



Journal of Sports Sciences

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rjsp20>

Winning matches in Grand Slam men's singles: An analysis of player performance-related variables from 1991 to 2008

Shang-Min Ma^a, Chao-Chin Liu^b, Yue Tan^c & Shang-Chun Ma^d

^a National Pingtung University of Science and Technology, Recreational Sport and Health Promotion, Pingtung, Taiwan

^b MeiHo University, Recreation Sports and Health Promotion, Pingtung, Taiwan

^c National Sun Yat-Sen University, Institute of Communication Management, Kaohsiung, Taiwan

^d National Cheng Kung University, Institute of Physical Education, Health & Leisure Studies, Tainan, Taiwan

Published online: 05 Mar 2013.

To cite this article: Shang-Min Ma, Chao-Chin Liu, Yue Tan & Shang-Chun Ma (2013) Winning matches in Grand Slam men's singles: An analysis of player performance-related variables from 1991 to 2008, Journal of Sports Sciences, 31:11, 1147-1155, DOI: [10.1080/02640414.2013.775472](https://doi.org/10.1080/02640414.2013.775472)

To link to this article: <http://dx.doi.org/10.1080/02640414.2013.775472>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

Winning matches in Grand Slam men's singles: An analysis of player performance-related variables from 1991 to 2008

SHANG-MIN MA¹, CHAO-CHIN LIU², YUE TAN³, & SHANG-CHUN MA⁴

¹National Pingtung University of Science and Technology, Recreational Sport and Health Promotion, Pingtung, Taiwan, ²MeiHo University, Recreation Sports and Health Promotion, Pingtung, Taiwan, ³National Sun Yat-Sen University, Institute of Communication Management, Kaohsiung, Taiwan, and ⁴National Cheng Kung University, Institute of Physical Education, Health & Leisure Studies, Tainan, Taiwan

(Accepted 8 February 2013)

Abstract

This study examines factors that lead to winning matches in men's singles Grand Slam tennis, and proposes guidelines for coaches and professional tennis players both in training and preparation for Grand Slam competitions. Using longitudinal data between 1991 and 2008 retrieved from the official website of the Association of Tennis Professionals, we analysed player performance over 9,144 matches in men's singles Grand Slam tournaments. To predict match outcome, 16 variables were classified into one of three categories: player skills and performance, player characteristics and match characteristics. The three categories were entered sequentially into a logistic regression model to predict the dependent variable: the chance of winning a men's singles Grand Slam match. The final altered model explains 79.4% of the variance (Nagelkerke's pseudo R^2) in match outcomes and correctly predicted 90.6% of cases. The importance of serving, receiving, and break points is further confirmed. The positive effect of stature diminishes when players are taller than 186 cm. We recommend more training in returning skills; to avoid overestimation of the positive impact of stature, left hand and professional experience; and that a male player begins his professional tennis career by participating in the US Open or Wimbledon.

Keywords: *Grand Slam men's singles, Association of Tennis Professionals, logistic model*

Introduction

As elite-standard tennis becomes increasingly competitive, players become motivated to improve their performance to a corresponding degree (Reid, Crespo, Lay, & Bery, 2007). The mean prize for each player in a match has risen from 47,084 USD in the 1990s (1991–1999) to 121,550 USD in the 2000s (2000–2008). Player performance has also improved for peak ball speed during the serve, aces, double faults, and returning ability (Cross & Pollard, 2009). At the completion of the 2012 Australian Open, the combined “rating points” (an indicator of player quality) of leading men's players Djokovic, Nadal, and Federer had reached an all-time peak for the sport of any three players at one time (Harris, 2012). A key question is: what are the determinants of outcomes of an elite-standard tennis match? Because of the complexity of determinants, a single answer is unlikely. However, the above metrics make it clear that tennis is played at a higher standard than it was 30 years ago. Hence, predictive factors have probably changed.

Longitudinal research therefore might identify determinants for match wins. Using longitudinal data for men's singles Grand Slam tournaments from 1991 to 2008, the aims of this study were to: (1) establish a logistic model that could identify determinants of wins in men's singles Grand Slam events (i.e., the Australian Open, the French Open, Wimbledon, and the US Open); and, (2) provide coaches and professional tennis players with recommendations for training and preparation for competition.

To examine conditions for winning matches in men's singles Grand Slam tournaments, studies have used true experimental designs (Hornery, Farrow, Mujika, & Young, 2007), quasi-experimental designs (Klaassen & Magnus, 2003), and questionnaire surveys (Scheibehenne & Brofer, 2007). These studies have provided important indicators, but fall short of producing a comprehensive answer to the research question. To build a model that controls for all the important factors necessarily demands a large sample size because the sampling error is decreased to 1% until the sample has 8800

cases. This study used a sample of 9,144 matches (including 845 players) in men's singles Grand Slam tournaments from 1991 to 2008, retrieved from the official open-access website (www.atpworldtour.com) of the Association of Tennis Professionals.

The outcome of a tennis match can be only "lose" or "win" yet tennis performance integrates physiological, biomechanical, psychological, and perceptual characteristics (Struder, Hollmann, & Duperly, 1995). Several researchers (Gould, Petlinchhoff, Simons, & Vevera, 1987; Jones, 1995; Rees, Ingledew, & Hardy, 1999; Vealey, 1994) have suggested the use of process measures to assess performance and reflect the complexity of sports. Therefore, we interpret players' performance as a mediator of a match outcome.

Four factors during matches influence tennis performance: (1) environmental conditions (Rees & Hardy, 2004); (2) match conditions (O'Donoghue & Ingram, 2001); (3) playing strategies (O'Donoghue & Brown, 2008); and, (4) personal characteristics (Loffing, Hagemann, & Strauss, 2009). This study focuses on match characteristics (such as court surface), personal characteristics (such as stature and age), and skills that can influence match outcomes.

Method

With approval from the National Science Council Project Review Board, Taiwan, R.O.C., a set of longitudinal data from 1991 to 2008 was retrieved from the official open-access website (www.atpworldtour.com) of the Association of Tennis Professionals (ATP). A manual search of the website provided data that were then entered into SPSS 17.0. For every 200 entries, the second author cross-checked 15% of the data entered to ensure accuracy. A total of 18,288 performances were recorded, and variables characterising professional men's singles players were identified. A total of 16 variables associated with those matches (including environment, match characteristics, personal traits, and playing skills) were tested simultaneously in our model. The ATP data have been examined by several studies, but most of them used only a small proportion and few variables to construct simulated models (Barnett & Clarke, 2005; Corral & Prieto-Rodriguez, 2010; Scheibehenne & Brofer, 2007). In contrast, the large and representative sample of this study ensures external validity of our model.

An example of a personal characteristic we considered is player stature. A taller player will generally have the advantage over a shorter player for serve speed (Cross & Pollard, 2009). Stature is an important physiological factor for a coach to consider when selecting future elite-standard tennis players (MacCurdy, 2006). Typically, a coach needs to

spend 10 years (or 10,000 hours) to train an elite-standard tennis player (Erricsson, Krampe, & Tesch-Romer, 1993; Helsen, Starkes, & Hodges, 1998), and therefore would be more productive investing such time and resources in training only players whose stature is optimal for winning Grand Slam matches. However, there is no consensus in the literature as to the optimum stature for an elite tennis player. Caution should be exercised when using stature as a selection criterion because predictions of adult stature from childhood stature (under 12 years old) have a mean error of ± 10 cm (Pluim, 2006). Although wrist x-rays can be used to determine skeletal age, they provide ambiguous outcomes "for very young players" (Pluim, 2006, p. 6). Corral and Prieto-Rodriguez (2010) reported no differences in stature between higher- and lower-ranked men's players in match outcomes whereas Cross and Pollard (2009) reported that match winners were "typically one or two cm taller than their opponents and serve $2\text{--}4 \text{ km} \cdot \text{hr}^{-1}$ faster on average" (p. 9). Most top-ranking tennis players in the period studied were neither unduly tall (i.e. above 190 cm) nor too short (i.e. under 180 cm), and while we doubted that there could be a linear or non-linear relationship between stature and match outcomes, stature was one of the variables assessed in this study.

Body mass is equally important as stature in match performance, but because such mass correlates with stature for this study ($r = 0.75$, $P < 0.001$), it was not included in the regression model, to avoid multicollinearity problems. High multicollinearity (the correlations between independent variables of above 0.8) produces untrustworthy b -values, limits the size of R , and makes it difficult to assess the individual importance of a predictor (Field, 2009, p. 224).

Age can also be an important factor in determining match outcomes. Corral and Prieto-Rodriguez (2010) reported that the probability of a higher-ranked player winning decreases when they compete against younger players. In particular, the younger that a higher-ranked player is, relative to the lower-ranked player, the greater the probability that the higher-ranked player will win. These findings however, did not consider the mediating effect of years of professional experience, between age and performance. Ageing results in a gradual loss of muscle function, with the size of type II muscle fibres beginning to reduce as early as age 20 (Kirkendall & Garrett, 1998). As players get older, they tend to be generally more experienced but in poorer physical condition. Professional year, which refers to the year when a player first achieves a professional ranking, seems to be an important mediator of the relationship between age and performance. This study examines the effect of years of professional experience rather than age on outcomes, because the high correlation between age

and professional years ($r = 0.87$, $P < 0.001$) generates multicollinearity problems. It is noteworthy that this variable is limited in that players may gain experience outside of professional tours.

Likewise, left-handed players enjoy an advantage in tennis matches, especially when serving against a right-handed player. Using the automated ball-tracking Hawkeye system, Loffing et al. (2009) reported that in international-standard tournaments, both for first and second serves, right- and left-handed servers differed in the distribution of ball landing points in the opponent's service box and the angle of lateral ball flight. Further, right-handed players need to adjust their habitual return stroke, because balls served left-handed differ in their spin from right-handed serves. Left-handed players are not as common as right-handed players. This is often an advantage over right-handed players, who are not accustomed to playing against left-handers and hence, have less experience with their style of play. In contrast, left-handed players are well accustomed to playing against right-handed opponents.

Ranking is an indicator of a player's overall performance during approximately one year before a given competition. The greater the ranking difference between contestants, the higher the probability of victory for the higher-ranked player. However, this effect decreases as the ranking of contestants drops (Boulrier & Stekler, 1999; Corral & Prieto-Rodriguez, 2010).

Skill and performance are important intervening variables between personal characteristics and match outcomes. This study focuses primarily on two performance indicators: serving and returning. Players who begin each rally by serving to the opponent, tend to win more points than receiving players. O'Donoghue and Brown (2008) reported a serving advantage in points lasting 3 to 4 shots on the first serve, and in points lasting 1 to 2 shots on the second serve. The first serve of any given point in men's singles tennis at Grand Slam tournaments gives the server an advantage such that 62.4% of points with duration of 3 to 4 shots are won: this is noticeably greater than the 49.7% of points won if the rally lasts for 5 or more shots.

Both serve and return of serve are strong indicators of match outcome (Chiu, 2010; Elliott, & Saviano, 2001). Good returns can overcome good service and neutralise the advantages of a good serve. Much less attention has been paid to the impact of returns, although Elliot and Saviano (2001) reported that the success of many professional tennis players is dependent, at least in part, on the return speed of the ball.

The Grand Slam tournament is entered into the model as a categorical variable. It is the court surface (i.e. grass, acrylic, or clay) that mainly affects match outcome. A slower surface is associated with more shots being played, longer rallies, and a wider distribution of shots (O'Donoghue & Ingram, 2001),

whereas on faster surfaces, serves and forehands are more successful (Collinson & Hughes, 2000). However, such an effect of court surface for men is not found by Corral and Prieto-Rodriguez (2010).

In addition, match round is entered into our model, to control its impact on match outcomes. This is because Gilsdorf and Sukhatme (2007) claimed that modelling is better at predicting outcomes in later-round matches.

Time period is the final match characteristic tested in our model. The long time span of our data allowed us to test the impact of time period on the probability of winning a match in a Grand Slam tournament. We divided the time span (1991 to 2008) into three periods according to the annual rankings of top players: (1) the time dominated by Pete Sampras and Andre Agassi (1991–1999); (2) the time dominated by Gustavo Kuerten, Lleyton Hewitt, and Andy Roddick (2000–2003); and, (3) the time dominated by Roger Federer and Rafael Nadal (2004–2008). Different time periods, with different top players and dominant playing styles, might be associated with different levels of difficulty in winning a Grand Slam match.

Statistical analyses

The analyses were conducted using the Statistical Package for the Social Sciences (SPSS 17.0). First, we used the t-test and chi-square test to compare match winners with match losers. Then, after deleting variables having multicollinearity problems, 16 predictor variables were determined for further analyses and were divided into three categories: (1) player's skills and performance (e.g. service); (2) player's characteristics; and (3) Grand Slam event. The three categories of variable were entered one at a time into the logistic regression model (a binominal logistic regression analysis was used to perform this hierarchical regression analysis). The comparative importance of the three categories, the model fitness, and the importance of individual predictors were evaluated and, based on this evaluation, an altered model was proposed for predicting the outcome of men's singles Grand Slam tournaments.

Because a player's performance in one match can influence his performance in other matches, to deal with the repeated-measures nature of the data, we used a 1-1 matched conditional logistic regression to double check the impact of personal characteristics as well as skills and performance on the chances of winning a match. In this pair-matched case-control design, cases were players who won matches and controls were players who lost the matches. However, the conditional logistic regression model is used only as a supplement to the unconditional logistic regression model because it cannot test the impact of match characteristics (round, timer period and game) as these variables are constant for each

pair of players. In addition, different from the purpose of this study, the conditional logistic regression model focused on the impact of the distinction between each pair of players on the match outcome.

Results

Descriptive analysis

The ATP dataset includes 18,288 performances from 9,144 matches involving 845 players held between 1991 and 2008. Most players were aged from 16 to 40 years (99%), with a mean age of 25.3 years ($s = 3.4$). The dominant hand of 85.2% of the players was their right hand. Players' mean stature was 184.5 cm ($s = 6.3$ cm), with a range of 165 to 208 cm. Players had participated in professional tennis matches for 6.7 years ($s = 3.3$) ranging from 0 to 31 years. A total of 716 players played in more than one match during the study period. The mean number of competitions for these 716 repeat players was 25.3 ($s = 32.2$) with a range of 2 to 237 matches each.

Table I summarises the results: according to the t-test and chi-square test results, the player who won the match was more likely to be right-handed, younger, taller, heavier, higher-ranked, more skilled than the opponent, and to have played more years as a professional.

The fit of the logistic model

Binary logistic regression assessed the impact of the 16 independent variables on the likelihood that

players would win a match. There were three blocks of independent variable: (1) player's skill and performance (number of aces, number of double faults, number of valid first serves, 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 break points converted and break points saved); (2) player's personal characteristics (stature, ranking, dominant hand, and number of years as a professional); and (3) match characteristics (Grand Slam tournament, rounds and time periods).

As Table II indicates, all three models differed from the model that shared only the intercept. In the first model, all the skill and performance measures predict match outcome: $\chi^2 = 14038.1$, $df = 9$, $P < 0.001$. The players' characteristics are influential predictors of winning a match, even after all other skill and performance variables have been entered (change of $\chi^2 = 23.5$, $df = 4$, $P < 0.001$). In addition, match characteristics predict the likelihood of winning a match. The inclusion of match characteristics increased the model chi-square (change of $\chi^2 = 38.2$, $df = 6$, $P < 0.001$). Model 3 was the most accurate in distinguishing between players who won and those who did not win: $\chi^2 = 14099.8$, $df = 19$, $P < 0.001$. The model as a whole explains between 59.6% (Cox-Snell R^2) and 79.4% (Nagelkerke R^2) of the variance in the final results of the match.

Fourteen of the 16 independent variables made unique contributions to the model. Conversely, the player's dominant hand ($B = 0.085$, $P = 0.313$), and

Table I. Descriptive statistics by match outcomes.

No.	Predictors	Match Outcomes		
		Losing Matches	Winning Matches	Independent samples t-test
1	Age (years)	25.4 (3.4)	25.2 (3.4)	-3.42***
2	Stature (cm)	184.2 (6.4)	184.7 (6.2)	6.17***
3	Mass (kg)	78.5 (6.7)	79.3 (6.6)	7.72***
4	Ranking	93.2 (101)	59.6 (85.9)	-24.17***
5	Years as a professional	6.6 (3.3)	6.8 (3.2)	3.94***
6	Aces	6.6 (5.7)	8.9 (6.4)	26.01***
7	Double faults	5.3 (3.4)	4.2 (3.1)	-21.32***
8	% 1 st serve	58.6 (8.4)	60.8 (8.3)	17.62***
9	% 1 st serve points won	66.0 (9.1)	76.3 (7.9)	82.17***
10	% 2 nd serve points won	43.0 (9.2)	53.4 (9.8)	74.08***
11	% 1 st serve return points won	23.4 (9.8)	33.5 (9.6)	69.78***
12	% 2 nd serve return points won	45.8 (9.9)	55.9 (9.4)	70.76***
13	% break points converted	36.0 (22.8)	47.0 (16)	36.98***
14	% break points saved	52.2 (15.9)	63.4 (22.8)	37.62***
15	% serve points won	56.4 (6.8)	67.4 (6.3)	112.38***
16	% return points won	32.1 (7.4)	42.8 (6.9)	100.16***
17	% total points won	44.5 (4.4)	54.7 (4.2)	160.65***
	Dominant hand			Chi-square
18	Right hand	84.5%	85.9%	7.09**
	Left hand	15.5%	14.1%	

Note: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table II. Winning men's Grand Slam singles matches - Hierarchical binomial logistic regression analysis (logit coefficients).

Predictors	Model 1		Model 2		Model 3	
	B	Exp(B)	B	Exp(B)	B	Exp(B)
<i>Skills & Performance</i>						
Aces	-0.033***	0.968	-0.030***	0.970	-0.029***	0.971
Double faults	0.037***	1.038	0.041***	1.042	0.046***	1.048
% 1 st serve	0.075***	1.078	0.076***	1.079	0.078***	1.081
% 1 st serve points won	0.238***	1.268	0.238***	1.269	0.241***	1.272
% 2 nd serve points won	0.158***	1.172	0.157***	1.171	0.159***	1.173
% 1 st serve return points won	0.144***	1.155	0.143***	1.154	0.143***	1.154
% 2 nd serve return points won	0.141***	1.152	0.140***	1.150	0.140***	1.151
% break points converted	0.031***	1.031	0.031***	1.031	0.031***	1.032
% break points saved	0.038***	1.038	0.038***	1.038	0.038***	1.038
<i>Personal Characteristics</i>						
Left hand	-	-	0.065	1.067	0.069	1.071
Stature	-	-	-0.011*	0.989	-0.013*	0.987
Ranking	-	-	-0.002***	0.998	-0.002***	0.998
Years as a professional	-	-	-0.006	0.994	-0.005	0.995
<i>Match characteristics</i>						
Grand Slam (ref. = French Open)						
Australia Open	-	-	-	-	0.151	1.163
US Open	-	-	-	-	0.347***	1.415
Wimbledon	-	-	-	-	0.287**	1.333
Round					-0.049*	0.953
Time period (ref. = 91-99)						
00-03					-0.084	0.919
04-08					-0.297***	0.743
Model Chi-square	14038.1		14061.6		14099.8	
Change in Model Chi-square	14038.1		23.5		38.2	
Sig. of Change	0.0000		0.0000		0.0000	
Cox-Snell R square	0.594		0.594		0.595	
Nagelkerke R square	0.792		0.792		0.794	

Note: The reference category for the dependent variable is winning a match. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, $N = 15,589$.

the number of years the player had played as a professional ($B = -0.007$, $P = 0.478$) did not influence match outcomes in the model.

After examining the results for individual predictors, two changes were made to improve the performance of the model: (1) because stature, number of aces and double faults revealed relationships (see Model 3 in Table II) but in opposite directions from the results of the t-tests in Table I, these variables were re-coded as categorical variables to probe their possible non-linear relationships with match result; and (2) the two non-predictors (dominant hand and years as a professional) were excluded from the altered model (see Table III).

The altered model successfully distinguishes between players who won and those who did not win a match in 90.6% of the men's singles Grand Slam tournament matches. The independent variables give adequate predictions compared with the null model: $\chi^2 = 14319.4$, $df = 20$, $P < 0.001$. The model as a whole explained between 59.5% (Cox-Snell R^2) and 79.4% (Nagelkerke R^2) of the variance in match outcomes.

As shown in Table III, the strongest predictor of match outcome was the percentage of first service

points won, recording an odds ratio (OR) of 1.27 ($B = 0.235$, $P < 0.001$, 95% confidence interval (CI), 1.25 to 1.279). This indicates that when the percentage of first serves won increased by one unit, a player was 1.27 times more likely to win the match, after controlling for all other factors in the model. Following this logic, as the chance of winning a point on the second serve increased by one unit, a player was 1.17 times more likely to win the match ($B = 0.155$, $P < 0.001$, 95% CI, 1.158 to 1.178). By increasing the percentage first serve return points won by one unit, the receiver was 1.16 times more likely to win the match ($B = 0.144$, $P < 0.001$, 95% CI, 1.146 to 1.164). In addition, as the percentage second serve return points won increased by one unit, the player was 1.15 times more likely to win the match ($B = 0.142$, $P < 0.001$, 95% CI, 1.143 to 1.161). The percentage of valid first serves ($B = 0.073$, $P < 0.001$, OR = 1.076, 95% CI, 1.067 to 1.085), break point saves ($B = 0.038$, $P < 0.001$, OR = 1.039, 95% CI, 1.035 to 1.042), and break point conversions ($B = 0.031$, $P < 0.001$, OR = 1.032, 95% CI, 1.028 to 1.035) also influenced the likelihood of winning a match. Players who served less than four aces in a match were less

Table III. Unconditioned logistic regression predicting the likelihood of winning or losing in a Grand Slam (altered model).

	<i>B</i>	SE	Wald	<i>df</i>	OR	OR 95% CI	
						lower	upper
<i>Match characteristics</i>							
Grand Slam (ref. = French Open)							
The Australian Open	0.111	0.093	1.44	1	1.12	0.932	1.339
The US Open	0.315	0.084	13.87***	1	1.37	1.161	1.616
Wimbledon	0.270	0.084	10.41**	1	1.31	1.112	1.544
Time period (ref. = 91–99)							
00–03	−0.041	0.075	0.30	1	0.959	0.828	1.112
04–08	−0.274	0.084	10.71**	1	0.761	0.646	0.896
Round	−0.058	0.024	5.94*	1	0.944	0.901	0.989
<i>Personal Characteristics</i>							
Stature (ref. = ≤ 180 cm)							
181–185 cm	0.182	0.076	5.74*	1	1.2	1.034	1.393
≥ 186 cm	0.038	0.071	0.285	1	1.04	0.904	1.194
Ranking	−0.002	0.000	21.52***	1	0.998	0.998	0.999
<i>Skills and Performance</i>							
Aces (ref. = 0–4)							
5–8	0.297	0.089	11.15**	1	1.35	1.131	1.602
≥ 9	0.174	0.075	5.45*	1	1.19	1.028	1.378
Double faults (ref. = 0–2)							
3–5	−0.218	0.092	5.59*	1	0.804	0.671	0.963
≥ 6	−0.183	0.070	6.79**	1	0.833	0.726	0.956
% 1st serve	0.073	0.004	310.68***	1	1.076	1.067	1.085
% 1st serve points won	0.235	0.006	1684.11***	1	1.265	1.250	1.279
% 2nd serve points won	0.155	0.004	1251.48***	1	1.168	1.158	1.178
% 1st serve return points won	0.144	0.004	1341.68***	1	1.155	1.146	1.164
% 2nd serve return points won	0.142	0.004	1230.00***	1	1.152	1.143	1.161
% break points converted	0.031	0.002	313.24***	1	1.032	1.028	1.035
% break points saved	0.038	0.002	468.75***	1	1.039	1.035	1.042
Constant	−43.22	0.81	2872.36***				
Cox-Snell $R^2 = 0.595$, Nagelkerke $R^2 = 0.794$							
Overall Model Fit $\chi^2 = 14319.4***$, $df = 20$							

Note: The reference category for the dependent variable is winning a match. OR = odds ratio; CI = confidence interval; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. $N = 15,833$.

likely to win a match than (a) those who served five to eight aces ($B = 0.297$, $P < 0.01$, OR = 1.35, 95% CI, 1.131 to 1.602) and (b) those who served more than nine aces ($B = 0.17$, $P < 0.05$, OR = 1.19, 95% CI, 1.028 to 1.378). Moreover, players who served fewer than two double faults were more likely to win a match than those who served three to five ($B = -0.218$, $P < 0.05$, OR = 0.804, 95% CI, 0.671 to 0.963) and those who served more than six ($B = -0.183$, $P < 0.01$, OR = 0.833, 95% CI, 0.726 to 0.956).

For personal characteristics, four independent variables had odds ratios of less than 1, indicating a negative relationship between the independent variable and the dependent variable. When all other variables were considered, the odds ratio of 1.20 for a player's stature between 181 cm and 185 cm meant that these players were 1.20 times more likely to win a match versus those of stature below 180 cm ($B = 0.182$, $P < 0.05$, 95% CI, 1.034 to 1.393). In the same manner, the odds ratio of 0.998 for a player's ranking indicated that when a player was ranked lower by one unit, they were 0.998 times as

likely to win, when all other variables were considered in the model.

Match characteristics were influential. The odds ratio of 1.37 for the US Open (with French Open as reference) indicated that players were 1.37 times more likely to win a match in the US Open versus those in the French Open when all other variables were controlled in the model ($B = 0.315$, $P < 0.001$, 95% CI, 1.161 to 1.616). Likewise, players were 1.31 times more likely to win a match in Wimbledon than in the French Open ($B = 0.27$, $P < 0.01$, 95% CI, 1.112 to 1.544). Because the coefficient B of the Australian Open did not differ from 0, we do not have enough evidence to conclude that players have any difference in their chances of winning a match in the Australian Open compared with the French Open. In terms of time period, a player was only 0.76 times as likely to win a match in the period 2004 to 2008 than pre-1999 ($B = -0.274$, $P < 0.01$, OR = 0.761, 95% CI, 0.646 to 0.896). However, there was no difference in the period 2000 to 2003 compared with pre-1999 ($B = -0.041$, $P = 0.58$, OR = 0.959). Not

surprisingly, it was more difficult to win a match in later rounds than in earlier rounds ($B = -0.058$, $P < 0.05$, OR = 0.944, 95% CI, 0.901 to 0.989).

The results in Table IV show that the impact of stature (Wald = 2.057, $P = 0.357$) and the performance of converting ($B = 0.026$, $P = 0.061$) and saving break points ($B = 0.13$, $P = 0.352$) disappeared when only the distinction between the two players in each match were considered.

Discussion

The main purpose of the present study was to establish a model that could effectively predict the chance of winning tennis matches in men's singles Grand Slams. The findings provide a model that explains meaningful variance in match outcome, predicts the winners in more than 90% of cases, and identifies characteristics of winning players from those of losing players. Specifically, service and return outcomes, players' ranking, stature, time period, and match characteristics were predictors of match outcomes.

This study confirms that serves and returns are the most important skills in tennis (Elliot & Saviano, 2001; O'Donoghue & Brown, 2008). First serves are the best predictors of match outcomes. Aces, valid first serves, and second serve points won, also increased the chances of winning. Winning first-serve return and second-serve return points increased the chances of winning matches. In addition, the chance of winning a match is also positively

associated with the chances of converting and saving break points, but this advantage disappeared when the distinctions (between a player and his opponent) were considered. The importance of returns has been overlooked. This finding suggests that the training of elite-standard men players should place more emphasis on improvements in return of service.

Another notable finding is that the positive impact of a player's stature on winning matches disappears when he is taller than 186 cm. When a player is between 181 cm and 185 cm, they are more likely to win a match than those below 180 cm. However, when a player is taller than 186 cm, the advantage of stature disappears compared with those below 180 cm. It is widely believed that stature tends to increase serve speed (and successful serving leads to winning). Nevertheless, this study finds that the positive correlation between stature and winning is not linear, and stature stops generating a positive effect when players are taller than 186 cm. In addition, the distinction of stature between a player and his opponent in a particular match is not a significant predictor of the match outcome. This finding has important implications for selecting young male tennis players, as their final stature might influence future performances in Grand Slam matches.

Surprisingly, the dominant hand was not a predictor of winning matches: left-handed elite players do not have an advantage when competing against right-handed opponents. As suggested by Loffing et al. (2009), there are more left-handed players so their

Table IV. 1:1 Matched conditional logistic regression predicting the likelihood of winning or losing in a Grand Slam.

	<i>B</i>	SE	Wald	<i>df</i>	OR	OR 95% CI	
<i>Personal Characteristics</i>							
Stature (ref. = ≤ 180 cm)							
181–185 cm	−0.069	0.089	0.596	1	0.934	0.785	1.111
≥ 186 cm	−0.116	0.081	2.041	1	0.891	0.760	1.044
Ranking	0.000	0.000	5.544*	1	0.999	0.998	1.000
<i>Skills and Performance</i>							
Aces (ref. = 0–4)							
5–8	0.350	0.107	10.69**	1	1.419	1.151	1.751
≥ 9	0.163	0.088	3.45	1	1.177	0.991	1.398
Double faults (ref. = 0–2)							
3–5	−0.336	0.115	8.51*	1	0.714	0.570	0.895
≥ 6	−0.274	0.085	10.478**	1	0.761	0.644	0.898
% 1st serve	0.092	0.006	268.19***	1	1.097	1.085	1.109
% 1st serve points won	0.22	0.016	197.69***	1	1.246	1.208	1.285
% 2nd serve points won	0.136	0.012	135.47***	1	1.146	1.12	1.172
% 1st serve return points won	0.057	0.015	14.59***	1	1.059	1.028	1.09
% 2nd serve return points won	0.043	0.011	14.46***	1	1.044	1.021	1.067
% break points converted	0.026	0.014	3.51	1	1.026	0.999	1.054
% break points saved	0.013	0.014	0.866	1	1.013	0.986	1.04
Overall Model Fit $\chi^2 = 4732.67^{***}$, $df = 14$							

Note: The reference category for the dependent variable is winning a match. OR = odds ratio; CI = confidence interval; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. $N = 15,619$.

former advantage has been reduced because right-handed players have more opportunity to play them. Based on this, we argue that the left-handed advantage disappears in international tournaments, as elite-standard players are now trained to accommodate the ball distribution and angle of lateral ball flight from left-handed players. However, Corral and Prieto-Rodriguez (2010) claimed an advantage in being left-handed when a lower-ranked player competed with a right-handed, higher-ranked opponent in a Grand Slam tournament. One possible reason is that higher-ranked players are not accustomed to the style of play of the lower-ranked left-handed opponent.

The number of years of experience as a professional player did not influence match outcomes. There are two possible explanations for this result. First, professional players have too many unpredictable interruptions in their careers, such as injuries, marriage, etc. Second, the performance of top players at such a high level cannot be markedly improved over time and associated experience. In contrast, the energy advantage that younger players have appears to be greater than the experience advantage that older players have. The probability of a higher-ranked player winning decreases as this player competes against younger players (Corral & Prieto-Rodriguez, 2010).

In terms of match characteristics, players are more likely to win a match in the US Open and Wimbledon than in the French Open under the predictions in the model. The French Open is the only tournament played on a clay surface. Clay surfaces produce lower ball speed, longer rallies and wider distribution, all of which demand additional energy. In addition, players slide on clay surfaces: this demands specific skills in approaching a flying ball. Furthermore, serves and forehands that can be an advantage on faster surfaces are ineffective on clay surfaces because of greater friction (Collinson & Hughes, 2000). This therefore increases the difficulty of winning a match in the French Open (clay courts) as opposed to the US Open (synthetic courts) and Wimbledon (grass courts) for players with less experience on clay courts. A male player should start his professional tennis career by participating in either the US Open or Wimbledon, which are easier to win and in turn could enhance his confidence. Even though fewer players dominate the French Open because of the specific demands of clay-court tennis, the domination can last longer than on other surfaces. One notable example is Rafael Nadal, who has won seven consecutive French Open titles.

We noted greater difficulty for a tennis player to win a Grand Slam match between 2005 and 2008, than before 1999. Compared with the 1990s,

matches in the 2000s were more competitive in terms of available prize money, performance and combined rating points of the top three players, and therefore it became harder for most players to win a match in a Grand Slam between 2005 and 2008, than before 1999.

According to goal-setting theory, specific goals increase performance through narrowing attention, mobilising effort, prolonging persistence, and fostering cognitive strategies (Locke & Latham, 2002). Therefore, tennis coaches can use the results of past Grand Slam tournaments to set specific goals for their players: this would help to improve the effectiveness of training. For example, the results show that winning players made 8.9 ($s = 6.4$) aces and 4.2 ($s = 3.1$) double faults for each match. Tennis coaches can record their players' skill and performance in the daily training, and compare them with those of the winning players. In this way, they can know exactly how far their skills and performance are from the top players, and make specific goals and training plans. In other words, the results in Table I can be used to guide tennis coaches with more accurate and specific data. In addition, the results show that the mean age and number of years' experience as a professional are 25.3 ($s = 3.4$) and 6.8 ($s = 3.2$) years, respectively. Hence, tennis coaches and player agencies can use this figure to plan players' careers and refine training and competition schedules.

This study is limited by the scope of the data kept at the ATP website. Therefore, the proposed model was constructed and tested only by the variables that were available. To increase a future model's prediction, research should include several important factors that are not tested in this study, such as unforced errors, serving and volleying. In addition, it is noteworthy that many of the player characteristics like stature or mass, are self-reported, and such data are not always up to date. Another limitation is that the findings of this study are applicable only to men players, so its generalisation to other groups, such as women players, needs to be tested. Finally, since some of the players played in men's singles Grand Slams more than once over the range of years, these observations would not be entirely independent. To confirm the findings obtained in this study, future research would probably benefit from the use of multilevel modelling that analyses units or measurements grouped at different levels. An example is the recent analyses made by Watts, Coleman, and Nevill in 2012.

In conclusion, the chances of winning a match in men's singles Grand Slams were improved through higher proportions of valid first serves, aces, second-serve points won, first serves returned, second-serves returned, and converted and saved break points.

Men players with stature less than 180 cm were less likely to win a match than those between 181 cm and 185 cm, but not less likely than those taller than 186 cm. The dominant hand and the number of professional years did not influence match outcomes. Finally, it is more likely for a player to win a match in the US Open and Wimbledon than in the French Open, and in the 1990s than in the period of 2004 to 2008.

Acknowledgement

This study was supported by a grant to the first author from the National Science Council, Taiwan (NSC99-2410-H-020-014-A1).

References

- Barnett, T., & Clarke, S. R. (2005). Combining player statistics to predict outcomes of tennis. *IMA Journal of Management Mathematics*, 16, 113–120.
- Boulrier, B., & Stekler, H. (1999). Are sports seedings good predictors? An evaluation. *International Journal of Forecasting*, 21, 565–576.
- Chiu, F. C. (2010). Comparative analysis on the skill characteristics of men's singles matches in 2008 Grand Slam tournaments. *Sports and Exercise Research*, 12, 83–95.
- Collinson, L., & Hughes, M. (2000). Surface effect on the strategy of elite female tennis players. *Conference Communications*, 266–267.
- Corral, J., & Prieto-Rorriguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*, 26(3), 551–563.
- Cross, R., & Pollard, G. (2009). Grand Slam men's singles tennis 1991–2009: Serve speeds and other related data. *ITF Coaching & Sport Science Review*, 16(49), 8–10.
- Elliott, B., & Saviano, N. (2001). Serves and returns. In P. Robert & J. Groppe, (Eds.), *World-class tennis technique* (pp. 207–222). Champaign, IL: Human Kinetics.
- Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.
- Gilsdorf, K., & Sukhatme, V. (2007). Testing Rosen's sequential elimination tournament model: Incentives and player performance in professional tennis. *Journal of Sports Economics*, 9, 287–303.
- Gould, D., Petlinchokoff, L., Simons, J., & Vevera, M. (1987). Relationship between competitive state anxiety inventory-2 subscale scores and pistol shooting performance. *Journal of Sport Psychology*, 9, 33–42.
- Harris, N. (2012, February 13). Exclusive: Djokovic, Nadal, Federer – as close to perfection as tennis has ever been. *Sporting Intelligence*. Retrieved from <http://www.sportingintelligence.com/2012/02/13/exclusive-djokovic-nadal-federer-as-close-to-perfection-as-tennis-has-ever-been-130201/>
- Helsen, W. F., Starkes, J. L., & Hodges, N. J. (1998). Team sports and the theory of deliberate practice. *Journal of Sport and Exercise Psychology*, 20, 12–34.
- Hornery, D. J., Farrow, D., Mujika, I., & Young, W. (2007). An integrated physiological and performance profile of professional tennis. *British Journal of Sports Medicine*, 41, 531–536.
- Jones, G. (1995). More than just a game: Research, developments and issues in competitive anxiety in sport. *British Journal of Psychology*, 86, 449–479.
- Kirkendall, D. T., & Garrett, W. E. (1998). The effects of aging and training on skeletal muscle. *The American Journal of Sports Medicine*, 26(4), 598–602.
- Klaassen, F., & Magnus, J. (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*, 148, 257–267.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717.
- Loffing, F., Hagemann, N., & Strauss, B. (2009). The serve in professional men's tennis: Effects of players' handedness. *International Journal of Performance Analysis in Sport*, 9, 255–274.
- MacCurdy, D. (2006). Talent identification around the world and recommendations for the Chinese Tennis Association. Retrieved from http://www.itftennis.com/shared/medialibrary/pdf/original/IO_18455_original.pdf
- O' Donoghue, P. G., & Brown, E. (2008). The importance of service in Grand Slam singles tennis. *International Journal of Performance Analysis in Sport*, 8, 70–78.
- O'Donoghue, P. G., & Ingram, B. A. (2001). Notational analysis of elite tennis strategy. *Journal of Sports Sciences*, 19, 107–115.
- Pluim, B. (2006). Medical considerations when identifying talent. *ITF Coaching & Sport Science Review*, 39, 6–7.
- Rees, T., & Hardy, L. (2004). Matching social support with stressors: Effects on factors underlying performance in tennis. *Journal of Sport and Exercise Psychology*, 5, 319–337.
- Rees, T., Ingledew, D. K., & Hardy, L. (1999). Social support dimensions and components of performance in tennis. *Journal of Sports Sciences*, 17, 421–429.
- Reid, M., Crespo, M., Lay, B., & Bery, J. (2007). Skill acquisition in tennis: Research and current practice. *Journal of Science and Medicine in Sport*, 10, 1–10.
- Scheibehenne, B., & Brofer, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, 23, 415–426.
- Struder, H. K., Hollmann, W., & Duperly, J. (1995). Amino acid metabolism in tennis and its possible influence on the neuroendocrine system. *British Journal of Sports Medicine*, 29, 28–30.
- Vealey, R. S. (1994). Knowledge development and implementation in sport psychology: A review of the sport psychologist, 1987–1992. *The Sport Psychologist*, 8, 331–348.
- Watts A. S, Coleman I., & Nevill, A. M. (2012). The changing shape characteristics associated with success in world-class sprinters. *Journal of Sports Sciences*, 30(11), 1085–1095.