

# How the interpretation of match statistics affects player performance

Tristan Barnett

## 1. Introduction

Commentators often refer to match statistics during a live match to make an assessment as to how the players are performing. For example, one player may have converted 80% of break point opportunities compared to 10% for the other player, and this is reflected in the scoreboard. At the completion of the match, the coaches and players are likely to use the match statistics to assess performance. Therefore, it is important to interpret match statistics correctly as misinterpretation could affect future performance. Andre Agassi lost to Marat Safin at the 2004 Australian Open semi-finals, even though Agassi won more points overall (171 for Agassi compared to 170 for Safin). This could be interpreted after the match as Agassi being “unlucky” in the sense that Agassi really performed better on the day by winning more points overall and lost as a result of two close tiebreaker sets that went in favour of Safin. The logic in the previous sentence is flawed. It is often the case that the weaker player serves more often and therefore has more opportunity to win more points. Therefore, a comparison of the percentage of points won on serve by each player should be used to interpret whether a player was “unlucky” to lose the match. As it turned out, Safin won a higher percentage of points on serve with 67.7% compared to 66.1% for Agassi.

This article begins with a summary of the statistics collected at the Australian Tennis Open. Section 3 provides calculations and analysis of statistics that can be obtained from broadcasted match statistics. Section 4 provides further statistics based on match statistics, where the information is calculated for players over many matches. Lastly, section 5 discusses how the match statistics can be used in mathematical models to optimize performance. The statistical information obtained from data collection at tennis events, broadcasted match statistics, commercial tennis software providers and mathematical models could be used by players and coaches to possibly improve performance.

## 2. Collection of statistics

Clarke and Norton (2002) discuss methods used to collect statistical data at the Australian Tennis Open. As quoted: “The recorded statistics at the Australian Open are reported back to players and coaches in the form of a two page summary forwarded after the match. This summary is much more detailed than the broadcasted television match statistics. In singles, the statistician does not enter the winner of the point, but the last player to make a play on the ball. Thus an entry of Nadal/forehand/forced error would result in the computer crediting the point to Nadal’s opponent and advancing the score appropriately. The person serving the first game of the match has to be entered, and the computer tracks the server for the remainder of the match. For singles, each serve is entered as one of in play/fault/winner/ace (lets are not recorded), the point conclusion is entered as the player to make the last play, one of forehand/backhand/overhead/volley, and one of unforced error/forced error/winner. In addition, if either or both players are at the net when the point is concluded, this is entered.”

### 3. Match statistics from an Australian Open 2004 semi-final match

Table 1 represents the match statistics for the 2004 Australian Open men's semi-final between Marat Safin and Andre Agassi. The Winning % on Serve for each player is not given directly, but can be calculated from the Receiving Points Won or using the 1<sup>st</sup> Serve %, Winning % on 1<sup>st</sup> Serve and Winning % on 2<sup>nd</sup> Serve. Calculations for the Winning % on Serve are given in Table 2 using the Receiving Points Won. There are other statistics given in Table 2 that can be obtained from calculations of the match statistics given from Table 1 and are as follows. The Winning % on 1<sup>st</sup> Serve is given in Table 1 as being conditional on the 1<sup>st</sup> Serve going in whereas the Winning % on the 2<sup>nd</sup> Serve is given in Table 1 as being unconditional on the 2<sup>nd</sup> Serve going in. The inconsistency between the 1<sup>st</sup> Serve being conditional and the 2<sup>nd</sup> Serve being unconditional should be identified when interpreting these two statistics. Calculations for the Winning % on the 1<sup>st</sup> Serve (unconditional on the 1<sup>st</sup> serve going in) are given in Table 2. Calculations for the Winning % on the 2<sup>nd</sup> Serve (conditional on the 2<sup>nd</sup> serve going in) requires calculating the 2<sup>nd</sup> Serve %, and both these statistics are given in Table 2. Calculations for the conditional and unconditional percentages for the Receiving Points Won on the 1<sup>st</sup> and 2<sup>nd</sup> Serve are given in Table 2. The outcome of a point is recorded as being an ace, double fault, winner, unforced error or forced error. It may be helpful for players and coaches to understand the subjectivity involved in how these statistics are recorded as detailed in Clarke and Norton (2002). The percentage of these different outcomes as a proportion of the the total number of points played are given in Table 2. The percentage of aces and double faults as a proportion of the total number of points served for each player can be obtained from the match statistics in Table 1. These values can then be used to obtain the percentage of errors and winners combined as a proportion of the total number of points served for each player and are given in Table 2. The percentage of net approaches and baseline play are given in Table 2 as well as the percentage of break point opportunities.

Match Statistic	Marat Safin	Andre Agassi
1 <sup>st</sup> Serve %	88 of 161 = 54%	107 of 180 = 59%
Aces	33	10
Double Faults	0	2
Unforced Errors	48	33
Winning % on 1 <sup>st</sup> Serve	68 of 88 = 77%	81 of 107 = 75%
Winning % on 2 <sup>nd</sup> Serve	41 of 73 = 56%	38 of 73 = 52%
Winners (including Service)	79	44
Break Point Conversions	2 of 7 = 28%	4 of 8 = 50%
Receiving Points Won	61 of 180 = 33%	52 of 161 = 32%
Net Approaches	13 of 24 = 54%	18 of 22 = 81%
Total Points Won	170	171
Fastest Serve	211 KPH	198 KPH
Average 1 <sup>st</sup> Serve Speed	195 KPH	176 KPH
Average 2 <sup>nd</sup> Serve Speed	148 KPH	148 KPH

Table 1: Match statistics for the 2004 Australian Open men's semi-final between Marat Safin and Andre Agassi.

Match Statistic	Marat Safin	Andre Agassi
Winning % on Serve	1-52/161 = 67.7%	1-61/180 = 66.1%
Winning % on 1 <sup>st</sup> Serve (uncond.)	$(88/161) * (68/88) = 42.2\%$	$(107/180) * (81/107) = 45.0\%$
2 <sup>nd</sup> Serve %	1-0/161 = 100%	1-2/180 = 98.9%
Winning % on 2 <sup>nd</sup> Serve (cond.)	$(41/73) / (1-0/161) = 56.2\%$	$(38/73) / (1-2/180) = 52.6\%$
Receiving Points Won on 1 <sup>st</sup> Serve (cond.)	1-81/107= 24.3%	1-68/88= 22.7%
Receiving Points Won on 1 <sup>st</sup> Serve (uncond.)	$1 - ((107/180) * (81/107)) = 55.0\%$	$1 - ((88/161) * (68/88)) = 57.8\%$
Receiving Points Won on 2 <sup>nd</sup> Serve (cond.)	$1 - ((38/73) / (1-2/180)) = 47.4\%$	$1 - ((41/73) / (1-0/161)) = 43.8\%$
Receiving Points Won on 2 <sup>nd</sup> Serve (uncond.)	1-(38/73)= 47.9%	1-(41/73)= 43.8%
Percentage of aces of total points played	33/(170+171)= 9.7%	10/(170+171)= 2.9%
Percentage of double faults of total points played	0/(170+171)=0%	2/(170+171)=0.6%
Percent. of unforced errors of total points played	48/(170+171)=14.1%	33/(170+171)=9.7%
Percentage of winners of total points played	79/(170+171)=23.2%	44/(170+171)=12.9%
Percentage of forced errors of total points played	$(170 - (33 + 48 + 79)) / (170 + 171) = 2.9\%$	$(171 - (10 + 2 + 33 + 44)) / (170 + 171) = 24.0\%$
Percentage of aces of points served	33/175=18.9%	10/180= 5.6%
Percentage of double faults of points served	0/175=0%	2/180= 1.1%
Percentage of errors and winners of points served	$1 - ((33 + 0) / 175) = 81.1\%$	$1 - ((10 + 2) / 180) = 93.3\%$
Percentage of net approaches of total points played	24/(170+171) = 7.0%	22/(170+171) =6.5%
Percentage of baseline play of total points played	$1 - (24 / (170 + 171)) = 93.0\%$	$1 - (22 / (170 + 171)) = 93.5\%$
Percentage of break point opportunities	7/(170+171)=2.1%	8/(170+171)=2.3%

Table 2: Statistics calculated from the match statistics given in Table 1.

The percentage of aces (A), double faults (DF), unforced errors (UE), winners (W) and forced errors (FE) of the total points played sum to 100%. These statistics are represented graphically in Figure 1 in the form of a pie chart. Similar graphical representations could be developed for the outcome of points as a percentage of points served by each player, and the percentage of net approaches and baseline play by each player.

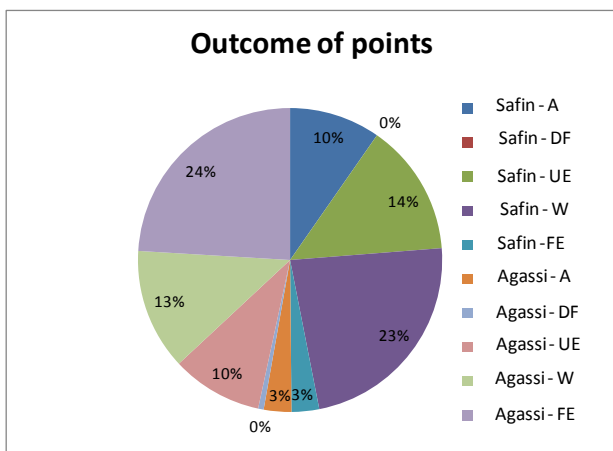


Figure 1: A graphical representation of the outcome of points from Marat Safin vs Andre Agassi match at the 2004 Australian Open

Agassi lost to Safin even though Agassi won more points overall (171 for Agassi compared to 170 for Safin). If the better performance on the day is determined by the player with the higher percentage of points won on serve, then Safin had the better performance with 109/161=67.7% compared to 119/180=66.1% for Agassi. The

reason Agassi still won more points overall is because he served more points (as is often the case for the weaker player). Agassi served 19 more points than Safin, and therefore had a greater opportunity to win more points. This needs to be taken into account when comparing two players. Safin's overall performance is 109/161 (his serve) plus 61/180 (Agassi's serve) equalling 101.6%. Agassi's overall performance is 119/180 (his serve) plus 52/161 (Safin's serve) equalling 98.4 per cent. It is now clear Safin was the better player overall. This paradox is more commonly referred to as Simpson's Paradox and can arise in statistical data when analyzing proportions. One entity can have a higher proportion or average of some desirable attribute in each of several categories but lower when the categories are combined. It might be argued that Safin was able to win the match as he won a higher percentage of the more "important" points (Morris, 1977). This is reflected by Safin winning the two tiebreaker sets, where a tiebreaker game has a high level of "importance".

There are many questions that can presently be answered using the data that is collected, but are not presented in the final statistics. Just a few examples are:

1. Safin made more unforced errors than Agassi, but were they on his service or Agassi's? Were they on break point opportunities?
2. Safin made more winners than Agassi, but were they on his service or Agassi's? Were they on break point opportunities?
3. Did faults on first serve for Agassi occur when his service was faster than his average speed?
4. Did net approach conversions for Agassi result more often from Safin's unforced errors than from Agassi's winners?

Given the large number of comparative and absolute statistics that can be relevant for an in-depth analysis of a match between two players (and for comparing that match with a 'general or average' match), it would be possible to program a computer to automatically output the 'relevant statistics' for any particular match. Such 'in-depth' output relevant to that particular match could be very useful to players, coaches and commentators.

#### **4. Comparison of career match statistics**

The OnCourt database ([www.oncourt.info](http://www.oncourt.info)) provides some match statistics since the 2003 French Open. Not all the match statistics for the ATP and WTA events are given. However the number of matches and tournaments included in the database has increased in recent years. The surfaces are categorized by grass, carpet, indoor hard, hard, acrylic and clay. Calculations of averaged match statistics over many matches can be conditioned by categories. For the Player category, calculations could be obtained for a particular player across all opponents or a particular player against a particular opponent. For the Surface/Tournament category, calculations could be obtained for a particular surface, a particular grand slam event, all grand slam events or all surfaces. For the Year category, calculations could be obtained for a particular year, particular years or across all years. The serving and receiving statistics averaged across all matches on the surfaces of grass, carpet, hard and clay are obtained for men (Table 3) and for women (Table 4), taken from the 6th Aug 2007. The results indicate that women serve a higher percentage of first serves in play compared to men for all four surfaces. However, the results indicate that men win a higher percentage of points on the first and second serve compared to the women on all four surfaces. Overall, the results indicate that men win a higher percentage of points on serve compared to the women on all four surfaces. This agrees with the results of Barnett and Pollard (2007). There is evidence to suggest that players on average win the highest percentage of

points on the first serve when on grass, followed by carpet, hard and clay. There is evidence to suggest that players on average win the highest percentage of points on serve when on grass, followed by carpet, hard and clay. This agrees with the results of Barnett and Pollard (2007) where there is a natural ordering of points won on serve from the fastest surface of grass to the slowest surface of clay.

Statistic	Grass	Carpet	Hard	Clay
Percentage of first serves in play	61.9%	61.3%	60.0%	61.5%
Percentage of points won on first serve (cond.)	74.1%	73.0%	71.0%	67.1%
Percentage of points won on second serve (uncond.)	51.8%	51.3%	50.9%	49.2%
Percentage of points won on serve	65.5%	64.6%	62.5%	60.2%
Percentage of points won on return on first serve (cond.)	25.9%	27.0%	29.0%	32.9%
Percentage of points won on return on second serve (uncond.)	48.2%	48.7%	49.1%	50.8%
Percentage of points won on return of serve	34.5%	35.4%	37.5%	39.8%
Number of matches	928	304	4319	3331

Table 3: Match statistics for men separated by court surface

Statistic	Grass	Carpet	Hard	Clay
Percentage of first serves in play	63.1%	63.4%	62.1%	63.4%
Percentage of points won on first serve (cond.)	65.4%	63.1%	62.0%	59.6%
Percentage of points won on second serve (uncond.)	46.1%	46.4%	45.3%	43.6%
Percentage of points won on serve	58.1%	57.0%	55.5%	53.5%
Percentage of points won on return on first serve (cond.)	34.6%	36.9%	38.0%	40.4%
Percentage of points won on return on second serve (uncond.)	53.9%	53.6%	54.7%	56.4%
Percentage of points won on return of serve	41.9%	43.0%	44.5%	46.5%
Number of matches	881	199	3432	2293

Table 4: Match statistics for women separated by court surface

Figure 2 presents a comparison of the average percentage of points won on the first serve (conditional on the first serve going in) between Roddick, Nadal and overall player averages across the surfaces of grass, hard and clay. It shows that both players are winning a higher percentage on the first serve on all surfaces compared with overall averages, and that Roddick is winning a higher percentage of points on the first serve compared to Nadal on all four surfaces. However, the difference between these two players is only marginal on a clay surface compared to the surfaces of grass and hard.

Figure 3 represents a comparison of the average percentage of points won on the return of the first serve (conditional on the first serve going in) between Roddick, Nadal and overall player averages across the surfaces of grass, hard and clay. It shows that Nadal is winning a higher percentage on the return of the first serve compared to Roddick on the surfaces of hard and clay and about the same on grass. However, Roddick is winning a higher percentage of points on the return of the first serve compared to overall player averages on grass, about the same on hard and less on clay.

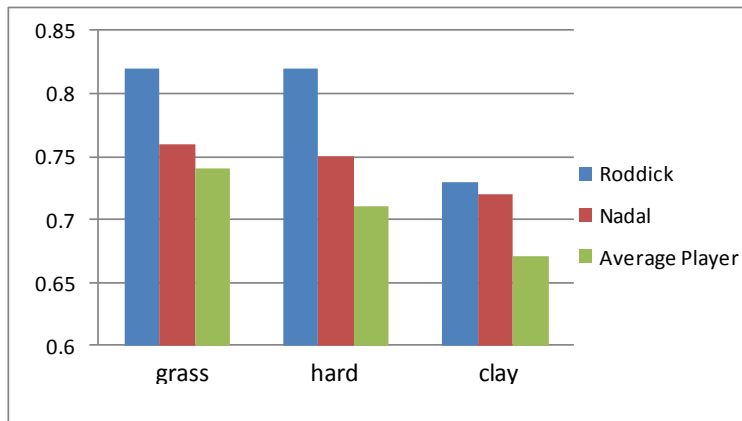


Figure 2: A graphical representation of the average percentage of points won on the first serve between Roddick, Nadal and overall player averages across the surfaces of grass, hard and clay.

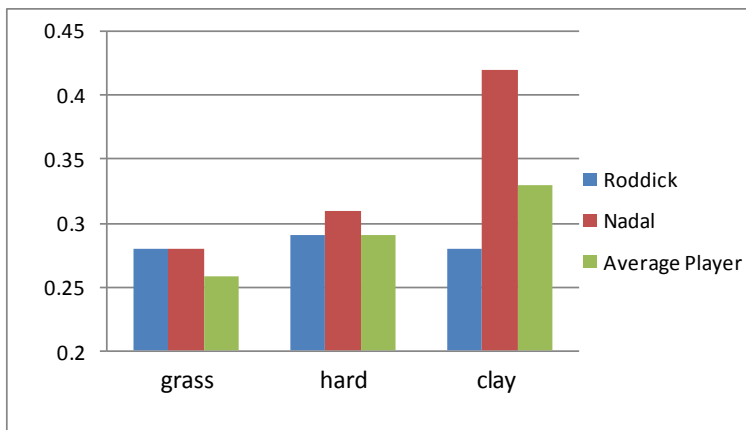


Figure 3: A graphical representation of the average percentage of points won on the return of the first serve between Roddick, Nadal and overall player averages across the surfaces of grass, hard and clay.

## 5. Mathematical modelling using match statistics

Match statistics have been applied to serving strategies to determine how much risk players should take on their serve to optimize performance (Barnett et al (2008), Pollard et al (2009)). Similar models could be developed to determine how often a player should serve-and-volley to optimize performance. This would require data on how often serve-and-volley was used and the relevant success rate. This data is currently not recorded in the official recorded statistics as detailed below.

In addition to the player's statistics given in section 4, further statistics averaged over many matches for players could include whether the outcome of a point was a forehand, backhand, volley or overhead, and the outcome of a first or second serve being an ace or a winner. These statistics could be made available on a tennis statistics software package such as OnCourt, and presented in a form that could be used by players and coaches to improve performance.

There are some statistics of interest that are not collected (Clarke and Norton, 2002) such as the number of strokes in a rally, the side (forehand or backhand) of any volleying winners or errors, the side of the opponents court to which winning shots are hit and whether a player approached the net directly after a serve (serve-and-volley) or during the rally. The latter statistic would be particularly useful for determining strategies as to how often players should serve-and-volley to optimize performance.

## Conclusions

Using a match from the 2004 Australian Open, this article has shown how the interpretation of player match statistics can affect player performance. By analyzing match statistics from commercial software providers, further statistics such as a player's average percentage of points won on serve (across many matches) can be obtained. Furthermore, these statistics could be conditioned by many factors such as the court surface. An example was given in a graphical format where a player's percentage wins on the return of the first serve across different court surfaces could be compared to another player and the average player. These types of comparisons in a numerical or graphical display could be useful to players and coaches for tactical and coaching regimes.

## References

- T. Barnett and G. Pollard, *How the Tennis Court Surface Affects Player Performance and Injuries* Med Sci Tennis (2007), 12(1): 34-37.
- T. Barnett, D. Meyer and G. Pollard, *Applying match statistics to increase serving performance*. Medicine and Science in Tennis. (2008) 13 (2), 24-27.
- S.R. Clarke and P. Norton, *Collecting Statistics at the Australian Open Tennis Championship*, In Proceedings of the 6<sup>th</sup> Mathematics and Computers in Sport conference, G. Cohen and T. Langtry eds. (2002), 105–111.
- C. Morris, *The Most Important Points in Tennis*, In Optimal Strategies in Sports, S.P. Ladany and R.E. Machol eds. (1977), Amsterdam: North– Holland, 131–140.
- G.N. Pollard, G.H. Pollard, T. Barnett and J. Zeleznikow, *Applying tennis match statistics to increase serving performance during a match in progress*. (2009), Medicine and Science in Tennis 14(3), 16-19.