Predicting and Understanding the Risk for Shoulder and Elbow Injuries in Major League Baseball Pitchers: A Game-Theory-Based Machine Learning Approach

Jacob Francis Oeding[1], Alexander M Boos, Aaron John Krych[2], Joaquin Sanchez-Sotelo[2], Jonathan D Barlow[2], Christopher L Camp[2]
[1]School of Medicine, Mayo Clinic Alix School of Medicine, [2]Mayo Clinic

INTRODUCTION:
Reliable prediction of future shoulder or elbow injuries in Major League Baseball (MLB) pitchers based on current demographic, workload, and pitching metrics can help modulate a player's current routine to reduce the risk of future injury. Understanding interactions between multiple risk factors is important to identify potential avenues by which risk can be reduced while minimizing impact on player performance. To better understand these interactions, a novel game-theory-based approach was taken to develop a machine learning model capable of predicting shoulder or elbow injuries in MLB pitchers. The purpose of this study was to develop a machine learning model predictive of next-season shoulder and elbow injuries in MLB pitchers and use this model to understand interdependencies and interaction effects between the most important risk factors.
METHODS:
MLB pitcher demographics, workload measures, injury data, and ball tracking pitch metrics from 2017 to 2022 were used to construct a database of MLB pitcher-years, where each item in the database corresponded to a pitcher's information and metrics for that year and whether the pitcher was placed on the injured list (IL) for a shoulder or elbow injury the following season. A machine learning model was trained to predict next-season shoulder and elbow injuries utilizing Shapley additive explanation (SHAP) values to quantify feature importance as well as interdependencies and interaction effects between predictive variables.
RESULTS:
A total of 3,808 pitcher-years were included in this analysis, and 606 (15.9%) of these involved a shoulder or elbow injury resulting in placement on the injured list. The machine learning model correctly predicted whether a pitcher would sustain a shoulder or elbow injury the following season with an accuracy of 84% (95% confidence interval [CI]: 0.83-0.85). Player demographics and workload metrics were much less predictive of injury compared to ball tracking metrics. Out of more than 65 candidate features, the most important contributors to predicting shoulder or elbow injury were increased: pitch velocity (of all pitch types), utilization of sliders, fastball spin rate, and fastball horizontal movement.
DISCUSSION AND CONCLUSION:
Ball tracking metrics were more predictive of future injury than the player demographics and workload metrics analzysed in this study, with the greatest predictors of injury being: increased pitch velocity (all pitch types), more frequent use of sliders, higher fastball spin rate, and increased fastball horizontal movement. Using these features, a machine learning model was able to predict next-season shoulder and elbow injuries in MLB pitchers with excellent accuracy (84%). Analysis of SHAP dependence plots revealed strong feature interdependencies among predictive features, with a pitcher's average fastball velocity demonstrating the strongest interdependencies with other predictive features. In addition, analysis of SHAP interaction values demonstrated strong interaction effects among some of the most important predictors of shoulder and elbow injury, which included the following: a higher FB velocity did not alter a younger pitcher's predicted risk of shoulder or elbow injury as substantially as it did for older pitchers, the risk for shoulder or elbow injury increases with the number of high velocity pitches thrown (regardless of pitch type and in an additive fashion), and average FB velocities below 95 mph demonstrated strong, negative interaction effects with higher SL percentages, suggesting that the overall predicted risk of injury for pitchers throwing a high number of SLs could be attenuated by throwing with a lower average FB velocity.