USING RANKINGS DATA TO PREDICT SUCCESS IN MEN'S PROFESSIONAL GOLF

Aaron Koenigsberg[1,2*], Jarred Pilgrim[2] and Joseph Baker[1]
[1]York University, [2]Golf Australia
email: ark99@yorku.ca

Introduction

The Official World Golf Rankings (OWGR) are a measure of career success in men's professional golf, with the attainment of a Top 50 ranking often regarded as a significant career milestone. Producing players capable of reaching top-ranking positions is a strategic priority for golf's National Sporting Organisations (NSOs) which see elite representation as an effective means to boost participation and attract funding. However, little evidence-based information exists on the typical pathways taken to reach golf's Top 50. Currently, only two peer-reviewed studies by Koenigsberg, Pilgrim and Baker (2020) and Koenigsberg, Pilgrim and Baker (2022) have explored this pathway. Yet while these studies describe top players' ranking pathways, they did not discriminate between players that reached the Top 50 and those who did not. In golf, the age and time to ranking milestones may be used by NSOs to determine emerging players' potential. Thus, more information on the key differences between successful and less-successful players can increase the accuracy of athlete selection and investment decisions, thereby maximising return on investment. Recent work in tennis has examined the age at which players from different peak career ranking groups (i.e., top 10, 11-20, 21-50, 51-100, 101-200, and 201-300) reached career ranking milestones. Results showed that early ranking milestones had a high value in forecasting the peak career ranking of higher-ranked (e.g., Top 10) and lower-ranked (Top 201-300) players (Li et al. 2020). The objective of the current study was to extend previous work by investigating the discriminatory potential of professional golf ranking milestones.

Methods

The names and peak career rankings of all players were retrieved from OWGR lists from 1990 to 2020. Players' date of birth and date of turning professional was retrieved from the public domain. For inclusion into the final sample, players must have attained at least a Top 500 ranking at some point during their career and be between the ages of 30 and 42 as of December 31st, 2020. The target sample of players were assigned to one of six groups according to their peak OWGR, 1-50 (n = 77), 51-100 (n = 51), 101-200 (n = 104), 301-300 (n = 119), 301-500 (n = 193). To examine players' ranking pathways, data was compiled on the date that they reach career ranking milestones. Ranking milestones were based on previous benchmarking work in golf (Koenigsberg et al., 2020) and included the date of first OWGR, top 1000, 750, 500, 400, 300, 200, 100, 50. The dates were then used to calculate the players' chronological age at each of the ranking milestones. The ages at which golfers declared professional status were also calculated. The first set of analyses aimed to compare the ranking pathways of players in different peak ranking groups. To test for differences in ranking pathways between regions, multiple one-way analysis of variance (ANOVA) tests were performed on all dependent variables ($p < 0.01$). Where a significant main effect was found, post-hoc comparisons using Tukey or Games-Howell were performed. The second set of analyses employed Chi-squared automatic interaction detection (CHAID) classification trees to model the relationship between the same ranking variables and peak ranking group. To prevent overfitting of the tree, a minimum of 10 cases were required in order for a node to split. A significance level of $p < 0.05$ was also required for the inclusion of a given variable. Five-fold cross-validation of all CHAID models constructed was undertaken using 70% of observations in the training set and 30% in the testing set at each fold.

Results and Discussion

Top 50 players reached ranking milestones at significantly earlier ages than all other ranking groups. Top 51-100 players reached each ranking milestone from first ranked to top 500 at significantly younger ages than Top 301-500 athletes. While at the top 300 and 400 ranking milestones, Top 51-100 players were significantly younger than players from the top 201-300 group. Similarly, the same relationship was seen for Top 101-200 players, as they also reached each ranking milestone from first ranked to top 500 at significantly younger ages than Top 301-500 athletes and were significantly younger than top 201-300 players at the top 300 ranking milestones. Results from the CHAID analysis revealed that the Top 500 ranking benchmark was the most accurate at classifying group membership. For instance, players who reach the Top 500 before the age of 24.03 were correctly classified as Top 50 players 84.0% of the time. Alternatively, players reaching the Top 500 after the age of 24.03 were correctly classified as Top 301-500 players 77.2% of the time. From a practical perspective, both analyses can be used to inform player selection initiatives and career mapping for developing players. The group comparison results can be used to form a 'performance funnel'. Performance funnels can be used to map the pathways of currently developing players within the Interquartile range of previous milestone data to monitor player progression along the ranking pathway. Further, while the group comparison offers many data points, the CHAID analysis offers a more parsimonious model with fewer dimensions. Such outputs are easier understood by coaches and stakeholders, while also allowing for probabilistic decisions to be made.

Significance

From a pragmatic approach the rankings data may be useful for discriminating between players who will make it to higher echelons of performance (i.e., top 50) or those who will not achieve a top 300 ranking. That is, NSOs seek to invest in high potential athletes at earlier stages of their professional careers while attempting to avoid wasting resources on 'lower potential' athletes. However, while the model produced is fairly accurate at discriminating between high and low-end ranked professional golfers, the rankings are only one variable and selection decisions should be increasingly comprehensive, based upon both other objective data and the knowledge of coaches/stakeholders. Future work should seek to incorporate additional data points to the current rankings model to create an increasingly comprehensive selection model. With a lack of objective data to inform selection decisions in golf, the current work begins to offer additional data to increase the accuracy of athlete selections in golf.