International Journal of Performance Analysis of Sport
2010, 10, 131-138.

Title: The relationship between match statistics and top 100 ranking in professional men's tennis

Machar Reid[1,2], Darren McMurtrie[1] and Miguel Crespo[3]

[1] Tennis Australia, Private Bag 6060, Richmond South, 3121, Victoria, Australia
[2] University of Western Australia, Perth, Australia
[3] International Tennis Federation, London, England

Abstract
Match statistics are supplied at the majority of professional tennis tour events. The governing body of the men's tour – the Association of Tennis Professionals (ATP) – updates players' statistical profiles on a weekly basis. The relationships between the rankings and fourteen statistics describing the match performance of the top 100 male professional players in 2007 were examined to determine which statistics were most related to playing success. Partial correlations determined the strength of these associations and selected variables were entered into a stepwise regression procedure to predict professional ranking. Five variables were significant predictors of top 100 ranking while only second serve return points won and second serve points won remained in the final prediction equation, which accounted for 52% of the variance in professional ranking: predicted men's professional ranking = 548.5 + -666.6 * second serve points won + -319.9 * second serve return points won. This analysis suggests that second serve points won and second serve return points are among the most relevant statistics commonly available to ATP players.

Key words: match notation, key performance measures

1. Introduction

Contemporary sport is replete with statistical data describing athlete or team performance. With heightened professionalism and integrated sport science information and staff, the volume of these data has increased, often leading to the delineation of discriminators of performance or key performance indicators. Increasingly, these data are used by coaches to shape in-game strategy and guide interventions.

In professional tennis, match statistics are supplied at the majority of tour events. On the men's tour, the governing body – the Association of Tennis Professionals, ATP – updates players' statistical profiles on a weekly basis. The utility of these data lies in assisting those involved in player development evaluate one-off performance or to establish performance trends over time. The ability of stakeholders to 'know where to look' and to interpret the data in context is instrumental. Indeed such is the wealth of data on offer that the paralysis by analysis maxim generally reserved for description of technique could be applied to statistics. The illogic of certain interpretations of match data formed the basis of a recent paper by Bedford *et al*. (2010).

Given the potential paradox between the need for some statistical support to supplement coaching practices, but not too much so as to inhibit them, the intent of this paper is to examine how the match statistics supplied by the ATP relate to ranking success in professional tennis. Or, in more practical terms, to determine the match statistics to which coaches should attend in an effort to mould elite professional players.

2. Methods

A summary of the 2007 Matchfact information (January 1 to December 31) of the top 100 male professional players was provided by the ATP via its website, www.atptennis.com. These data are compiled and updated weekly by the ATP, in partnership with IBM. They statistically describe player performance in ATP-sanctioned tour events in the following areas: total matches won, total matches lost, total tie-breaks won, total tie-breaks lost, total aces, aces per match, total double faults, double faults per match, first serve percentage, (percentage of) first serve points won, (percentage of) second serve points won, serve games won, break points saved, (percentage of) first serve return points won, (percentage of) second serve return points won, break points won and return games won.

The structure of professional men's tennis provides players with different competitive avenues to earn ranking points (Crespo *et al*., 2003). For example, a player may attain a top 100 ranking by participating in a high number of Challenger – but not ATP Tour – events. Consequently, as data were derived from a varying number of matches for each player (n=5-85), only the variables expressed as percentage terms or capable of being normalised to ATP matches played were explored using SPSS Version 18.0. Descriptive statistics were generated on tie-breaks won per match and tie-breaks lost per match as well as data describing specific serve performance (aces per match, double faults per match, first serve percentage, first serve points won, second serve points won, break points saved), specific return performance (first serve return points won, second serve return points won, break points won) and general serve (serve games won) and return performance (return games won). The normality of each variable was examined. 'Aces per match' was not normally distributed and was corrected with a log transformation.

Between variable associations were examined using two-tailed Pearson correlation coefficients ($p<0.05$). The strength of the associations were interpreted as trivial (0.0-0.1), small (0.1-0.3), moderate (0.3-0.5), large (0.5-0.7), very large (0.7-0.9) and nearly perfect (0.9-0.99) (Hopkins, 2010). All variables with the exception of service games won and return games won were then entered in a stepwise regression procedure to predict top 100 ATP ranking. The decision to exclude service games won and return games was made because they can be considered the product of executing well on serve and return.

3. Results

Table 1 presents the descriptive statistics of the variables reported in the ATP's Match Fact records. Top 100 male players win approximately 4 out of every 5 service games and 1 out of

every 5 return games that they play. They tend to serve more aces than double faults per match and also demonstrate a greater likelihood of winning points with their first serves than when relying on their second deliveries. The probability of success in a point increases when facing second serves rather than first serves.

Table 1. Descriptive statistics of the matchplay performance of top 100, male professional players (adapted from ATP Matchfacts).

Variable	Mean	Std Deviation
Serve Games Won (%)	78.7	5.2
Return Games Won (%)	22.4	5.0
Tie-breaks won per match	0.24	0.11
Tie-breaks lost per match	0.23	0.09
Double faults per match	2.55	0.79
Aces per match	5.84	3.13
First Serve Percentage (%)	61.2	5.0
First serve points won (%)	71.6	4.2
Second serve points won (%)	51.0	2.5
Break points saved (%)	60.5	4.7
First serve return points won (%)	29.3	3.0
Second serve return points won (%)	49.3	3.2
Break points won (%)	40.2	4.5

The correlation coefficients between the different performance variables of the top 100 male players are detailed in Table 2. The following variables shared very large correlation coefficients: service games won with first serve points won (0.85) and log transformed aces per match (0.70); return games won with first serve return points won (0.88), second serve return points won (0.79) and break points won (0.77); and log transformed aces per match with first serve points won (0.82). The association between first and second serve return points won was stronger (0.53) than that between first and second serve points won (0.39). Measures of serve and return performance were significantly and positively correlated to break points saved and won respectively. Ranking was highly correlated with points won on second serve (-0.64, $p<0.05$) and shared significant, more moderate negative associations with serve games won (-0.49), return games won (-0.47), second serve return points won (-0.45) and first serve return points won (-0.35).

Table 2. Correlation matrix of the ATP Matchfact variables (* $p<0.05$).

Variable	Ln Aces/ Match	Return Games Won	Break Points Won	Points won 2nd Return	Points won 1st Return	Break Points Saved	Serve Games Won	Points won 2nd Serve	Points won 1st Serve	First Serve %	Double Faults/ Match	Tie Breaks Lost/ Match	Tie Breaks Won/ Match
Ranking	-.16	-.47*	-.28*	-.45*	-.35*	-.26*	-.49*	-.64*	-.28*	.06	.13	.24*	-.15
Tie Breaks Won/Match	.38*	-.24*	-.38*	-.27*	-.13	.38*	.45*	.29*	.46*	-.01	-.10	.19	
Tie Breaks Lost/Match	.38*	-.58*	-.46*	-.44*	-.60*	.37*	.29*	-.02	.30*	-.05	.29*		
Double Faults/Match	.26*	-.13	-.04	-.17	-.05	-.03	-.04	-.26*	.17	-.44*			
First Serve %	-.37*	.01	-.12	-.02	.02	.01	-.04	.05	-.43*				
Points won 1st Serve	.82*	-.31*	-.25*	-.24*	-.29*	.51*	.85*	.39*					
Points won 2nd Serve	.29*	.16	.02	.17	.07	.44*	.69*						
Serve Games Won	.70*	-.23*	-.29*	-.15	-.25*	.68*							
Break Points Saved	.48*	-.26*	-.28*	-.19	-.32*								
Points won 1st Return	-.41*	.88*	.57*	.53*									
Points won 2nd Return	-.40*	.79*	.59*										
Break Points Won	-.36*	.77*											
Return Games Won	-.47*												

Table 3 outlines the results of the stepwise regression procedure, where tie-breaks won per match, tie-breaks lost per match, double faults per match, first serve percentage, first serve points won, second serve points won, break points saved, first serve return points won, second serve return points won, break points won and log transformed aces per match were entered as predictor variables. The variables shown to be significant predictors of top 100 performance were, in order of percentage of variance explained, second serve points won, second serve return points won, log transformed aces per match, first serve return points won and first serve percentage. As the latter three variables only explained a further 7.3% of the variance, the percentage of second serve points won and the percentage of second serve return points were used in the following prediction equation:
Predicted men's professional ranking = 548.5 + -666.6 * Second Serve Points Won + -319.9 * Second Serve Return Points Won.

Table 3. Results of the step-wise regression predicting top 100 professional ranking

Independent Variable	Regression co-efficient	Beta Weight	R square	Adjusted R square
Second serve points won *	-666.6	-0.574	0.521	0.511
Second serve return points won *	-319.9	-0.350		
Constant	548.5			

* $p < 0.01$

4. Discussion

The volume of match notation information has increased markedly in tennis in recent years. The fact that Hawkeye generates real-time coordinate position of the ball during point play (Hawk-eye Innovations, 2007) suggests that it is a trend unlikely to abate. To this end, a parallel increase in the methods used to interpret these data has been observed (O'Donoghue, 2001, Barnett and Clarke, 2005, Takahashi *et al*., 2008, Gillet *et al*., 2009). For example, a sizeable body of work has been devoted to the link between serving strategy, score and the probability of winning (Pollard and Pollard, 2007, Barnett *et al*., 2008). The approach taken within this article, however, has been to systematically evaluate the Matchfact data presented to the competitors and coaches on the ATP tour.

Players and coaches continually pursue competitive advantages over their peers and so knowing what match statistics are most relevant; what data best accounts for playing/ranking success, can help to shape match strategy and player development. For example, the top 100 male players in 2007 won almost 80% of their service games and only 22% of their return games. Closer inspection of the data revealed the top 10 players won a mean 83 and 28 percent of serve and return games respectively, which a further insight into the key role of serve and, particularly, return performance in the upper echelons of the game. First serve percentages as well as first and second serve and return points won are consistent with those reported in past research (Barnett and Pollard, 2007; Barnett *et al*., 2008) pointing to a certain homogeneity in the serve and return performance in large samples of male professional players.

Players win more points on their first serves than second serves, while they are more likely to win points when receiving an opponent's second delivery. More specifically, where the average top 100 player wins 51% of their second serve points, the game's best players appear to win approximately 10% more points when they miss their firsts serves (e.g. Roger Federer 59%: highest in top 100). Similarly, the average top 100 ATP player wins 49% of second serve return points, yet top 10 players like David Ferrer are more effective (57%: highest in top 100). These data are supported through the moderate to large correlations observed between ranking and second serve points won and second serve return points won.

The strength of the association between players' first and second serve return points won as compared to the association between first and second serve points won is instructive. That is, the stronger association between the players' first and second serve return winning percentages could be interpreted to suggest that the practice of first and second serve returns is more transferrable (i.e. get better at one, and you're more likely to improve at other) than the practice of first and second serves. This is likely overly simplistic as it discounts the influence of all subsequent shots and point play. However, it offers practical support to the modelling of players' first and second serve strategy with a view to arriving at an optimal, individualised serving approach for each player (Barnett *et al*., 2008).

That the percentage of break points saved sits between the percentage of first and second serve points won is also noteworthy. Although some players have been shown to lift their play on certain points (Pollard *et al*. 2006), to the knowledge of the authors, no study has reported whether serve performance deteriorates (i.e. players hit more seconds than firsts) on break points. The more important points in tennis have been shown to occur on the second court (Morris, 1977) – the side of more break points – and so side of the court may represent another variable to be considered in any analysis of serve performance on break points.

Serve games won shared a stronger association with first serve points won (0.85) than second serve points won (0.69). A similar trend was observed when return games won were correlated with first serve return points won (0.88) and second serve return points won (0.79). However, as abovementioned, it was both second serve points won and second serve return points won rather than that the respective first serve/return statistics that were more strongly correlated to ranking. This was consistent with the results of the regression analysis: predicted men's professional ranking = 548.5 + -666.6 * Second Serve Points Won + -319.9 * Second Serve Return Points Won, which saw these two variables explained 52% of the variance in top 100 professional ranking. These data would certainly appear to support to the time-honoured coaching maxim that "players are only as good as their second serves". Equally however, it may need to be modified to include reference to the "second serve return" also.

This paper offers an insight in to how match statistics can be interpreted in relation to professional ranking in tennis. It is anticipated that future work will expand on this work, exploring whether the unearthed interactions between ranking and match statistics transcend other forms of the game (e.g. professional women's tennis and junior tennis) and remain consistent year-to-year.

5. Conclusion

Match statistics are readily available for players and coaches to interpret in professional men's tennis. The need to interpret these statistics is important yet scant research exists to assist with this analysis. This research article has demonstrated second serve points won and second serve return points won as significant predictors of top 100 professional ranking, accounting for 52% of its variance in the men's game. These data would appear to comprise the two of the most relevant statistics among the many available to ATP players in their match reports.

6. References

Barnett, T. and Clarke, S.R. (2005). Combining player statistics to predict outcomes of tennis matches. **IMA J Management Math**,16, 113-120.

Barnett, T., Meyer, D. and Pollard, G. (2008). Applying match statistics to increase serving performance. **Medicine and Science in Tennis**, 1, 2.

Bedford, A., Barnett, T., Pollard, G. and Pollard, G. (In press). How the interpretation of match statistics affects player performance. **Medicine and Science in Tennis.**

Crespo, M., Reid, M., Miley, D. and Atienza, F. (2003). The relationship between professional tournament structure on the national level and success in men's professional tennis. **Journal of Science and Medicine in Sport**, 6, 1, 3-13.

Gillet, E., Leroy, D., Thouvarecq, R. and Stein, J. F. (2009). A notational analysis of elite tennis serve and serve-return strategies on slow surface. **Journal of Strength and Conditioning Research***, 23*, 2, 532-539.

Hawk-Eye Innovations. (2007). *Tennis: Hawk-eye officiating system*. http://www.hawkeyeinnovations.co.uk/?page_id=1011.

Hopkins, W. (2010). A Scale of Magnitudes for Effect Statistics. http://www.sportsci.org/resource/stats/effectmag.html

Morris, C. (1977). The important points in tennis. In: **Optimal strategies in sports. Studies in Management Science and Systems** (Vol 5). (edited by S.P. Ladany and R.E. Machol), pp. 131-140. Amsterdam: North Holland Publishing Company.

O'Donoghue, P. (2001). The most important points in grand slam singles tennis. **Research Quarterly for Exercise and Sport**, *72*, 2, 125-131.

Pollard, G., Cross, R. and Meyer, D. An (2006). An analysis of ten years of four grand slam men's singles data for lack of independence of set outcomes. In: **Proceedings of the Eighth Australasian Conference on Mathematics and Computers in Sport**. (edited by J. Hammond and N. de Mestre). pp. 239-46, Queensland, Australia: MathSport (ANZIAM).

Pollard, G.H. and Pollard, G.N. (2007). Optimal risk taking on first and second serves. In: **Tennis Science & Technology 3** (edited by S. Miller, J. Capel-Davies), pp. 273-280. London: International Tennis Federation.

Takahashi, H., Wada, T., Maeda, A., Kodama, M., & Nishizono, H. (2008). An analysis of the time duration of ground strokes in grand slam men's singles using the computerised scorebook for tennis. **International Journal of Performance Analysis in Sport**, 8, 3, 96-103.