A simple new method for identifying performance characteristics associated with success in elite tennis

FITZPATRICK, Anna, STONE, Joseph <http://orcid.org/0000-0002-9861-4443>, CHOPPIN, Simon <http://orcid.org/0000-0003-2111-7710> and KELLEY, John <http://orcid.org/0000-0001-5000-1763>

Article type: Original Research Article

Corresponding Author:

Anna Fitzpatrick, S001 Chestnut Court, Sheffield Hallam University, Collegiate Crescent, Sheffield, S10 2BP, UK. Tel: +44 114 225 2355; E-mail: anna.fitzpatrick@shu.ac.uk.

Article Title: A simple new method for identifying performance characteristics associated with success in elite tennis.

Authors

Anna Fitzpatrick.[1] Joseph A. Stone.[2] Simon Choppin.[1] John Kelley.[1]

[1]Centre for Sports Engineering Research, Sheffield Hallam University, UK

[2]Academy of Sport and Physical Activity, Sheffield Hallam University, UK

Abstract

Performance analysis and identifying performance characteristics associated with success are of great importance to players and coaches in any sport. However, while large amounts of data are available within elite tennis, very few players employ an analyst or attempt to exploit the data to enhance their performance; this is partly attributable to the considerable time and complex techniques required to interpret these large datasets. Using data from the 2016 and 2017 French Open tournaments, we tested the agreement between the results of a simple new method for identifying important performance characteristics (the Percentage of matches in which the Winner Outscored the Loser, PWOL) and the results of two standard statistical methods, to establish the validity of the simple method. Spearman's rank-order correlations between the results of the three methods demonstrated excellent agreement, with all methods identifying the same three performance characteristics (*points won of 0-4 rally length*, *baseline points won* and *first serve points won*) as strongly associated with success. Consequently, we propose that the PWOL method is valid for identifying performance characteristics associated with success in tennis, and is therefore a suitable alternative to more complex statistical methods, as it is simpler to calculate, interpret and contextualise.

Keywords

Elite tennis strategy; clay courts; match statistics; successful performance; winning in tennis

Introduction

Performance analysis is of great applied importance to players and coaches in any sport,[1] enabling identification of strengths and areas for improvement,[2] assessing technical and tactical effectiveness,[3] opposition analysis,[1] guiding the development of athletes' training programmes,[4,5] injury rehabilitation,[6] identifying successful patterns of play[7,8] and predicting future match outcomes.[9] However, while it is a valued discipline and well-established tool in sports such as soccer[10] and cycling,[11] the comparative progress of performance analysis within tennis has been slow.[12] In a 2012 poll, tennis was ranked as the second least progressive sport with regards to its use of performance data.[13] Since then, tennis performance analysis research has advanced, however, with recent studies examining the prediction accuracy of different types of tennis match forecasting models,[9] comparing the physical demands and performance characteristics of professional tennis to those of the junior game,[14] and developing a comprehensive tennis shot taxonomy based on spatiotemporal data.[12]

To continue the development of performance analysis research in tennis, theoretical match investigations, whereby data from multiple matches are used to identify the typical performance characteristics of a sport,[15] could be employed. Theoretical match investigations are helpful for identifying performance characteristics associated with success, monitoring the characteristics of a sport over time, on different playing surfaces and for varying playing positions, as well as investigating the influence of rule changes. For example, a theoretical match investigation in soccer showed that the number of passes and pass success rates have increased over time, but that the magnitudes of these increases depend on playing position; thus, providing benchmark requirements for current elite players.[16] Similarly, Murray et al.[17] found that recent rule changes in elite squash (the new 11 point-per-rally scoring system and reduced tin height) decreased mean match duration and reduced the time players have to perform strokes. Murray et al.[17] also identified that more attacking strategies were adopted by players after these rule changes were implemented. Consequently, implications for training and conditioning had to be reconsidered, to ensure that players' training behaviours were representative of their match-play behaviours.[18]

Published investigations of elite tennis match-play have demonstrated that the characteristics of performance differ depending on court surface and sex.[5] Reid et al.[5] revealed several sex-based differences in match-play characteristics at the Australian Open (i.e. on hard court), and subsequently endorsed sex-specific training designs. In this context, such investigations facilitate a better understanding of the way tennis is played, allowing coaches to better prepare their players for matches.[19] In turn, the identification of performance characteristics associated with success on different surfaces enables the periodisation of training according to court surface, whereby a player's tournament schedule is built around sub-seasons (e.g. the clay court season, the grass court season), which are characterised by surface-specific training methods.[19] For example, if winning a higher percentage of net-points is found to be more closely associated with success on grass courts than on any other surface, this should be reflected in training sessions, with approach shots and net-play afforded more practice time around the grass court season. The serve is generally considered to be the most important stroke in tennis[20,21] and several studies have described the serve and serve-return as key factors to overall success,[22,23] but few attempts have been made to objectively identify the performance characteristics most strongly associated with success.

With the development of increasingly sophisticated methods for monitoring and recording aspects of tennis match-play performance, data collection has become more prevalent.[20] Consequently, huge amounts of data are readily (and freely) available to players. Typically, however, complex data processing and analysis techniques are required to uncover more useful information,[24] particularly when the data were recorded using motion tracking systems such as Hawk-Eye (Hawk-Eye Innovations Ltd., Basingstoke, UK). For example, a recent study analysing "big Hawk-Eye data" featured several stages of data cleaning and further processing before a magnus linear model was produced to examine the relationship between serve ball trajectory and winning-point probabilities.[20] Other studies have also used point-based probabilistic modelling[25,26] and "common opponent models"[27] to predict match outcomes. Even analysing and interpreting relatively simple sports data, such as notational analysis data, requires time and a degree of statistical knowledge. Notational analysis-based studies aiming to identify performance characteristics associated with success can incorporate data from 100+ matches and have employed methods including Pearson's correlation coefficients (with additional log transformations) and stepwise regression procedures (tennis[28]), Student's *t*-tests (tennis,[29] basketball[30]) discriminant analysis (rugby[31]) and Kruskal-Wallis H (soccer[32]).

It has become commonplace to employ full time performance analysts in many sports,[33] but analysts are not cheap, so even at tennis' elite level, it is uncommon for players to work with an analyst or attempt to exploit the vast amounts of data available to them.[24] Instead, coaches (before and after matches) and players (during matches) typically attempt to fulfil the role of analyst themselves, but this match-by-match approach lacks objectivity and risks neglecting long-term performance development, in favour of identifying short-term solutions. Taking this and the complex analysis methods typically required into account, coaches are not likely to have the time[4] nor often the desire[24] to analyse the large quantities of data, or potentially the expertise to transform these large datasets into meaningful interpretations with respect to tennis.[34] Therefore, a simple method of identifying performance characteristics associated with success, that is easy to understand and contextualise, may help make performance analysis more accessible and user-friendly for players and coaches. Before a method can be considered appropriate for use within elite tennis, however, it must first be compared to existing methods to assess its validity.

The aim of this study was to establish the validity of a simple new method for identifying performance characteristics associated with success in tennis, by testing the agreement between the results of the simple method and results of two standard statistical methods; 1) paired *t*-tests, as used by O'Donoghue[29] and Scanlan et al.[30] to identify match-play characteristics associated with success in tennis and basketball, respectively, and 2) point-biserial correlations, recently used by Cowden,[35] to investigate the association between mental toughness and match outcome in tennis, and Scanlan et al.[30] to assess the association between basketball match-play characteristics and match outcome. Analysis was undertaken using sample data from men's and women's elite tennis match-play at the French Open.

Method

Matches

With institutional ethics approval, performance characteristics for the 2016 and 2017 French Open men's (n = 244) and women's (n = 250) singles matches were obtained from the Roland Garros website.[36] All performance characteristics available on the website were included. Incomplete matches (i.e. walkovers, retirements and defaults) were excluded from the study.

Performance characteristics

The following performance characteristics were collected for both players in each match:

- number of aces, number of double faults
- number of first serves in
- average (i.e. mean) first serve speed*
- number of first serve points won, number of second serve points won
- number of first serve-return points won, number of second serve-return points won
- number of baseline points won, number of net points won
- number of break points won
- number of winners, number of forced errors, number of unforced errors
- number of points won of 0-4, 5-8 and 9+ rally length, respectively*.

* Collected only for those matches where a serve speed radar was available.

Data were classified by match outcome (i.e. winning player or losing player) and normalised using the equations in Table 1, before being reduced to mean values (± *sd*).

Table 1. Normalised performance characteristic equations, derived from O'Donoghue and Ingram[37] and O'Donoghue.[38]

Performance characteristic	Equation
Aces (%)	Number of aces/number of serves performed x 100
Double faults (%)	Number of double faults/number of points served x 100
Successful first serves (%)	Number of first serves in/number of first serves attempted x 100
First serve points won (%)	Number of first serve points won/number of first serve points played x 100
First serve-return points won (%)	Number of first serve-return points won/number of first serve-return points played x 100
Second serve points won (%)	Number of second serve points won/number of second serve points played x 100

Second serve-return points won (%)	Number of second serve-return points won/number of second serve-return points played x
Break points won (%)	Number of break points won as returner/number of break points played as returner x 100
Net points won (%)	Number of net points won/number of net points played x 100
Baseline points won (%)	Number of baseline points won/number of baseline points played x 100
Winners (%)	Number of winners/number of rally points played x 100
Forced errors (%)	Number of forced errors/number of rally points played x 100
Unforced errors (%)	Number of unforced errors/number of rally points played x 100
Points won of 0-4 rally length (%)	Number of points won of 0-4 rally length/number of points played of 0-4 rally length x 10
Points won of 5-8 rally length (%)	Number of points won of 5-8 rally length/number of points played of 5-8 rally length x 10
Points won of 9+ rally length (%)	Number of points won of 9+ rally length/number of points played of 9+ rally length x 100

Data analysis

Statistical correlation-based method: Normalised data were imported into in SPSS (v23.0, SPSS Inc, USA). Point-biserial correlations between match outcome and each performance characteristic were calculated, to identify which characteristics were associated with match outcome, for both sexes.

Statistical paired t-test method: For each performance characteristic, a paired *t*-test was used to compare winning and losing players' data, for each sex, with the simplifying assumption of normality of the winner-loser differences. The *t* values were used to identify the performance characteristics that best distinguished between winning and losing players.[29]

Proposed method: For each performance characteristic, the winning player's performance was compared to that of their opponent (i.e. the losing player), to identify which player 'outscored' the other. Then, the *Percentage of matches in which the Winner Outscored the Loser* (PWOL) was calculated for each performance characteristic. For example, if the winning player hit *more* aces than (i.e. outscored) the losing player in 200 out of 250 matches, the PWOL for aces would be 80.0%. Similarly, if the winning player hit *more* unforced errors than (i.e. outscored) the losing player in 100 out of 250 matches, the PWOL for unforced errors would be 40%; this would mean that the losing player hit *more* unforced errors than the winning player in 60% of matches. To the authors' knowledge, the method has not previously been applied within sports performance analysis.

A PWOL value of 50% for a particular performance characteristic means that players who outscored their opponent won the match in 50% of cases, indicating no association with match outcome (i.e. success). As the PWOL value increases towards 100%, this indicates a stronger positive association with match outcome or success. Correspondingly, as the PWOL value decreases towards 0%, this indicates a stronger negative association with match outcome (i.e. a stronger association with losing). Therefore, performance characteristics with either a high PWOL value or a low PWOL value are considered important in terms of winning, whereas those with PWOL values close to 50% are considered less important. PWOL values are simple to calculate, with no need for statistical software packages, such as SPSS. Furthermore, users do not require a comprehensive understanding of data analysis techniques to apply the method or interpret the results, so it may be more suitable for coaches and other sports practitioners.

The results of each of the three methods were used to indicate the relative importance of each performance characteristic. To assess the agreement between the results of the methods (i.e. establish the validity of the PWOL method), pairwise comparisons between the PWOL values, *t* values and point-biserial correlation coefficients were performed using Spearman's rank-order correlations.

Results

Table 2 (men) and Table 3 (women) show the mean and standard deviation of the performance characteristics for winning and losing players. Table 2 and 3 also show the results of each statistical method; point-biserial correlations between each performance characteristic and match outcome (r_{pb}), *t* values from paired *t*-tests comparing winning and losing players (*t*) and the Percentage of matches in which the Winner Outscored the Loser (PWOL).

Table 2. Men's performance characteristics (presented as mean ± *sd*), point-biserial correlations with match outcome, *t* values and PWOL values (± 95% confidence intervals); sorted by r_{pb}

Performance characteristic	Winning players	Losing players	r_{pb}	t_{dof}	PWOL
Points won of 0-4 rally length (%)	55.2 ± 5.0%	44.8 ± 5.0%	0.72	44.47_{180}	89.0 (± 4.6%)
Baseline points won (%)	53.1 ± 6.5%	42.2 ± 6.3%	0.65	48.43_{243}	82.4 (± 4.8%)

First serve points won (%)	74.7 ± 8.1%	65.1 ± 8.3%	0.51	45.64_{243}	85.2 (± 4.5%)
First serve-return points won (%)+	34.9 ± 8.3%	25.3 ± 8.1%	0.51	45.64_{243}	85.2 (± 4.5%)
Points won of 5-8 rally length (%)	54.4 ± 8.1%	45.6 ± 8.1%	0.48	29.27_{180}	65.2 (± 6.9%)
Second serve points won (%)	56.5 ± 9.4%	46.2 ± 9.5%	0.47	42.18_{243}	76.6 (± 5.3%)
Second serve-return points won (%)+	53.8 ± 9.5%	43.5 ± 9.4%	0.47	42.18_{243}	76.6 (± 5.3%)
Points won of 9+ rally length (%)	55.6 ± 14.7%	44.4 ± 14.7%	0.35	27.57_{180}	65.7 (± 6.9%)
Winners (%)	18.1 ± 4.7%	15.1 ± 4.4%	0.31	16.67_{243}	63.9 (± 6.0%)
Break points won (%)	46.1 ± 16.0%	33.7 ± 22.5%	0.30	35.83_{230}	70.9 (± 5.9%)
Net points won (%)	67.6 ± 13.1%	61.4 ± 13.4%	0.23	22.22_{243}	61.9 (± 6.1%)
Aces (%)	5.1 ± 4.0%	3.7 ± 2.8%	0.21	10.46_{243}	59.4 (± 6.2%)
Successful first serves (%)	61.7 ± 7.4%	60.3 ± 7.2%	0.09	6.71_{243}	55.7 (± 6.2%)
Average first serve speed (km/h)	181.8 ± 9.7 km/h	180.9 ± 10.7 km/h	0.04	3.21_{180}	50.8 (± 7.3%)
Double faults (%)	3.1 ± 2.1%	3.4 ± 2.2%	-0.08	-3.09_{243}	43.9 (± 6.2%)
Unforced errors (%)	13.8 ± 4.5%	17.2 ± 5.0%	-0.34	-19.39_{243}	32.8 (± 5.9%)
Forced errors (%)	16.4 ± 3.3%	19.5 ± 3.9%	-0.40	-22.80_{243}	22.1 (± 5.2%)

+ A player’s first (or second) serve-return points won (%) = 100 – opponent’s first (or second) serve points won (%), hence identical associations with match outcome.

Note: Degrees of freedom differ, as some performance characteristics (e.g. first serve speed and rally lengths) are only recorded on some match courts.

Table 3. Women's performance characteristics (presented as mean ± *sd*), point-biserial correlations with match outcome, *t* values and PWOL values (± 95% confidence intervals); sorted by r_{pb}

Performance characteristic	Winning players	Losing players	r_{pb}	t_{dof}	PWOL
Baseline points won (%)	54.1 ± 7.0%	42.6 ± 7.0%	0.64	49.58_{249}	83.9 (± 4.6%)
Points won of 0-4 rally length (%)	55.5 ± 6.8%	44.5 ± 6.8%	0.63	39.57_{172}	84.5 (± 5.4%)
First serve points won (%)	67.5 ± 9.7%	56.1 ± 10.1%	0.50	46.83_{249}	82.8 (± 4.7%)
First serve-return points won (%)+	43.9 ± 10.1%	32.5 ± 9.7%	0.50	46.83_{249}	82.8 (± 4.7%)
Points won of 5-8 rally length (%)	54.8 ± 10.8%	45.2 ± 10.8%	0.41	27.25_{172}	67.8 (± 7.0%)
Second serve points won (%)	50.2 ± 11.0%	40.5 ± 10.9%	0.41	38.78_{249}	76.4 (± 5.3%)
Second serve-return points won (%)+	59.5 ± 10.9%	49.8 ± 11.0%	0.41	38.78_{249}	76.4 (± 5.3%)
Winners (%)	18.0 ± 6.1%	13.9 ± 5.2%	0.34	20.47_{249}	68.0 (± 5.8%)
Points won of 9+ rally length (%)	54.3 ± 16.4%	45.7 ± 16.4%	0.25	19.57_{171}	55.7 (± 7.4%)
Break points won (%)	53.7 ± 17.3%	43.0 ± 24.9%	0.24	29.64_{241}	66.0 (± 6.0%)
Aces (%)	2.7 ± 2.6%	1.7 ± 1.8%	0.23	9.47_{249}	57.2 (± 6.1%)
Net points won (%)	66.8 ± 17.5%	59.9 ± 19.1%	0.18	20.19_{249}	53.6 (± 6.2%)
Successful first serves (%)	64.5 ± 8.1%	62.8 ± 8.7%	0.10	7.83_{249}	58.0 (± 6.1%)
Average first serve speed (km/h)	155.3 ± 10.5 km/h	154.7 ± 9.9 km/h	0.03	2.13_{172}	51.7 (± 7.4%)
Double faults (%)	4.0 ± 3.2%	4.5 ± 3.4%	-0.08	-3.82_{249}	45.6 (± 6.2%)
Forced errors (%)	14.7 ± 4.1%	17.5 ± 4.0%	-0.26	-12.27_{249}	34.4 (± 5.9%)
Unforced errors (%)	15.7 ± 5.1%	19.6 ± 6.3%	-0.32	-19.81_{249}	33.6 (± 5.9%)

+ A player’s first (or second) serve-return points won (%) = 100 – opponent’s first (or second) serve points won (%), hence identical associations with match outcome.

Note: Degrees of freedom differ, as some performance characteristics (e.g. first serve speed and rally lengths) are only recorded on some match courts.

Point-biserial correlations, *t* values and PWOL values all identified *points won of 0-4 rally length, baseline points won* and *first serve points won* as the performance characteristics most strongly associated with match outcome (i.e. success), for both sexes. *Forced errors* and *unforced errors* were the performance characteristics most negatively associated with match outcome, i.e. associated with losing. Serve-related performance characteristics including *aces*, *double faults*, *successful first serves* and *average first serve speed* were least associated with match outcome.

Agreement between methods

Table 4 displays the Spearman's rank-order correlation coefficients for point-biserial correlation coefficients, *t* values and PWOL values for men's and women's data, respectively.

Table 4. Spearman's rank-order correlation coefficients for point-biserial correlation coefficients, *t* values and PWOL values, for both sexes.

Pairwise comparison	Men	Women
r_{pb} and t	0.95	0.95
r_{pb} and PWOL	0.96	0.95
t and PWOL	0.98	0.94

Note: all correlations were significant at p < 0.001.

All Spearman's rank-order correlation coefficients demonstrated excellent agreement between the results of the different methods.[39]

Discussion

The aim of this paper was to establish the validity of a simple new method for identifying performance characteristics associated with success in tennis, by testing the agreement between the results of the simple method and two previously used statistical methods. Spearman's rank-order correlations between results of the new PWOL method and those of two statistical methods (paired *t*-tests and point-biserial correlations) demonstrated excellent agreement (r_s between 0.94 and 0.98),[39] for men's and women's datasets. These high correlations show that the PWOL method can identify performance characteristics associated with success as effectively as more complex statistical methods; and is therefore a valid method. Accordingly, we suggest elite coaches consider employing the PWOL method, as it is simpler to calculate, interpret and contextualise than standard statistical methods.

Points won of 0-4 rally length demonstrated the highest PWOL value (89.0%) for men; this corresponded to a point-biserial correlation with match outcome of 0.722 and a *t* value of 44.47. In simple terms, a PWOL value of 89.0% means that players who won more points of 0-4 rally length than their opponent won the match in almost 9 out of 10 cases. Similarly, in the women's event, *points won of 0-4 rally length* demonstrated the highest PWOL value (84.5%), a point-biserial correlation with match outcome of 0.633 and *t* value of 39.57. So, players who won more points of 0-4 rally length than their opponent won the match in 84.5% of cases. *Baseline points won* and *first serve points won* were also associated with success for men and women, exhibiting PWOL values of above 80%. These values corresponded to moderate point-biserial correlation coefficients and *t* values. Collectively, these results imply that *points won of 0-4 rally length*, *baseline points won* and *first serve points won* may be considered the three most important performance characteristics at the French Open, with superior performance in these areas closely associated with success for men and women. These three performance characteristics all pertain to 'points won' and often comprise a large proportion of the total points played within a match, so in a tennis context, it may be considered unsurprising that they demonstrated associations with success.

For both sexes, *forced errors* and *unforced errors* were the performance characteristics with the lowest PWOL values (between 22.1% and 34.4%), corresponding to negative point-biserial correlation coefficients and negative *t* values. For example, male players who hit more forced errors than their opponent won the match in only 22.1% of cases and female players who hit more forced errors than their opponent won the match in 34.4% of cases. These results indicate a negative association with match outcome, i.e. an association with losing, and show that hitting fewer forced and unforced errors is advantageous in terms of winning on clay.

Serve-related performance characteristics including *aces, double faults, average first serve speed* and *successful first serves* were among the least associated with match outcome for both sexes, exhibiting PWOL values between 43.9% and 59.4%. The PWOL values for *aces* and *successful first serve percentage* were above 50%, demonstrating weak associations with success, whereas the PWOL value for *double faults* was below 50%, demonstrating a weak association with losing. Together, results of the serve-related performance characteristics indicate that a player's serve performance is not closely linked with success on clay courts and may therefore be considered less important in terms of winning.

The PWOL method has several advantages compared to standard statistical methods. Calculating the PWOL value of a performance characteristic is straight forward and the process does not require a comprehensive understanding of statistical methods or a software package. A pen, notepad and basic mobile phone calculator (if necessary) are sufficient; to tally the number of matches in which the winner outscored the loser on a relevant performance characteristic and calculate the tallied number of matches as a percentage of total number of matches. Furthermore, the PWOL value of a single performance characteristic can be interpreted in isolation, whereas the result of a paired *t*-test (for example) is more difficult to contextualise, as a series of values are required to gauge relative importance.

We propose that analysts adopt the PWOL method to investigate important performance characteristics, as its ease of interpretation means that it can be effectively fed back to coaches directly.[40] Clearly, representative tournaments should be chosen for analysis, i.e. coaches of elite female players should only consider elite level women's tournaments in their sample, rather than lower level or men's events, and the court surface should also be considered. The PWOL confidence intervals displayed in Table 2 and Table 3 demonstrated that a smaller *n* resulted in a larger confidence interval. In terms of the sample size required to draw valid conclusions using the PWOL method, our results suggest that analysing data from 200 matches will give confidence intervals of approximately ± 5% for PWOL values of 80 - 90%. As such, Grand Slam events are ideal for this type of analysis, as performance data from over 100 men's and women's matches are readily and freely available to all players, coaches and analysts at every Grand Slam. Further work is needed to establish whether PWOL can be used on a smaller scale (i.e. using data from fewer matches); if so, the method is such that coaches could adopt it themselves if they so wished. The process would also be efficient with an appropriately designed spreadsheet template.

In addition to its simple calculation and interpretation in a coaching context, the PWOL method offers a further benefit compared to standard statistical methods. Paired *t*-tests and point-biserial correlations both consider the magnitude of the differences between winning and losing players' values on a particular performance characteristic, when establishing that characteristic's association with match outcome. In contrast, the PWOL method simply acknowledges the fact that the winning player's value was either higher or lower than the losing player's value, irrespective of the magnitude of the difference. For this reason, the PWOL is more robust than paired *t*-tests and point-biserial correlations in the case of extreme values or outliers. For example, in this study, men's *average first serve speed* exhibited a *t* value of 3.21, which indicates a *significant* difference between winning and losing players' values. However, if we were to remove from the dataset the *two* matches in which the lowest average first serve speeds occurred, the *t* value would decrease to a *non-significant* 1.76, which demonstrates the strong influence of extreme values in a paired *t*-test. While the PWOL may be more robust to outliers (by disregarding the magnitude of the difference between two players' values), it should be noted that in some instances, the magnitude of the difference can offer relevant information.

As we are proposing the PWOL method as a more user-friendly alternative for coaches, statistical significance was not incorporated in the results here, as it is unlikely to be relevant to coaches. It is worth highlighting, however, that if performance analysts or other users wish to calculate statistical significance for PWOL values, this can be done using a binomial distribution, with parameters n and *p*, where n is the sample size and *p* is the probability of, in this case, the winning player outscoring the losing player in a single match.

In previous studies attempting to identify performance characteristics associated with success in tennis,[28,29] methodological differences such as the performance characteristics included in the analyses and their respective calculations appear to have contributed to inconsistent results. For example, rally length statistics have not previously been included, but here, *points won of 0-4 rally length* was most closely associated with success. This study has shown that the PWOL method, correlation-based method and paired *t*-test method demonstrate excellent pairwise agreement. Therefore, if performance characteristics, operational definitions and calculations are consistent between studies in future, we can assume that any differences in the characteristics identified as strongly associated with success are attributable to differences in the context of the performances analysed (e.g. court surface, sex, time etc.), and not differences in the data analysis methods. The list of performance characteristics presented here comprises a more comprehensive selection than those in previous studies. Future research aiming to identify performance characteristics associated with success in tennis should endeavour to incorporate a comprehensive range of characteristics, including rally length statistics. Accordingly, a standardised list of tennis strategy performance characteristics, calculations and definitions would be beneficial. In this context, the PWOL method can be used to help prevent 'paralysis by analysis', in that a performance analyst or future studies analysing the sport (e.g. on different surfaces) may employ the method to narrow down or 'filter' a full list of performance characteristics and highlight those most strongly associated with success on a particular surface. Subsequently, the filtered list would provide a concise summary of relevant areas for coaches to focus on during training sessions. Additionally, with more sophisticated methods of performance tracking (e.g. Hawkeye Innovations Ltd., Basingstoke, UK) now commonplace in elite tennis,[5] performance characteristics not typically reported (e.g. distance travelled, average and maximum movement speed) could become more accessible. Future work should investigate the potential use of the PWOL method for new (and currently under-used) performance characteristics.

In conclusion, results of the PWOL method demonstrated excellent agreement with results of the point-biserial correlation and paired *t*-test methods. As such, this study has shown that the PWOL method is able to successfully identify performance characteristics associated with success in elite tennis. The method is simple to calculate and does not require statistical software or the expertise of a performance analyst to understand the results; this may encourage players and coaches to begin to engage with performance analysis as a discipline and recognise its potential benefits. Furthermore, the PWOL method is robust in the case of extreme values. We therefore propose the PWOL method as a suitable, more user-friendly alternative to common statistical methods of data analysis in elite tennis.

Acknowledgements

The author(s) would like to acknowledge and thank Roland Garros and Brain Game Tennis for facilitating this study.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Groom R, Cushion CJ and Nelson LJ. The delivery of video-based performance analysis by England youth soccer coaches: towards a grounded theory. *J Appl Sport Psychol* 2011; 23: 16-32.

2. Smith DJ, Norris SR and Hogg JM. Performance evaluation of swimmers. *Sports Med* 2002; 32(9): 539-554.

3. Passos P, Araujo D and Volossovitch A. *Performance Analysis in Team Sports*. Routledge: London, 2017.

4. Reilly T, Morris T and Whyte G. The specificity of training prescription and physiological assessment: a review. *J Sport Sci* 2009; 27: 575-589.

5. Reid M, Morgan S and Whiteside D. Matchplay characteristics of Grand Slam tennis: implications for training and conditioning. *J Sport Sci* 2016; 34(19): 1791-1798.

6. Blobel T, Pfab F, Wanner P, Haser C and Lames M. Healthy Reference Patterns (HRP) supporting prevention and rehabilitation process in professional football. In: World conference on science and soccer, Rennes, France, 31 May-2 June 2017, pp. 183-184.

7. Hughes M. Computerised notation of racket sports. In: Reilly T, Hughes M and Lees A (eds) *Science and Racket Sports*. London: E & FN Spon, 1995, pp. 249-256.

8. Unierzyski P and Wieczorek A. Comparison of tactical solutions and game patterns in the finals of two grand slam tournaments in tennis. In: Lees A, Kahn J-F and Maynard I (eds) *Science and Racket Sports III*. Oxon: Routledge, 2004, pp. 169-174.

9. Kovalchik SA. Searching for the GOAT of tennis win prediction. *J Quant Anal Sports* 2016; 12(3): 127-138.

10. Thompson KG. *Pacing Individual Strategies for Optimal Performance*. Champaign: Human Kinetics, 2015.

11. Trimble L. et al. *Sport in the UK*. Exeter: Learning Matters, 2010.

12. Kovalchik SA and Reid M. A shot taxonomy in the era of tracking data in professional tennis. *J Sport Sci* 2018; 36(18): 2096-2104.

13. Martin T, Annacone P, O'Shannessy C and Stein M. Tennis Analytics. http://www.sloansportsconference.com/content/tennis-analytics/. (2012, accessed 18 July 2018).

14. Kovalchik SA and Reid M. Comparing matchplay characteristics and physical demands of junior and professional tennis athletes in the era of big data. *J Sport Sci Med* 2017; 16: 489-497.

15. O'Donoghue P. Match analysis in racket sports. In: Lees A, Kahn J-F and Maynard I (eds) *Science and Racket Sports III*. Oxon: Routledge, 2004, pp.155-162.

16. Bush M, Barnes C, Archer DT, Hogg B and Bradley PS. Evolution of match performance parameters for various playing positions in the English Premier League. *Hum Mov Sci* 2015; 39: 1-11.

17. Murray S, James N, Hughes MD, Pers J, Mandeljc R and Vuckovic G. Effects of rule changes on physical demands and shot characteristics of elite-standard men's squash and implications for training. *J Sport Sci* 2017; 34(23): 2170-2174.

18. Pinder RA, Davids K, Renshaw I and Araujo, D. Representative learning design and functionality of research and practice in sport. *J Sport Exerc Psychol* 2011; 33: 146-155.

19. Over S and O'Donoghue P. What's the point – tennis analysis and why. *ITF Coach Sport Sci Rev* 2008; 15(45): 19-21.

20. Mecheri S, Rioult F, Mantel B, Kauffmann F and Benguigui N. The serve impact in tennis: first large-scale study of big Hawk-Eye data. *The ASA Data Science Journal* 2016; 9(5): 310-325.

21. Bollettieri N. *Nick Bollettieri's Tennis Handbook*. 2nd ed. Champaign: Human Kinetics, 2015.

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

23. Klaus A, Bradshaw R, Young W, O'Brien B and Zois J. Success in national level junior tennis: tactical perspectives. *Int J Sports Sci Coach* 2017; 12(5): 618-622.

24. Briggs S. Why tennis is ready to embrace its inner geek. *The Telegraph,* https://www.telegraph.co.uk/tennis/2018/03/27/tennis-ready-embrace-inner-geek/ (2018, accessed 27 March 2018).

25. Barnett T and Clarke SR. Combining player statistics to predict outcomes of tennis matches. *IMA J Manag Math* 2005; 16(2): 113-120.

26. Spanias D and Knottenbelt WJ. Predicting the outcomes of tennis matches using a low-level point model. *IMA J Manag Math* 2012; 24(3): 311-320.

27. Knottenbelt WJ, Spanias D and Madurska AM. A common-opponent stochastic model for predicting the outcome of professional tennis matches. *Comput Math Appl* 2012; 64(12): 3820-3827.

28. Reid M, McMurtrie D and Crespo M. The relationship between match statistics and top 100 ranking in professional men's tennis. *Int J Perform Anal Sport* 2010; 10: 131-138.

29. O'Donoghue P. Performance models of ladies' and men's singles tennis at the Australian Open. *Int J Perform Anal Sport* 2002; 2(1): 73-84.

30. Scanlan TA, Teramoto M, Delforce M and Dalbo JV. Do better things come in smaller packages? Reducing game duration slows game pace and alters statistics associated with winning in basketball. *Int J Perform Anal Sport* 2017; 16: 157-170.

31. Ortega E, Villarejo D and Palao JM. Differences in game statistics between winning and losing rugby teams in the Six Nations tournament. *J Sport Sci Med* 2009; 8(4): 523-527.

32. Lago-Penas C, Lago-Ballesteros J, Dellal A and Gomez M. Game-related statistics that discriminated winning and losing teams from the Spanish soccer league. *J Sport Sci Med* 2010; 9(2): 288-293.

33. Hughes M and Bartlett R. What is performance analysis? In: Hughes M and Franks IM (eds) *Essentials of Performance Analysis in Sport*. 2nd ed. London: Routledge, 2015, pp. 18-28.

34. Hughes M and Franks IM. *Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport*. 2nd ed. London: Routledge, 2004.

35. Cowden RG. Competitive performance correlates of mental toughness in tennis: a preliminary analysis. *Percept Mot Skills* 2016; 123: 341-360.

36. Roland Garros. Completed Matches, www.rolandgarros.com (2017, accessed 30 June 2017).

37. O'Donoghue P and Ingram B. A notational analysis of elite tennis strategy. *J Sport Sci* 2001; 19(2): 107-115.

38. O'Donoghue P. Normative profiles of sports performance. *Int J Perform Anal Sport* 2005; 5: 104-119.

39. Hahs-Vaughn DL and Lomax RG. *An Introduction to Statistical Concepts*. 3rd ed. New York: Routledge, 2012.

40. Sigrist R, Rauter G, Reiner R and Wolf P. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychon Bull Rev* 2013; 20: 21-53.