Playoffs Round 1: Racquetball

An analysis on the dominance of Kane Waselenchuk: What factors lead to his success?

By Aayush Bhatt, Connor Ollio, Cole Henry, Garrett McKinstry, Caelyn Barrett

Introduction

Created in 1950, racquetball is one of the few indoor racket sports widely played today. Often described as a “rich man's sport” and widely confused with Squash, racquetball has a variety of factors that make it a unique sport and worthy of analysis. Racquetball consists of two players for singles, or two teams of two players for doubles. For the sake of this analysis, we will be referring to the singles game. A point can only be scored when you are serving, games are played to 15 points, and, in professional matches, a player must win the best 2 out of 3 games to win the overall match. Scoring usually varies across tournaments but this is the general majority of how racquetball scoring is conducted. Though nearly 4 million people play racquetball annually, there isn't much coverage of the sport professionally. Consequently, our analysis focused on the most dominant and widely covered racquetball player: Kane Waselenchuk. The winningest player of all time, being ranked number one for over 13 seasons in a row, Waslenchuck’s dominance can be attributed to a multitude of factors. As such, we thought it would be insightful to conduct observations on a couple of Waslenchuck’s games (over 130 rallys) to analyze which variables are the best predictors in winning a rally and eventually an overall game. Since you can only score while you are serving, we initially hypothesized that the service type and the number of faults would be most deterministic in winning a rally. To investigate this and other factors in play, we decided to record 130 observations of gameplay with 15 box score variables. Our analysis implores to better uncover the way racquetball is played and if surprising correlations or predictors exist for winning rallies.

Opponent	Rally	Serve?	Serve Type	Starting Serve Position	Faults	Walls on Winning Hit	Length of Rally	Direction Changes	Opponent Direction Change	Forehands	Backhands	Opponent Forehands	Opponent Backhands	WinServe?
Andres Acuna	1	Yes	Backhand Drive	Center	0	1	1	0	0	1	0	0	1	Yes
Andres Acuna	2	Yes	Backhand Drive	Center	0	1	5	1	2	2	1	0	2	Yes
Andres Acuna	3	Yes	Backhand Lob	Opponent Backhand	1	1	2	0	0	1	0	0	1	No
Andres Acuna	4	No	Backhand Lob	Forehand	1	2	2	0	0	0	1	1	0	Yes
Andres Acuna	5	Yes	Backhand Drive	Center	1	1	9	2	3	3	2	3	1	No
Andres Acuna	6	No	Forehand Lob	Backhand	1	2	4	2	2	2	0	2	0	Yes
Andres Acuna	7	Yes	Forehand Drive	Opponent Backhand	0	2	5	2	3	2	1	1	1	No
Andres Acuna	8	No	Backhand Lob	Forehand	1	1	5	0	0	0	2	3	0	No
Andres Acuna	9	No	Backhand Lob	Forehand	1	1	3	0	0	0	1	2	0	Yes
Andres Acuna	10	Yes	Backhand Drive	Center	0	2	3	0	2	1	1	0	1	Yes

For our observations, we watched two recorded games on Youtube: one played in the 2019 John Pelham Memorial TOC and the other in the 2020 Longhorn Open. Each group member watched one game in a match, and all group members watched the leftover unseen game. The data was recorded in a shared Google Sheets file in a 130 x 15 dimension table that included all 130 rallys and 15 variables of interest. Each observation represents one rally. Since only 2 games were watched, there are only two opponents: Andres Acuna and Rodrigo Montoya. There was a combination of categorical and numerical variables in the data. The categorical variables included: Opponent, whether or not Kane served, Serve Type, Starting Serve Position, and whether or not Kane won the rally. The rest were the numerical variables: Faults, Walls Hit on Winning Hit, Rally Length, Kane's direction changes, Kane's opponent's direction changes, Kane's forehands, Kane's backhands, Kane's opponent's forehands, and Kane's opponent's backhands.

Each variable was collected by simply watching the rally and recording the action. The opponent, whether or not Kane served, serve type, and serve position were recorded before the rally began. Opponents and whether or not Kane served are fairly obvious, but the serve type and serve position are more complex. For the serve type, there are five types of serves in racquetball: the forehand drive, the backhand drive, the forehand lob, the backhand lob, and the wrap-around. The forehand and backhand drive are simply hard-hit serves that aim for the opponent's forehand or backhand, respectively. In contrast, the forehand and backhand lob are softly hit serves meant to bounce high to the opponent's forehand or backhand, respectively. These lob serves are most common after a fault since they are easier to place and thus avoid a double fault (which gives the serve to the opponent). Finally, the wrap-around serve is a hard-hit serve meant to hit three walls, the front, side, and back, before the opponent can return the ball. This type of serve seems to be the least common. Next, the numerical variables were largely recorded during gameplay. Since racquetball moves so quickly, the video had to be slowed down and stopped multiple times to accurately gather the data. The data is majority small integer values since the rallies generally did not last many hits. Describing the data, faults are simply the number of times the server had an illegal serve. This variable had values of zero or one since the server cannot fault more than once before losing the serve. Rally length was simply the number of total hits before the ball hit the ground twice before the player could return. Direction changes and opponent direction changes were more subjective variables. They measured how many times the player, Kane or his opponent, was forced to change his body's movement quickly in order to return the ball. Finally, forehands, backhands, opponent forehands, and opponent backhands were recorded rather easily, just by counting how many times during the rally the player hit a forehand or backhand shot. WinServe was the final variable recorded since it was at the end of the rally. It states whether or not Kane won the rally which was our purpose for analyzing these videos.

Summary of Data

Categorical Variables	Possible Values	Count	Relative Frequency
Serve?	Yes	79	61.2%
	No	50	38.8%
Serve Type	Backhand Drive	35	27.1%
	Backhand Lob	44	34.1%
	Forehand Drive	30	23.3%
	Forehand Lob	12	9.3%
	Wrap Around	8	6.2%
Starting Serve Position	Center	10	7.8%
	Opponent Backhand	58	45.0%
	Opponent Forehand	10	7.8%
	Backhand	41	31.8%
	Forehand	10	7.8%
WinServe?	Yes	81	62.8%
	No	48	37.2%
Faults	Yes	51	39.5%
	No	78	60.5%

Table 1: Table 1 summarizes our five categorical variables we analyzed while studying racquetball. Firstly, in our table we display the type of outcomes for our categorical variables. For example, Serve Type can be one of five different serves, while WinServe? is either yes Kane won the serve or no he did not. Additionally, we calculated the relative frequencies of each variable to show how likely each outcome was.

Numerical Variables	Min	Q1	Median	Q3	Max	Mean	St.D.	Total Count	95% Confidence Interval	
Walls on Winning Hit	0	1	1	2	3	1.52	0.63	196	1.41	1.63
Length of Rally	0	2	3	5	13	3.57	2.14	461	3.21	3.94
Direction Change	0	0	0	1	3	0.50	0.71	65	0.38	0.63
Opponent Direction Change	0	0	1	1	4	0.78	0.85	101	0.64	0.93
Forehands	0	1	1	2	4	1.22	0.88	158	1.07	1.38
Backhands	0	0	1	1	4	0.70	0.81	90	0.56	0.84
Opponent Forehands	0	0	1	1	3	0.84	0.82	108	0.70	0.98
Opponent Backhands	0	0	1	1	6	0.84	0.87	108	0.69	0.99

Table 2: Table 2 summarizes the eight quantitative variables we recorded when watching racquetball. We calculated summary statistics of our data, such as minimum, first quartile, median, third quartile, maximum, mean, standard deviation, total count of each variable and lastly 95% confidence intervals. As seen from our standard deviations most of our data appears to be relatively constant and tends not to vary from rally to rally.

	Walls on Winning Hit	Length of Rally	Direction Changes	Opponent Direction Change	Forehands	Backhands	Opponent Forehands
Length of Rally	0.003						
Direction Changes	-0.084	0.556					
Opponent Direction Change	0.037	0.540	0.379				
Forehands	0.099	0.544	0.382	0.350			
Backhands	-0.042	0.687	0.351	0.395	-0.146		
Opponent Forehands	-0.108	0.546	0.291	0.219	0.062	0.422	
Opponent Backhands	0.042	0.746	0.361	0.385	0.436	0.496	-0.059

Figure 1: Figure 1 is a correlation matrix between our eight quantitative variables. The variables with highest correlation are Length of Rally and Opponent Backhands and this is expected because the opponent's backhand returns directly influence the length of the rally. A noteworthy correlation is the correlation between Opponent Direction Change and Backhand. Despite the correlation being relatively weak, it shows that Kane was able to make his backhand shots difficult to return by forcing the opponent to change direction.

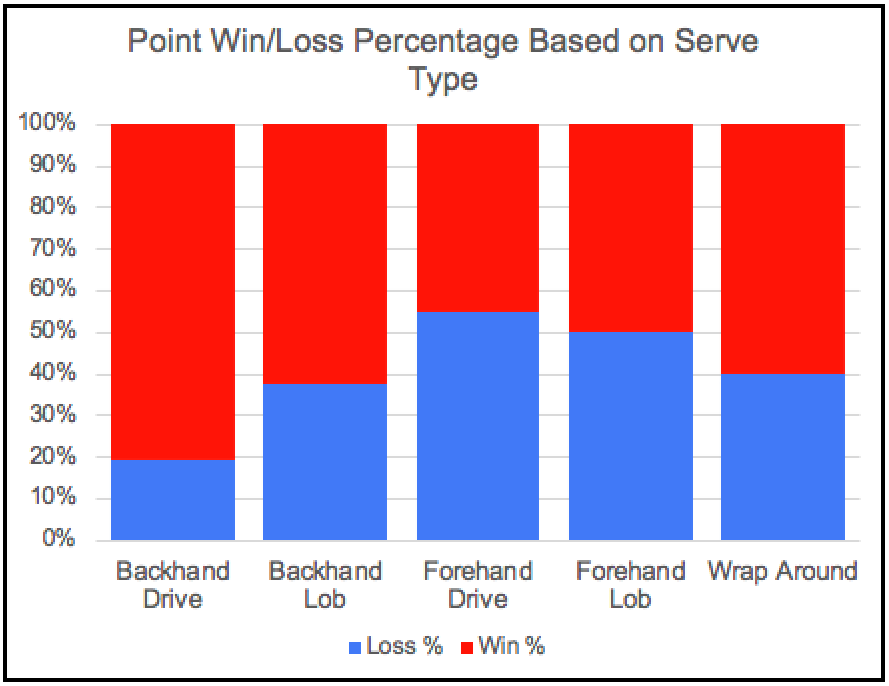

Figure 2: Figure 2 summarizes the percentage of points won for each serve type. The backhand drive serve appears to be the most effective serve, while forehand drive is the least effective. Overall it appears the serves directed towards a player's backhand are the most successful. Lastly, it is difficult to analyze the effect of the wrap around serve because it rarely occurred.

Rally Number \ Length of Rally	1 --2	3 -- 4	5 -- 6	>6
0 -- 10	7	7	4	1
10 -- 20	8	5	5	2
20 -- 30	8	9	2	1
30 -- 40	8	7	2	3
40 -- 50	3	11	4	2
50 -- 60	7	8	3	2
>60	2	5	1	1

Figure 3: Figure 3 is a heat map between the length of the rally and the rally number. The heat map shows that the rally number did not seem to have any influence on the rally length.

Insights

After studying the results of our data analysis, the first thing we realized was that WinServe? is not highly correlated with any other variable. The variable with the highest correlation to Kane winning the point is the number of opponent backhands. However, just because no other variables are highly correlated with WinServe? does not mean some aren't statistically significant. In order to do calculations with WinServe ?, each result must be converted into a binary (yes / no) result or in the case of other variables like Serve type, each serve is represented by a number 1-5. After turning all the categorical variables into numerical ones, we can create a GLM model predicting the now binary WinServe? using all of our predictors. This model then gives us the p-values of the Z-tests for each variable, proving 'Direction Changes', 'OpponentDirectionChange', 'Backhands', and 'Forehands' as significant variables in the model with respective P-values 0.00371, 0.00164, 0.01860, 0.01188. We can then refine the model just to contain these predictors which gives us a new set of p-values where Direction Changes proves to be the most significant with a P value of 0.00451, with opponentDirectionChange close behind at 0.00632. From this, we can conclude that from our research and data set, the best predictor for whether Kane wins a rally is how many times he and his opponent are forced to change directions. Although we originally predicted that Type of Serve and number of faults would be the most significant predictors, both direction change variables make a lot of sense, especially because of their corresponding coefficients. The coefficient for 'Direction Changes' is -0.5127 (seen in figure 4), meaning every time Kane is forced to change directions, his predicted chance of winning the point decreases. Meanwhile, 'OpponentDirectionChange' has a coefficient of 0.5470 (seen in figure 5), meaning if Kane forces the opponent to change directions, his predicted chance of winning the point increases. In summary, we concluded that the most beneficial thing for Kane to do in a game would be to force his opponent to change directions, i.e. hit the ball to the other side of the court. With this information, our first question was how often does Kane change the direction of his opponent? The mean number of 'Opponent Direction Changes' per rally is 0.7829, meaning Kane does not always do this and is the reason why we think this insight could be helpful.

It turns out that the serve does not matter as much as we originally thought; this maybe could have been predicted by the fact that the mean Length of Rally is 3.57 hits. It turns out that length of rally equals 1 only 10.08% of the time, meaning only 10% of serves are aces.

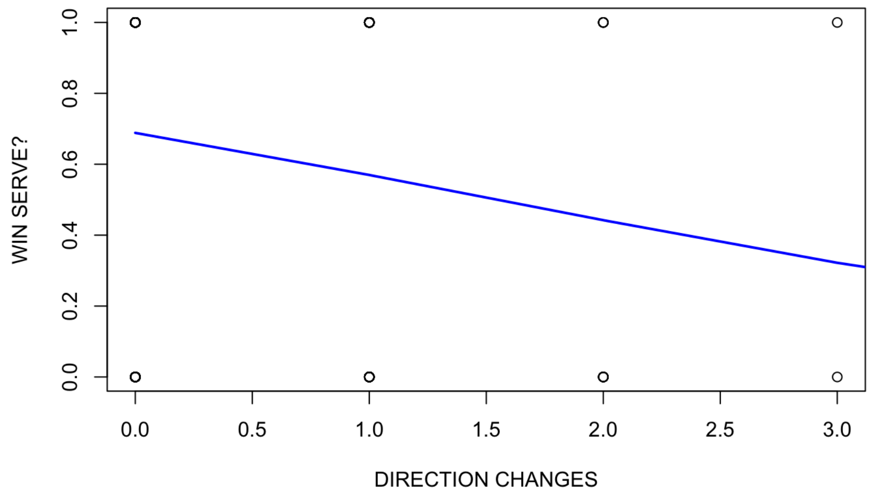

Figure 4: Plot of the negative and statistically significant relationship between Kane's number of direction changes and his likelihood of winning the rally.

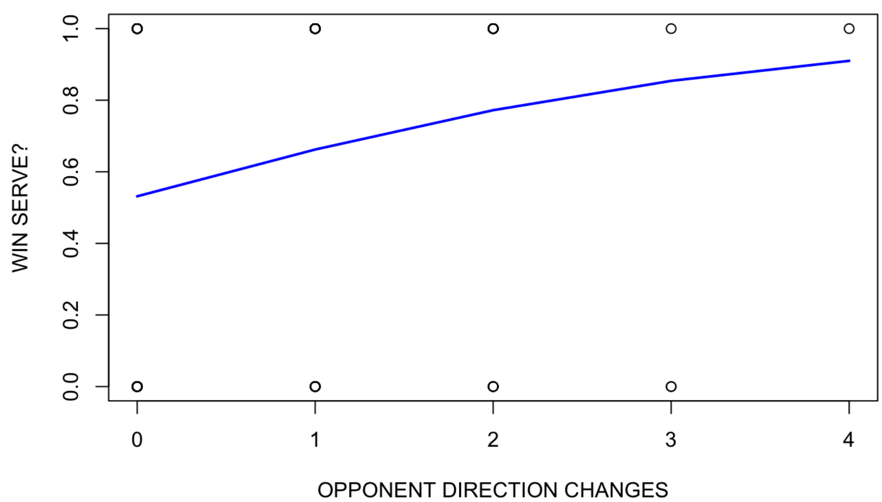

Figure 5: Plot of the positive and statistically significant relationship between Kane's opponent's number of direction changes and his likelihood of winning the rally.

After determining which variables had the most impact on winning a point, we were still interested in what impact serving had on the game. It was found that serve type is a significant predictor of faults, i.e. each serve type has a significantly different chance of resulting in a fault. For example, a forehand drive has a 0% chance of faulting whereas backhand lob has 86% chance of faulting. Fitting serve type as a predictor of faults with a linear model gives a p value of 0.0156 from a t-test (seen in figure 6). For a player like Kane, this information wouldn't be necessary for his first serve but if he faulted, it would be best to do a forehand drive on the second attempt so as to not double fault.

```
Call:
glm(formula = data$Faults ~ `Serve Type`, family = gaussian,
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5153  -0.4339  -0.2710   0.5661   0.8105

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.59675    0.09243   6.456 2.07e-09 ***
`Serve Type` -0.08145    0.03321  -2.452   0.0156 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 6: Model formula with intercepts and T-test stats for Serve type predicting Faults

Although we were able to make some conclusions about our data, we certainly don't believe our project and analysis was perfect. Firstly, based on the nature of our data, the only conclusions we are able to make are those based on Kane Waselenchuk because we only collected data from his racquetball matches. It was determined that direction change was the most significant variable to predict a win or loss of point, however we can't make conclusions on any other racquetball players or the sport as a whole with this variable. To widen the scope of this project, we would need to collect data from a number of different matches with different players because it could be true that direction change is only significant in Kane's style of play. If we had taken data on more players, we could also have added more variables such as age, current record, win/loss streak per player etc. Additionally, if we were to do this project again, spatial data would likely lead to better and more helpful conclusions. For example, instead of analyzing the type of serve for success rate, we could have used an X-Y plane to plot where each serve landed and use its coordinates as a predictor. The location of the serve is likely more significant than the type of serve. Although the type of serve did turn out to predict faults, this wouldn't be as helpful for a player because you always get a second chance after a fault (it doesn't affect the game that much if you don't double fault). Overall, we believe our data is a good start for studying the metrics of racquetball and by analyzing more players' skill sets we can determine true predictors of success in racquetball.