency from most efficient to least is knock-out, D&P, partial RR and RR*. The fact that there is more than one measure of efficiency for each 'reduction' is not seen as a problem. For example, in statistics there are several measure of central tendency such as the mean, the trimmed mean, the median and the mode, and several measures of spread such as the standard deviation, the mean deviation, the range and the interquartile range.

2.2. Analysis of the Efficiency of 8-Player Tournament Structures

We now consider tournaments with 8 players, denoted by 1, 2, 3 … 8 from best to weakest. The knock-out (K-O) considered here takes the form ((1, w), (4, x)), ((3, y), (2, z)), where players 1 and 4 are drawn to meet in the semi-finals and where w, x, y and z is a random permutation of players 5 to 8. In the D&P, the draw is the above knock-out and the process is ((1, z), (3, x)), ((4, y), (2, w)). In the RR each player plays each other player (28 matches), and in the partial RR, the best 4 players play the weakest 4 players (16 matches).

Simulations of 1,000,000 tournaments were carried out in order to produce the results in the following tables. All ties were split on an equal probability basis. Various checks on the simulations (such as observing the results when all players were equal) were carried out. In order to test the sensitivity of the results to the probability values assumed, in these simulations two sets of probability were considered, the second set of probabilities reflecting a much closer tournament.

Table-3. Probabilities that Better Player Wins for a Match between Two Players

Difference in ranks for 2 players	1	2	3	4	5	6	7
Less Competitive Tournament	.54	.58	.62	.66	.70	.74	.78
More Competitive Tournament	.52	.54	.56	.58	.60	.62	.64

There are 4 measures for reducing from 8 to 1 player. Measure 1 has a win to player 1 as the favourable outcome, and a win to player 8 as the unfavourable outcome. Measure 2 has player 1 or 2 winning as the favourable outcome, and player 8 or 7 winning as the unfavourable outcome. It can be seen from Figure 1 and Table A1 in the Appendix that on all measures the order of relative efficiency from most efficient to least is K-O, D&P, partial RR and RR and that this order is preserved for closer tournaments with a difference of only 0.02 in the better player's winning probability for successive ranks. Relative efficiency is lower in the case of closer tournaments. It is clear that the RR is really quite inefficient relative to the other three structures. This is not surprising as the RR makes no use of prior information. However, the inferiority of the RR is less obvious in the case of very close tournaments.

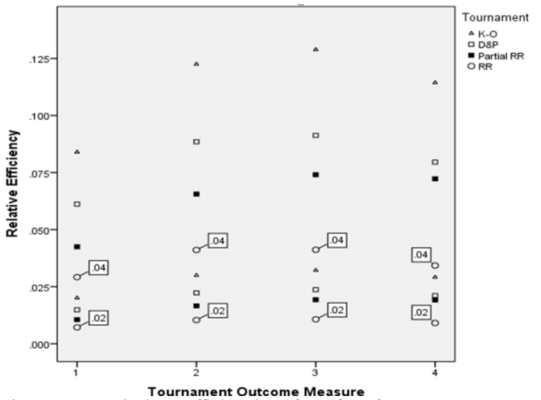

Figure-1. Relative efficiencies for the four tournament outcome measures for the reduction from 8 players to 1 player with differences in probability of winning of 0.02 (below) and 0.04 (above) for successive player rankings

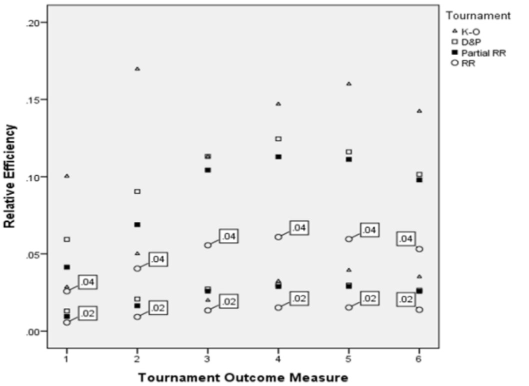

Figure-2. Relative efficiencies for the six tournament outcome measures for the reduction from 8 to 2 players with differences in probability of winning of .02 (below) and .04 (above) for successive player rankings

There are 6 measures for the reduction from 8 to 2 players. Measure 1 has players 1 and 2 remaining as the favourable outcome, and players 8 and 7 remaining as the unfavourable outcome. Measure 2 has players 1 and 2 or 1 and 3 remaining as the favourable outcomes, and players 8 and 7 or 8 and 6 remaining as the unfavourable outcomes. Measures 3 to 6 can be identified as indicated below. These measures can also be described in terms of the maximum sum of ranks for a favourable outcome and the minimum sum of ranks for an unfavourable outcome. For example as shown in Table 4, for measure 6 ranks 1 and 7 in the final would be regarded as a favourable outcome while ranks 8 and 2 in the final would be regarded as an unfavourable outcome. (Note for example that the outcomes (11) and (88) are only possible in the D&P, and therefore a maximum of 2 for the favourable sum of ranks and a minimum of 16 for the unfavourable sum of ranks are not shown in Table 4.

Table-4. Definition of measures for a favourable and unfavourable outcome for the reduction from 8 to 2 players

Measure	1	2	3	4	5	6
Maximum Favourable Sum of Ranks	3	4	5	6	7	8
Minimum Unfavourarable Sum of Ranks	15	14	13	12	11	10

The results for the reduction from 8 players to 2 are given in Figure 2 and Table A2 in the Appendix. It can be seen that on all measures the order of relative efficiency from most efficient to least is essentially K-O, D&P, partial RR and RR for both levels of competition (2% and 4% winning probability differences for successive ranks). It is noted that on measures 3 and 4 the D&P has a comparable efficiency to the K-O. Interestingly, while on measures 3, 4, 5 and 6 the relative efficiency of the partial RR is not hugely less than that of the D&P. Again it is clear that the RR is really quite inefficient relative to the other three structures. Of particular interest is the heightened relative efficiency of the K-O on measure 2. Because of the seeding employed in the K-O, measure 2 suggests that the chances of two of the first three seeds reaching the final is greatly enhanced.

There are 8 measures for reducing to 4 players. Measure 1 has players 1, 2, 3 and 4 remaining as the favourable outcome, and players 8, 7, 6 and 5 remaining as the unfavourable outcome. (Note that the D&P has additional outcomes in Measure 1. The favourable outcomes are ijkl where i and j are the two players who reach the final of the draw, k and l are the two players who reach the final of the process, and the sum i+j+k+l is less than or equal to 10). Measure 2 has players 1, 2, 3 and 4 or players 1, 2, 3 and 5 remaining as the favourable outcomes, and players 8, 7, 6 and 5 or 8, 7, 6 and 4 remaining as the unfavourable outcomes. (Again, the D&P has additional outcomes in this measure). Thus, these measures can be described in terms of the maximum and minimum sum of the ranks as shown in Table 5. This table shows for each measure the largest sum of the rankings for the last 4 players if a favourable outcome is to be achieved in a tournament and the smallest sum of the rankings for the last 4 players if an unfavourable outcome is to be achieved in a tournament.

Table-5. Definition of measures for a favourable and unfavourable outcome for the reduction from 8 to 4 Players

Measure	1	2	3	4	5	6	7	8
Maximum Favourable Sum of Ranks	10	11	12	13	14	15	16	17
Minimum Unfavourable Sum of Ranks	26	25	24	23	22	21	20	19

We have not done the simulations for the cases of the RR and the partial RR. Considering the RR for example there are 28 match results in total. Suppose for example that players A and B tie, C and D tie, and E, F and G tie, this being a possibility even if unlikely. It is clear that there are a large number of other ways in which ties might occur in general, and that there are 28 possible selections just for players A and B in the above. Thus, it can be seen that in total there are a very large number of ways in which ties might occur. For this reason, it was decided to leave this analysis for a later study.

It can be seen from Figure 3 and Table A3 in the Appendix that across both probability structures the K-O is more efficient than the D&P in the reduction from 8 to 4 players.

Nevertheless, the D&P has quite reasonable relative efficiency on these measures even though it is a comparatively unknown tournament structure.

It is interesting to again note that if a tournament structure is more efficient than another at identifying the best of 8 players, then it is typically also comparable or more efficient than the other at reducing the number of players from 8 to 2 or 4 players. That is, there is no particular conflict in selecting a tournament structure based on its efficiency in reducing to 1, 2 or 4 players.

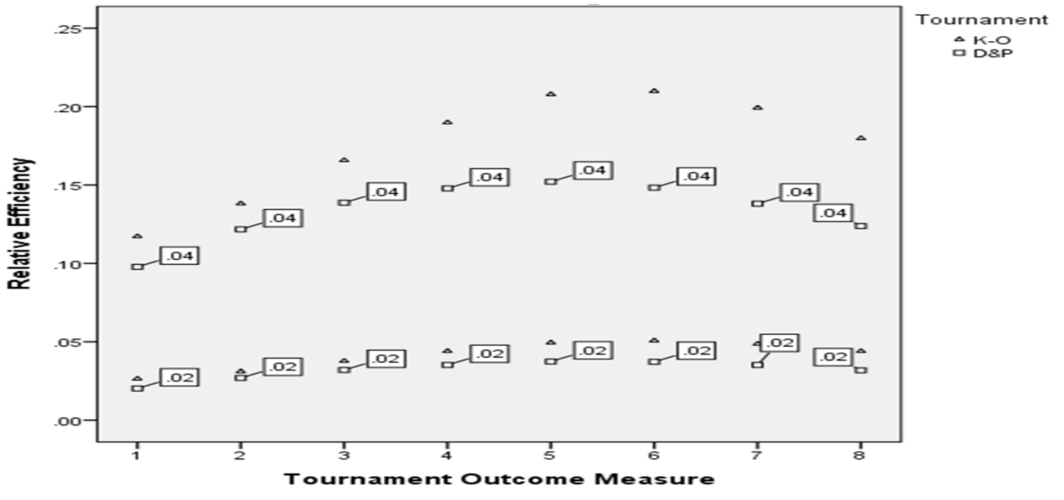

Figure-3. Relative efficiencies for the six tournament outcome measures for the reduction from 8 to 4 players with differences in probability of winning of .02 (below) and .04 (above) for successive player rankings

2.3. Some Notes on Measuring Absolute Efficiency

In the previous section we have considered relative efficiency measures for various tournament structures. We now show how the actual (or absolute) efficiency rather than the relative efficiency can be calculated in an analogous manner to that of Miles (1984) for the tournament situation of reducing 4 players to 2 The method is also similar to the method used by Pollard and Pollard (2008).

In the 4-person tournament situation we suppose players 1 and 2 are the strongest players and 3 and 4 are the weakest, and in a corresponding manner to the earlier work mentioned above, we set up the *standard family of tournament scoring systems with unit efficiencies* Wn(W2(13, 14), W2(23, 24)), (n = 1, 2, …). Note that the case when n=1 is one tournament structure, the case when n=2 is another tournament structure with larger expected duration, etcetera. Here the notation 13 represents a match between players 1 and 3, etcetera, and W2(13, 14) represents a tournament scoring system in which matches 13 and 14 alternate until player 1 is 2 matches ahead of players 3 and 4 combined (a favourable outcome), or until players 3 and 4 combined are 2 matches ahead of player 1 (an unfavourable outcome), when this first section of above scoring system is completed. The second section of the above tournament scoring system W2(23, 24) then commences and is won by player 2 (a favourable outcome) or by players 3 and 4 combined (an unfavourable outcome). Thus, players 1 and 2 combined win both these sections of the above scoring system (and win overall when n=1), or players 3 and 4 combined win both sections (and win overall when n=1), or there is a draw between players 1 and 2 combined and players 3 and 4 combined, in which case the tournament scoring system starts over again and continues until it is won by players 1 and 2 combined or players 3 and 4 combined. Note that at the completion of this tournament scoring system the winning pair will be 4 matches ahead of the losing pair when n=1. This tournament scoring system is set up, not because it has any particular practical relevance, but because it provides a (natural) family of scoring systems (with unit efficiency) against which the efficiency of practical and relevant tournament scoring systems can be evaluated (in an absolute manner).

We consider an example of the above tournament structure (for the case when n = 1), supposing for simplicity that the probability player i beats player j , p_{ij}, is equal to 0.6 if i equal to 1 or 2 and j equal to 3 and 4. Then, using the results of Pollard (1992) (Theorem 5 with n=2, t=1, $p_a = p_{13}$ and $q_b = p_{14}$, on page 280), and making use of the constant probability ratio property which this tournament scoring system possesses, it follows that, for this first section W2(13, 14), $P_1/Q_1=(p_{13}*p_{14})/(q_{13}*q_{14})$ and $(P_1-Q_1)/\mu_1 = (p_{13}+p_{14}-1)/2$ where P_1 is the probability player 1 wins this section, $Q_1=1-P_1$, $q_{ij}=1-p_{ij}$, and μ_1 is the mean number of matches in this section. Thus, for these parameter values, W2(13, 14) has $P_1=9/13$ and $\mu_1=50/13$, and correspondingly the second section W2(23, 24) also has $P_2=9/13$ and $\mu_2=50/13$, where P_2 is the probability player 2 wins the second section, $Q_2=1-P_2$ and μ_2 is the mean number of matches in the second section. Thus, for the complete scoring system (when n = 1), analogous to the methods of Pollard and Pollard (2010), (1)), we have $P_{12}=P_1*P_2/(P_1*P_2+Q_1*Q_2)$ and $\mu_{12}=(\mu_1+\mu_2)/(P_1*P_2+Q_1*Q_2)$ where P_{12} is the probability players 1 and 2 win under the complete system and μ_{12} is the mean number of matches in the complete system, and $P_{12}=81/97$ and $\mu_{12}=1300/97$ for the above parameters. It is noted from these values of P_{12} and μ_{12} that the relative efficiency measure for this tournament scoring system is equal to 0.0811, and that this is in fact the same value as the relative efficiency of a single point-pair with each p-value equal to 0.6, namely (0.36-0.16)/2)ln(0.36/0.16). Also, it can be shown that the relative efficiency measure $((P_{12}-Q_{12})/\mu_{12})\ln(P_{12}/Q_{12})$ is equal to the value for the first tournament outcome measure (12 : 43) for K-O1.

Thus, for this case of reduction from 4 to 2 players, K-O1 has an absolute efficiency of 1. In this example we have assumed that the probability player i beats player j , p_{ij}, is equal to 0.6 if i equal to 1 or 2 and j equal to 3 and 4. However, it can be shown using the same method that, for the case of 4 players reduced to 2, K-O1 must have unit efficiency when $p_{13} = p_{23}$ and $p_{14} = p_{24}$, and that for more general parameter values it has an efficiency close to unity.

Further, again using the methods of Pollard and Pollard (2010) (1), the above approach for evaluating the actual efficiency rather than relative efficiency when reducing 4 players to 2, can be extended to reducing 8 players to 4, and when reducing 16 players to 8.

3. LIMITATIONS

There are at least two limitations in this study. Importantly, only two sets of pairwise probabilities are selected for the better player winning. The results suggest that relative tournament efficiency is not affected by these probabilities, but it is clear that the differences in efficiency between tournament structures is less noticeable in the case of close tournaments. Secondly, no attempt has been made to simulate an unseeded K-O tournament. The results for a seeded K-O indicate that the efficiency of the tournament in terms of two of the best three players reaching the final is greatly enhanced. Clearly an unseeded K-O would be less efficient than a seeded K-O but the effect on each of the measures, particular for the reduction to 2 players would be of interest.

4. CONCLUSIONS

A methodology for determining the relative efficiency of tournament structures has been described in this paper. The method is a very general one, and can be used for evaluating the efficiency of all tournament structures, be they for sports such as the various types of football, for the various racket sports, for sports like volleyball, for indoor sports like darts and snooker, or even for games like chess. In this paper simulation is used to show how this method can be used for assessing the relative efficiency of tournament structures at reducing 2^n players to 2^{n-1} players, 2^{n-2} players, … and 1 player. The round robin, the partial round robin (a new structure), the knock-out and the draw and process have been considered in detail for the cases of 4-player and 8-player tournaments. It has been shown that the most efficient tournament structure (for a typical parameter set) is the knock-out and that the round robin is quite inefficient. In fact the order of efficiency from highest to lowest is knock-out, draw and process, partial round robin and round robin. Further, it has been demonstrated that the knock-out structure is a very efficient structure in absolute terms. These results have been confirmed for the case of a close very competitive tournament and for the case of a less competitive tournament. The knock-out structure has been shown to be very efficient in absolute terms, rather than just relative to other commonly used tournament structures. Thus, there is no particular need in terms of efficiency for anyone to think about tournament structures that could possibly be more efficient than the knock-out structure.

However, this paper also acknowledges that tournament efficiency is not the only concern when choosing an appropriate tennis structure. The accuracy of rankings should also be taken into consideration. The KO-1 system is particularly sensitive to ranking inaccuracy while the RR is not at all sensitive to ranking inaccuracy. In order to allow for ranking inaccuracy it is recommended that instead of just considering measure 1 (efficiency for highest ranked player or team winning as opposed to the lowest ranked player or team winning), more inclusive measures should be used (e.g. efficiency for either of two highest ranked players or teams winning as opposed to the two lowest ranked players or teams winning). A further concern of tournament organisers relates to the profitability of the tournament and the satisfaction of tournament followers and players. Ideally every player/team should have the opportunity to have more than one match, making the knockout not the ideal structure despite its high efficiency.

Funding: This study received no specific financial support.

Competing Interests: The authors declare that they have no competing interests.

Contributors/Acknowledgement: All authors contributed equally to the conception and design of the study. The authors thank a reviewer for suggestions made on their submitted version of the paper.

REFERENCES

Appleton, D.R., 1995. May the best man win. The Statistician, 44(4): 529-538.

Feller, W., 1957. An introduction to probability theory and its applications. New York: John Wiley & Sons.

McGarry, T. and R.W. Schutz, 1997. Efficacy of traditional sport tournament structures. Journal of Operational Research Society, 48(1): 64-74.

Miles, R.E., 1984. Symmetric sequential analysis: The efficiencies of sports scoring systems with particular reference to those of tennis. Journal of the Royal Statistical Society B, 46(1): 93-108.

Pollard, G.H., 1992. The optimal test for selecting the greater of two binomial probabilities. Australian Journal of Statistics, 34(2): 273-284.

Pollard, G.H. and G.N. Pollard, 2010. The efficiency of scoring systems with outcome dependent points. In: Proceedings of the Tenth Australasian Conference on Mathematics and Computers in Sport, edited by A. Bedford and M. Ovens, Mathsport (ANZIAM). pp: 35-42.

Pollard, G.H. and G.N. Pollard, 2012. A general measure for the relative efficiency of any two scoring systems. In: Proceedings of the Eleventh Australasian Conference on Mathematics and Computers in Sport, edited by A. Bedford and Adrian Schembri, Mathsport (ANZIAM). pp: 162-169.

Pollard, G.N. and G.H. Pollard, 2008. The efficiency of doubles scoring systems. In: Proceedings of the Ninth Australasian Conference on Mathematics and Computers in Sport, edited by J. Hammond, Mathsport (ANZIAM). pp: 45-51.

Pollard, G.N. and G.H. Pollard, 2010. The efficiency of doubles scoring systems. Journal of Sports Science and Medicine, 9(3): 393-397.

Wald, A., 1947. Sequential analysis. New York: Wiley.

Wald, A. and J. Wolfowitz, 1948. Optimal character of the sequential probability ratio test. Ann. Math. Statist., 19(3): 326-339.

APPENDIX

Table-A1. Results for the four relative efficiency measures for the reduction from 8 to 1 player with differences in probability of winning of 0.04 for successive player rankings

Tournament Outcome Measure	Tournament	P	Q	# matches	Relative Efficiency
1)1 : 8	K-O	0.2833	0.02826	7	0.0840
	D & P	0.3246	0.01702	14.817	0.0612
	Partial RR	0.2847	0.02145	16	0.0425
	RR	0.3119	0.01917	28	0.0292
2)12 : 87	K-O	0.5014	0.06900	7	0.1225
	D & P	0.5636	0.04503	14.817	0.0885
	Partial RR	0.5149	0.05300	16	0.0656

	RR	0.5402	0.05137	28	0.0411
3)123 : 876	K-O	0.6683	0.1265	7	0.1289
	D & P	0.7285	0.08815	14.817	0.0913
	Partial RR	0.6983	0.09716	16	0.0741
	RR	0.7036	0.1029	28	0.0412
4)1234 : 8765	K-O	0.7952	0.2048	7	0.1144
	D & P	0.8460	0.1540	14.817	0.0796
	Partial RR	0.8434	0.1566	16	0.0723
	RR	0.8185	0.1815	28	0.0343

Table-A2. Results for the six relative efficiency measures for the reduction from 8 to 2 players with differences in probability of winning of 0.04 for successive player rankings

Tournament Outcome Measure	Tournament	P	Q	# matches	Relative Efficiency
1	K-O	0.1885	0.006855	6	0.1003
1	D & P	0.2029	0.003156	14	0.0594
1	Partial RR	0.1688	0.003097	16	0.0414
1	RR	0.1663	0.002034	28	0.0258
2	K-O	0.3447	0.01562	6	0.1697
2	D & P	0.3443	0.008003	14	0.0904
2	Partial RR	0.3042	0.007407	16	0.0689
2	RR	0.2870	0.005121	28	0.0405
3	K-O	0.3447	0.03784	6	0.1130
3	D & P	0.4894	0.01714	14	0.1131
3	Partial RR	0.5211	0.01878	16	0.1043
3	RR	0.4635	0.01447	28	0.0556
4	K-O	0.4947	0.06387	6	0.1470
4	D & P	0.6175	0.03403	14	0.1245
4	Partial RR	0.6442	0.03347	16	0.1129
4	RR	0.5898	0.02830	28	0.0609
5	K-O	0.6536	0.1113	6	0.1600
5	D & P	0.7272	0.06294	14	0.1161
5	Partial RR	0.7701	0.06236	16	0.1112
5	RR	0.7234	0.05881	28	0.0596
6	K-O	0.7405	0.1669	6	0.1424
6	D & P	0.8172	0.1097	14	0.1015
6	Partial RR	0.8354	0.0994	16	0.0979
6	RR	0.8150	0.1015	28	0.0531

Table-A3. Results for the eight relative efficiency measures for the reduction from 8 to 4 players with differences in probability of winning of 0.04 for successive player rankings

Tournament Outcome Measure	Tournament	P	Q	# matches	Relative Efficiency
1	K-O	0.1861	0.01250	4	0.1172
1	D & P	0.2591	0.002661	12	0.0978
2	K-O	0.2252	0.01605	4	0.1381
2	D & P	0.3563	0.005539	12	0.1217
3	K-O	0.2923	0.02460	4	0.1657
3	D & P	0.4533	0.01055	12	0.1387
4	K-O	0.3778	0.03992	4	0.1898
4	D & P	0.5420	0.01832	12	0.1478
5	K-O	0.4904	0.06840	4	0.2078
5	D & P	0.6358	0.03109	12	0.1521
6	K-O	0.5779	0.09987	4	0.2098
6	D & P	0.7132	0.04896	12	0.1483
7	K-O	0.6679	0.1454	4	0.1991
7	D & P	0.7857	0.07600	12	0.1381
8	K-O	0.7389	0.1964	4	0.1797
8	D & P	0.8420	0.1105	12	0.1238