

# Comparing different tournament systems in their probability that the best player wins

Norico Groeneveld, Jozef Siu

March 30, 2017

## Abstract

This paper compares the performance of different tournament systems. The performance is measured by looking at the probability that the best player wins, and the average rank of the winning player. This is important when selecting a tournament system that meets the requirements and expectations of the people who organise a tournament. The results are determined by running computer simulations for each tournament setup.

**Keywords**— Tournament systems, Elo rating, elimination, round-robin, Swiss-system, multi-stage, win probability

## 1 Introduction

In competitive games or sports in which tournaments are played, competitors are playing to find out who is the best among them. To satisfy that, it's necessary to provide a tournament format that allows for the best player to win and reflect his true skill. Szymanski also said "Designing an optimal contest is both a matter of significant financial concern for the organizers, participating individuals, and teams, and a matter of consuming personal interest for millions of fans." [Szy03] There are many variables that contribute to the result of a tournament. The two major categories are the tournament setup and player skill, and then there are some uncontrollable variables that some would contribute to luck, such as fatigue. The tournament setup is determined by the organisers and as a player you have little control over that factor. For the organisers it's therefore their responsibility to provide a tournament system that accurately reflects the player's skill. To be able to use a tournament system that accurately reflects the player's skill, we are researching the probability that the best player wins in different tournament systems.

The variable of tournament setup will be measured in this paper, as for the player skill, we'll be emulating that through Elo rankings as is used in chess for example.

In our paper we will be researching the following tournament systems: single elimination, double elimination, single round-robin, double round-robin, Swiss-system and the multistage system. We've chosen these as elimination, round-robin and swiss are common tournament systems played in chess. [Fed] In our preliminary research we've found several studies that compares the performance of some setups. P. Scarf et al. found that the round-robin system performs better than the multistage system. [SYB09] While W. Elmenreich et al. found that the round-robin system performs better than Swiss-system. [EIF09] McGarry and R.W. Schutz show that the round-robin system performs better than the double elimination system, which performs better than the single elimination system. They also show that double elimination is sometimes better than round-robin, when looking at certain metrics that aren't used in our research. [MS97] It's likely that we'll be finding similar results as we're aiming to gain an insight at the relative performance of all these tournament systems.

In this research, we aim to look at the relative performance of all these tournament systems.

## 2 Methods

The data that is required for this research topic should first of all contain data of winners of tournaments. Second, to determine their ranking, their elo rating is necessary. To gather this data there is the possibility of gathering historic tournament results of chess tournaments. However

that would require crawling the online database and stitching pieces of data. This would also limit our research to the round robin, swiss and elimination modes.[Fed] Instead of using historic data we've chosen to simulate our tournaments. This allows for more control over the data and results in complete data. It also gives us the opportunity to compare some other tournament modes.

## 2.1 Tournament systems

The tournament systems we simulate are based on matches between two players, from which one loses and one wins.

In an elimination tournament, players play one match every round. An elimination tournament can be a single or double. In a single elimination tournament, only the winner of a match continues in the tournament. At the final round two players are left of which the winner wins the tournament. In a double elimination tournament, the winner of a match goes to the next round, just like in a single elimination tournament. However, the loser still has a chance to win the tournament, because he enters a parallel bracket of matches where all the losers compete. The losers of the first round are divided in pairs, who each play one match. The winners of those matches compete with the losers of the second round, again divided in pairs and with one match per pair. The winners of those matches compete against the losers of the third round in the same way, and so on. At the end, there is one winner of the main bracket and one winner of the loser bracket. If the one from the main bracket wins, he's declared the winner. If the winner is the player from the losers bracket, the two play another match, the winner of which is the winner of the tournament. To increase the diversity of the matches in the losers bracket, players in the losers bracket aren't matched the same way every time. In the odd rounds, the first half of the players who are already in the losers bracket compete against the first half of the players who are just coming into the losers bracket and the second half competes against the second half. In the even rounds, the first half competes against the second half and the second half competes against the first half.

In a round-robin tournament, every player plays one or two matches against everyone else, depending on if it's a single or a double round-robin tournament. After this, the player with the highest number of points is the winner. It's possible to have multiple players with the same number of wins at the end of the tournament. We handle that by letting them compete in a round-robin tournament with all those players. Depending on the original tournament, that finale is also a single or a double round-robin tournament.

In a Swiss-system tournament, players only play against opponents with a similar tournament score. We use a simplified version where players can possibly play against the same competitor multiple times. We use two simple ways to create an initial pairing. In the first variation, players are paired up randomly. In the second variation, players are sorted by their Elo rating and paired up, starting at the highest rating. After every set of matches, players are sorted by the number of wins in the tournament and are paired up, starting at the highest number of wins. The tournament stops when there is a winning player who has won more matches than all the other players.

In a multistage tournament, sometimes called a group tournament, players are divided in groups. There are one or more stages where the players in a group compete in the same way as a single round-robin tournament. The best player in each group advances to the next stage. At a certain moment, the group phase stops and the tournament continues like a normal single elimination tournament. We use four different multistage tournaments, called A, B, C and D. The parameters are shown in Table 1.

Identification	A	B	C	F
Group size	32	32	16	16
Number of rounds with groups	2	1	2	1
Number of remaining players who compete using single elimination	None	32	4	64

Table 1: The multistage tournament parameters

## 2.2 Winner

When two players compete, the better player is more likely to win, in accordance with how much better he is than his competitor. Using Elo's rating, we have a method of representing the difference in skills between two players. The formula for the expected score takes as input the two Elo scores

of two players and returns the probability that one player wins. A random number generator then determines which of the two players wins the game.[E109]

$$ExpectedscoreP1 = \frac{1}{1 + 10^{(P2-P1)/400}}$$

## 2.3 Elo ratings

In our simulations, we use 1024 players, with an Elo rating ranging from 1000 to 2000. They can be ranked according to their Elo rating, where the player with rank 1 has the best Elo rating. The Elo ratings are spaced out evenly, so two consecutive players are 1000/1024 points apart. A difference of one rank between two players would result in a 50,141% win chance for the higher ranked player when using the formula for the expected score. The difference in Elo rating between two players is surely a factor in determining the probability that the best player wins, however this paper focuses on the effect the tournament system has on that probability. For this reason we've kept the Elo range on a constant 1000 to 2000.

Rank difference	1	10	100	1000
Elo points difference	0,97656	9,7656	97,656	976,56
Expected score/win chance	50,141%	51,405%	63,695%	99,639%

Table 2: Expected scores

## 2.4 Simulation

Each and every tournament is simulated a 100.000 times with the same players and same rankings. Each tournament simulation outputs the ranking of the winner, which are then counted for each player. At the end of the simulation the result will be the number of wins per player sorted on their Elo rank. We divide this by the total number of simulations to get a win probability

## 3 Results

The results show a convergence to 0 wins for the lower ranked players as would be expected. And as we focus on the probability that the best player wins, the lower end of the rankings are less significant. For this reason we've put our full results in a logarithmic scale. Figure1

We've also put the cumulative win rate of the players into perspective, with that we can easily determine the convergence and spread of the each tournament system. Figure2

### 3.1 Descriptive statistics

	Mean	Median	Std. Deviation	$\sigma/\sqrt{(N-1)}$	Worst winning player
Round-robin double	6,71	5	5,53	0.0175	48
Round-robin single	8,86	7	7,5	0.024	61
GroupA	23,19	17	20,694	0.065441	191
GroupB	38,13	29	33,991	0.10749	295
GroupC	32,37	24	29,01	0.09174	259
GroupD	42,44	31	38,401	0.12144	337
Double elimination	42,54	31	38,594	0.12205	367
Single elimination	46,5	34	42,176	0.13337	458
Swiss-random	46,37	35	41,898	0.13249	399
Swiss-sorted	59,2	42	58,181	0.18399	570

Table 3: Descriptive statistics

The worst winning player is the player with the lowest rank, and thus the lowest Elo rating, who has won the tournament at least once.

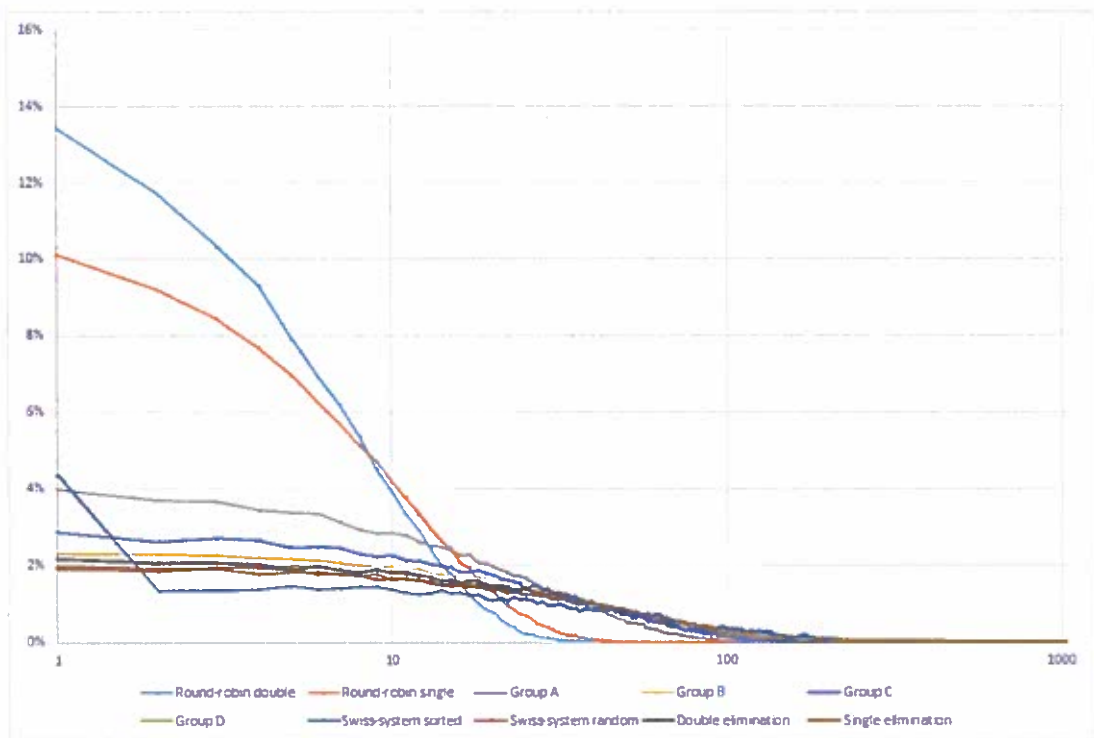


Figure 1: Probability of winning the tournament on a logarithmic scale per rank

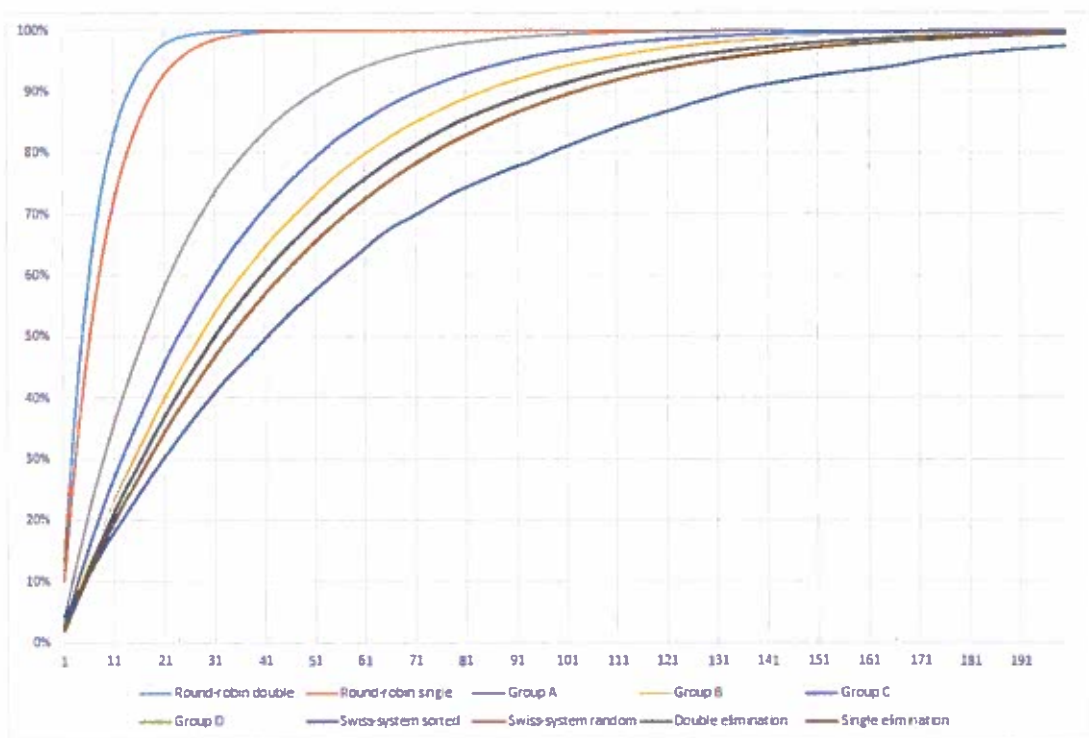


Figure 2: Cumulative probability of winning the tournament for each system

## 4 Discussion

The tournament systems we researched can be divided in different groups: round-robin, multistage, elimination and Swiss-system. We can compare the tournament systems in different ways: all the systems, the systems in the same group, and the different groups of systems.

### 4.1 All tournament systems

The results show that the player who has the most wins is also the player with the highest Elo rating.

Figure 1, the winning probability graph, shows that the round-robin system allows for a significantly higher win probability for the best player. The double round-robin results in the highest win rate for the best players. Which is followed by the single round-robin. Figure 2, the cumulative win rate graph, also supports the double round-robin as best tournament system, as the double round-robin converges the fastest, which means that fewer lower ranked players have a chance of outright winning the entire tournament.

### 4.2 Tournament systems in the same group

All results show that of the multistage tournaments, group A is the best, followed by C, B and D. This makes sense, because that is the order of similarity with round-robin and unsimilarity with single elimination.

For the Swiss-system with sorted initial pairing we found that it has a higher win probability for the best ranked player than the Swiss-system with random initial pairing. However, this only holds for the best player. According to the results the system heavily favors only the single best player. For the remainder of the ranks it has worse results than the randomized pairing system. The Swiss-random system also converges faster, which means that overall the Swiss-random performs better.

As for the elimination system, all results show that double elimination is slightly better than single elimination. But not significantly. This is likely due to the fact that the very best player gets favorable pairings, however to clarify this, more research is needed.

### 4.3 Different groups of tournament systems

Looking at the cumulative winning probability and the average winning rank, round-robin is by far the best of the different groups of tournament systems. After that, the second best group is the multistage system, closely followed by the elimination and Swiss-systems. Notably, the winning probability of the best ranked player shows something different. For the first rank, the Swiss-system with sorted initial pairing is best, after round-robin. And, the Swiss-system with random initial pairing performs worst at rank one. Therefore, it's not possible to tell where this group belongs in the order of tournament system groups, looking at this result. A downside of round-robin system tournaments are that they require a lot of matches. It may be better to sacrifice a bit of quality of the tournament, in favor of less required matches.

### 4.4 Our simulation

Our simulations use a small Elo difference between the ranks. Because of this, the win probability between two consecutive players is also quite small. If the Elo difference is larger, all the tournaments will most likely be better at letting the best player win the tournament. The cumulative winning probability converges to 100% faster as well. However, the relative performance of the tournament systems doesn't change. So this should not be of relevance to our results.

Our simulations use an Elo difference that is evenly spaced out. This is of course not in accordance with real life. But for the sake of controllability and statistical modelling this was left out as we mentioned before. For further studies it may be interesting how the spacing in Elo ratings or rather skill, may impact the tournament results in the various tournament systems.



## 4.5 Further research

We only gathered the number of players who won a tournament. Players who came in second or later, are not taken in account in our results. This could distort how well the results show which tournaments are better. It doesn't show how the other positions in the tournament result are spread out. This is a topic for further research.

When simulating the multistage tournaments, there are two variables. These are the size of the groups and the number of players who compete in a single elimination tournament (instead of continuing the group phase). We only simulated tournaments with four combinations of the two variables. Another thing that can be changed is the amount of players who continue after all the matches in a group are played. In our simulation, only the best player continues. In a more complicated system, multiple players can continue. There could be further research about what the effect of these parameters on the performance of the tournament system is, and what this means for the relative performance, compared to other tournament systems.

In our simulation, we used a simplified version of the Swiss-system tournaments. More advanced variations exist and are used in real life. These variations differ both in the method for initial pairing as for pairing during the rest of the tournament. Besides that, they don't allow two players to play more than one match. When these variations are researched, they might turn out to perform better.

P. Scarf et al. put the performance of their tournament systems, round-robin and multistage tournament, against the amount of matches that is required. For further research it may be of interest to do the same for our results, which may lead to even more usable data. In our research we've not taken this into account as the scope is limited to the performance of each system. [SYB09]

## 5 Conclusion

With our constraints and controlled variables, which allows us to only evaluate the influence of the tournament mode on the win probability of the best players. We've found that the (double)round robin tournament system performs best. Both at the win probability for the best player as at the speed at which it converges. This makes it the overall best performing system. The other tournament systems perform significantly less. Of these we conclude that the multi stage tournament systems perform best, followed by the elimination systems and swiss systems.

## References

- [EIF09] Wilfried Elmenreich, Tobias Ibounig, and István Fehérvári. Robustness versus performance in sorting and tournament algorithms. *Acta Polytechnica Hungarica*, 6(5):7–18, 2009.
- [Fed] World Chess Federation. Fide handbook. [https://www.fide.com/handbook?option=com\\_handbook](https://www.fide.com/handbook?option=com_handbook). [Online; accessed 24-March-2017].
- [MS97] T McGarry and RW Schutz. Efficacy of traditional sport tournament structures. *Journal of the Operational Research Society*, 48(1):65–74, 1997.
- [SYB09] Philip Scarf, Muhammad Mat Yusof, and Mark Bilbao. A numerical study of designs for sporting contests. *European Journal of Operational Research*, 198(1):190–198, 2009.
- [Szy03] Stefan Szymanski. The economic design of sporting contests. *Journal of economic literature*, 41(4):1137–1187, 2003.
- [Él09] Árpád Enrick Élő. *The rating of chessplayers, past and present*. Ishi Press Internat, Bronx, NY, 2009.