The American Journal of Sports Medicine
http://ajs.sagepub.com/

Predictors of Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers

David Whiteside, Douglas N. Martini, Adam S. Lepley, Ronald F. Zernicke and Grant C. Goulet

Am J Sports Med published online May 6, 2016
DOI: 10.1177/0363546516643812

The online version of this article can be found at:
http://ajs.sagepub.com/content/early/2016/05/04/0363546516643812

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
American Orthopaedic Society for Sports Medicine
AOSSM

AJSM PreView, published on May 6, 2016 as doi:10.1177/0363546516643812

Predictors of Ulnar Collateral Ligament Reconstruction in Major League Baseball Pitchers

David Whiteside,*†‡§ PhD, Douglas N. Martini,‖ PhD, Adam S. Lepley,¶ PhD, ATC, Ronald F. Zernicke,†#** PhD, and Grant C. Goulet,† PhD
Investigation performed at the University of Michigan, Ann Arbor, Michigan, USA

Background: Ulnar collateral ligament (UCL) reconstruction surgeries in Major League Baseball (MLB) have increased significantly in recent decades. Although several risk factors have been proposed, a scientific consensus is yet to be reached, providing challenges to those tasked with preventing UCL injuries.

Purpose: To identify significant predictors of UCL reconstruction in MLB pitchers.

Study Design: Case control study; Level of evidence, 3.

Methods: Demographic and pitching performance data were sourced from public databases for 104 MLB pitchers who underwent UCL reconstruction surgery and 104 age- and position-matched controls. These variables were compared between groups and inserted into a binary logistic regression to identify significant predictors of UCL reconstruction. Two machine learning models (naïve Bayes and support vector machine) were also employed to predict UCL reconstruction in this cohort.

Results: The binary linear regression model was statistically significant ($\chi^2(12) = 33.592$; $P = .001$), explained 19.9% of the variance in UCL reconstruction surgery, and correctly classified 66.8% of cases. According to this model, (1) fewer days between consecutive games, (2) a smaller repertoire of pitches, (3) a less pronounced horizontal release location, (4) a smaller stature, (5) greater mean pitch speed, and (6) greater mean pitch counts per game were all significant predictors of UCL reconstruction. More specifically, an increase in mean days between consecutive games (odds ratio [OR], 0.685; 95% CI, 0.542-0.865) or number of unique pitch types thrown (OR, 0.672; 95% CI, 0.492-0.917) was associated with a significantly smaller likelihood of UCL reconstruction. In contrast, an increase in mean pitch speed (OR, 1.381; 95% CI, 1.103-1.729) or mean pitches per game (OR, 1.020; 95% CI, 1.007-1.033) was associated with significantly higher odds of UCL reconstruction surgery. The naïve Bayes classifier predicted UCL reconstruction with an accuracy of 72% and the support vector machine classifier with an accuracy of 75%.

Conclusion: This study identified 6 key performance factors that may present significant risk factors for UCL reconstruction in MLB pitchers. These findings could help to enhance the prevention of UCL reconstruction surgery in MLB pitchers and shape the direction of future research in this domain.

Keywords: Tommy John; injury; surgery; pitch count; elbow; machine learning

As of May 2015, at least 113 Major League Baseball (MLB) pitchers have undergone at least 1 ulnar collateral ligament (UCL) reconstruction surgery (also known as Tommy John surgery) in the decade beginning in 2010.[57] This number already exceeds that of the previous decade, 2000 to 2009 (101 pitchers), and is considerably higher than what was recorded during the 1990s (33 pitchers). More alarmingly, this UCL injury "epidemic," as it has been labeled by some researchers,[5,18] affects pitchers at all levels of baseball.[6,22] Considering that UCL reconstruction surgery has been reported to impose an average of 16.8 to 20.5 months away from MLB competition[12,34] and possibly diminish performance,[28] there is an obvious need to curtail the current injury trend and protect the integrity of the elbow joint in pitchers at all levels.

In light of the above statistics, sports medicine research has afforded considerable attention to UCL injuries in baseball since the turn of the century. Investigations of

*Address correspondence to David Whiteside, PhD, Tennis Australia, Olympic Boulevard, Melbourne, VIC 3000, Australia (email: david.whiteside@gmail.com).
†School of Kinesiology, University of Michigan, Ann Arbor, Michigan, USA.
‡Game Insight Group, Tennis Australia, Melbourne, Australia.
§Institute of Sport, Exercise and Active Living, Victoria University, Melbourne, Australia.
‖Department of Neurology, School of Medicine, Oregon Health and Science University, Portland, Oregon, USA.
¶Department of Kinesiology, University of Connecticut, Storrs, Connecticut, USA.
#Department of Orthopaedic Surgery, University of Michigan Medical School, Ann Arbor, Michigan, USA.
**Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan, USA.
The authors declared that they have no conflicts of interest in the authorship and publication of this contribution.

The American Journal of Sports Medicine, Vol. XX, No. X
DOI: 10.1177/0363546516643812

youth and adult pitchers have variously reported that factors such as pitch count[32,33] (number of pitches thrown in a game), innings pitched per season,[18,32] pitch speed,[4,11] throwing mechanics,[1,8,47] fatigue,[23,25,33,40] and throwing breaking pitches (during youth years)[32,44] may relate to UCL injury risk. The latter has since been refuted, with elbow loading in the curveball proving lower than the fastball[11,21,39] and showing no significant relation to injuries.[18] The link between fatigue and injury remains inconclusive, with elbow loading having been shown to either remain stable[14] or decrease as a game progresses.[38] Mechanically, research has reported that peak elbow valgus, a posture that applies tensile force to the UCL, occurs when the shoulder transitions into internal rotation at the commencement of the arm acceleration phase.[19] However, this appears to be a nonmodifiable risk factor as this posture is an inevitable precursor to velocity generation in maximal-effort overhead projectile movements.[3,15,19,20]

The geographic location of a pitcher's high school has recently been proposed as another factor that relates to injuries,[13] under the premise that pitchers who spend their formative years in warmer climates tend to throw more pitches during their youth and, therefore, are predisposed to UCL rupture earlier in their career. As opposed to longer term factors, pitch counts per game arguably garner the most attention from managers and sports medicine personnel, with an arbitrary maximum of 100 pitches per game commonly adhered to in MLB.[7,54] Despite the advancements that this body of research provides, it is apparent that a consensus on injury risk is yet to be reached, with the best evidence of this perhaps provided by the incidence rate.

The increasing trend of UCL reconstruction surgeries at the MLB level emphasizes the need to improve injury prevention protocols for this at-risk population. With the intention of identifying significant risk factors, the objective of the current study was to examine how demographics and performance factors relate to UCL reconstruction in MLB pitchers. Although this surgery is generally successful and allows pitchers to return to the sport,[5,9,24,27,41,42] it was envisaged that identifying risk factors may enhance the current prediction, prevention, and management of UCL injuries in this population. In accordance with previous research, it was hypothesized that reduced time between games, greater pitch speed, and greater throwing volumes would be significantly associated with UCL reconstruction surgery in MLB pitchers.

METHODS

Before the commencement of the project, an institutional review board approved the experimental protocol. An initial search for MLB pitchers who had undergone UCL reconstruction surgery since 2010 identified a cohort of 113 pitchers, along with each pitcher's surgery date (referred to as "index date").[57] To be included in the study, a pitcher was required to have pitched at least 10 innings in MLB competition in the 12 months preceding his last game before surgery (referred to as "index year"). In the case where a pitcher underwent follow-up surgery, only his first surgery was considered for analysis. These delimitation criteria yielded a cohort of 104 pitchers who composed the surgery group in this study.

A custom MATLAB (MathWorks) code was composed in the software's native programming language to execute the analyses in this study. MATLAB was designed to accommodate large-scale numerical computation and has web-based data mining and sophisticated analytic capabilities. As such, it was conducive to harvesting, transforming, and analyzing the data in this study.

Each pitcher's unique Fangraphs player identification (FangraphsID) number was manually garnered from his player page on the Fangraphs website.[16] The MATLAB code then iterated these FangraphIDs into a uniform resource locater to harvest demographic and playing information (date of birth, height, mass, innings pitched, fielding-independent pitching [FIP]). The same code also identified the last game in which the pitcher competed prior to surgery, before parsing his game logs to quantify games played and games started during his index year. If the pitcher started over 90% of the games he competed in, he was classified as a starting pitcher; otherwise, he was classified as a relief pitcher. The number of days between consecutive games played was also calculated from the game log data (excluding consecutive games either side of the off-season).

The second section of the MATLAB code identified a matched control for each pitcher in the surgery group. To qualify as a suitable control, a pitcher had to (1) have been born within 365 days of his counterpart, (2) have pitched at least 10 innings in MLB during his counterpart's index year, (3) play the same position as his counterpart (starting or relief), (4) have never undergone UCL reconstruction surgery, and (5) not have been already selected in the control group. A sample of suitable controls was initially identified for each pitcher and a single control then selected at random. Using the same index year as his counterpart, each control pitcher's demographic and playing information was subsequently harvested using the same method described above.

After both the surgery and control groups had been identified and their demographic and playing data harvested, the third section of the MATLAB code harvested ball tracking data from the MLB Advanced Media (MLBAM) PITCHf/x database.[37] Ball tracking data first became comprehensively available in MLB after the competition-wide installation of the PITCHf/x system (Sportvision) in 2008. This provided the rationale for only investigating surgeries from 2010 onward (which allowed every pitcher's index year to include a full quota of PITCHf/x data). PITCHf/x records 4-dimensional (time, x, y, z) ball tracking data for every pitch thrown in MLB, with a reported system error of <1.60 km/h and <1.02 cm. These data are then uploaded to the MLBAM website at the end of each day and stored in XML format. Pitch type and physical pitch parameters were harvested directly from this website and used to reconstruct the ball trajectories of 318,117 pitches. The initial y-

TABLE 1
Variables Analyzed in This Study

Age[a] (at index date)
Height[a]
Mass[a]
Position
Innings pitched in index year
Fielding-independent pitching in index year[a]
Number of pitches in repertoire[a]
Mean days between games[a]
Mean pitches per inning[a]
Mean pitches per game[a]
Mean pitch speed[a]
Mean spin rate[a]
Mean horizontal release location[a,b]
Mean vertical release location[a,b]

[a]Served as a predictor variable in the logistic regression and machine learning models.
[b]Release locations normalized to standing height.

displacement was also adjusted to a more realistic location of 16.764 m from the home plate to maximize accuracy of the reconstructed trajectory.[17,35,50] A series of equations was then solved to calculate the release speed of the ball as well as the vertical (z) and horizontal (x) positions of the ball at release (which were normalized to standing height). The horizontal release location values were normalized to a right-handed pitcher.

From the data harvesting process, 14 variables of interest (ie, predictor variables) were ultimately collected and/or calculated for analysis in this study (Table 1). FIP was selected as the criterion measure of pitching performance, as it controls for competition-wide differences in fielding ability (which are beyond the pitcher's control). Consequently, FIP isolates performance outcomes that are exclusively within the pitcher's control, thereby providing the most representative measure of his effectiveness.[26] A lower FIP is indicative of better performance. While the 18 possible pitch types were initially grouped into "hard," "breaking," and "off-speed" categories,[52] an initial mixed-model analysis of variance (within-group factor: pitch category; between-group factor: UCL reconstruction) revealed no interaction effects for pitch speed, spin rate, or release location, thus permitting pitches to be grouped together for the purpose of this study.

A binary logistic regression revealed how the predictor variables affected the likelihood that pitchers underwent UCL reconstruction. Because strong correlations between independent variables can distort regression models, a correlation matrix was initially constructed to detect collinearity. Using a standard threshold[10] of $|r| > 0.7$, it was determined that position, innings pitched during the index year, and mean pitches per game were all highly correlated, and thus, the former 2 predictors were excluded from the regression analysis. The binary logistic regression yielded standardized beta coefficients (β) and odds ratios (ORs) with 95% CIs, which were used to interpret each predictor variable's contribution to the model, using a significance level of $P < .05$. These statistical procedures were performed in SPSS Statistics 21 (IBM Corp).

To provide further novel, predictive insights in the UCL injury problem, the 12 predictor variables entered into the logistic regression were used to train 2 machine learning models: (1) a naïve Bayes classifier and (2) a linear support vector machine classifier. These supervised machine learning approaches determine the most probable outcome associated with a given set of predictor variables. Successful applications of naïve Bayes classifiers include filtering spam email (based on text within the email)[2] and, more pertinent to the current study, diagnosing medical conditions[30,43,49] (based on symptoms, history, demographics, etc). Support vector machines have been used to classify proteins[53,56] and for facial recognition.[48] In this study, the classifier was predicting whether a pitcher would undergo UCL reconstruction surgery based on his demographics and performance metrics across the previous 12 months.

To develop a machine learning classifier, predictor and response data from the 208 pitchers were first randomly partitioned into 5 "folds" (ie, partition), with an even split of control and surgery pitchers in each fold. The machine learning model was then trained using 4 of the folds (ie, the "training set" of ~166 pitchers) and cross-validated on the fifth fold (ie, the "test set" of ~42 pitchers). To optimize the model, an exhaustive (ie, brute force) approach was adopted, whereby every possible subset ($N = 2^{12} - 1 = 4095$) of predictors was used to train a classifier. Each of the 4095 trained classifiers was cross-validated on the test set (unseen to the model during training) to determine how accurately it classified pitchers (as "surgery" or "no surgery"). The optimized classifier was that which produced the greatest classification accuracy. This process was repeated 5 times, such that each fold served as the test set once (Figure 1).

The mean predictive accuracy of these 5 models was then used to evaluate the overall efficacy of classification. The frequencies with which each predictor variable appeared in the optimized model were also computed to identify those variables that were most valuable for classifying pitchers. All machine learning was conducted in MATLAB.

RESULTS

Descriptive demographic and data are presented according to group and position in Tables 2 and 3. The binary linear regression model was statistically significant ($\chi^2(12) = 33.592$; $P = .001$), explained 19.9% of the variance in UCL reconstruction surgery, and correctly classified 66.8% of cases (Table 4). According to the model, fewer days between games, fewer pitch types employed, more pitches per game, greater pitch speed, a less pronounced horizontal release location, and a smaller stature all increased the likelihood of undergoing UCL reconstruction surgery.

According to the ORs (and, as such, assuming all other factors remain equal), an increase in mean days between consecutive games (OR, 0.685; 95% CI, 0.542-0.865) or number of unique pitch types thrown (OR, 0.672; 95% CI, 0.492-0.917) was associated with significantly lower odds

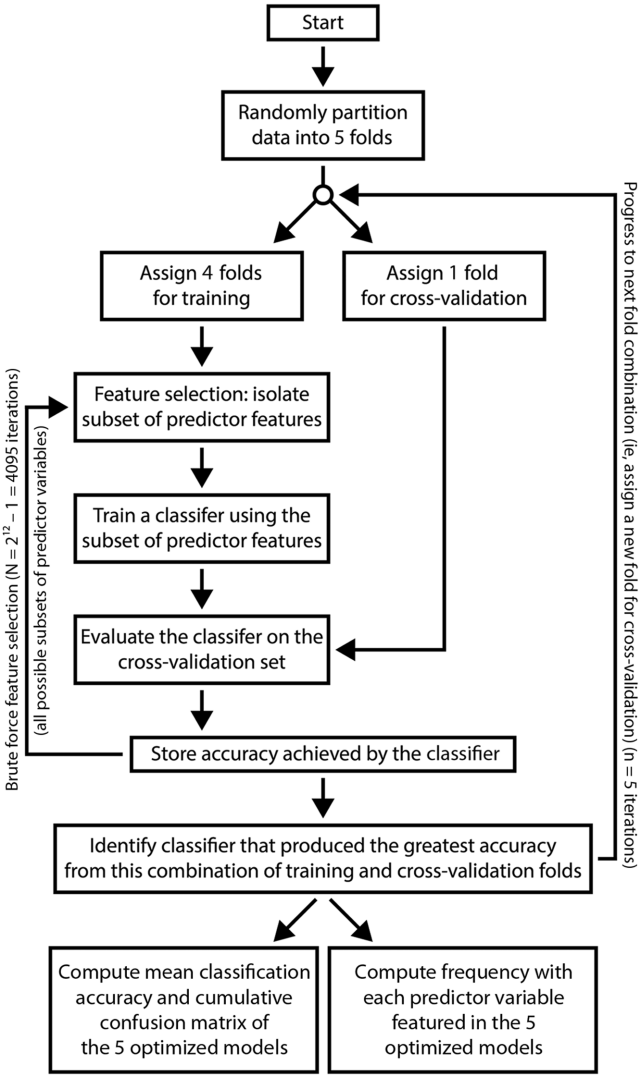

Figure 1. The machine learning process.

of UCL reconstruction. Likewise, an increase in standing height (OR, 0.941; 95% CI, 0.897-0.988) or mean horizontal release location (OR, 0.028; 95% CI, 0.001-0.642) was associated with a significantly smaller likelihood of surgery. On the contrary, the likelihood of undergoing UCL reconstruction surgery increased by 38% for every unit (m/s) increase in mean pitch speed (OR, 1.381; 95% CI, 1.103-1.729) and increased by 2% for every unit increase in mean pitches per game (OR, 1.020; 95% CI, 1.007-1.033).

The classification accuracy of the naïve Bayes classifier was 72% (Table 5) and the support vector machine classifier 75% (Table 6). The most frequently selected feature in the optimized machine learning models was mean days between consecutive games, which appeared in 9 of the 10 models. The next most prevalent features all appeared in 7 of the 10 models and were (1) pitches in repertoire, (2) mean pitch speed, and (3) horizontal release location.

DISCUSSION

This study sought to identify significant predictors of UCL reconstruction in MLB pitchers. Mean days between consecutive games, number of pitch types in repertoire, mean pitch speed, mean pitch count per game, stature, and mean horizontal release location were all identified as significant predictors of UCL reconstruction. These findings are instructive for MLB stakeholders, who could possibly use these data to better identify at-risk pitchers and/or to modify pitchers' activity to reduce the risk of catastrophic UCL injury.

The current findings support the theory that recovery periods after MLB pitching bouts are critical for reducing injuries, a point previously specified at the youth level[29] and part of the rationale for multiple-pitcher rotations in MLB.[36] On average, starting and relief pitchers who underwent UCL reconstruction surgery had 0.8 and 0.7 fewer days, respectively, between consecutive games compared with controls. Pitchers' practice activity was not captured in this study, so it would be inappropriate to describe these results as 1 extra day of "rest." Nevertheless, these data equated to controls being afforded 12% to 19% more time between competitive games and offer potential implications for clubs and policy makers in MLB. Most decisively, if an extra day between games reduces UCL injuries, the value of 6-man pitching rotations (as opposed to the conventional 5) and/or larger bullpens is obvious. Current MLB policy, a finite talent pool, and an overarching pursuit of success renders both concepts unattainable at present. However, this finding warrants consideration as MLB stakeholders continue to explore solutions to this injury problem.

While previous research has investigated the potential injuriousness of individual pitch types,[11,21,39] no study has considered the size of a pitcher's repertoire in the context of injury risk. The OR indicated that a pitcher's odds of undergoing UCL reconstruction surgery decreased by 33% for each unique pitch type that he possessed in his repertoire (all other factors remaining equal). More descriptively, the repertoire of the average control pitcher contained 0.4 more pitches than a pitcher who underwent UCL surgery. "Identifying and preventing overuse" has been proposed as the key to avoiding injuries in pitchers.[31] Considering that overuse injuries are thought to be the result of repetitive microtrauma to connective tissue,[46] the mechanical differences between pitch types[14,21] may dictate that throwing a greater variety of pitches thwarts repetitive, uniform loading of the UCL and helps to explain this finding. However, this remains a hypothesis that future research should look to test. Practically, it would be overly simplistic to suggest that an MLB pitcher can readily add a new type of pitch to his repertoire, leaving junior pitchers the most logical beneficiaries of these data in that they remain in a position to develop variety.

Horizontal release location, perhaps this study's best surrogate for mechanics, was also a significant contributor to the regression model and was less pronounced in pitchers who underwent UCL reconstruction surgery. Considering that PITCHf/x does not factor in the pitcher's starting

TABLE 2
Participant Demographics[a]

	Starting Pitchers (n = 74)		Relief Pitchers (n = 134)	
	Surgery	Control	Surgery	Control
Age, y	27.3 ± 3.8	27.8 ± 3.7	28.3 ± 3.3	28.7 ± 3.4
Height, cm	191.0 ± 4.9	190.2 ± 5.7	188.9 ± 5.2	191.1 ± 6.0
Mass, kg	97.8 ± 8.5	98.1 ± 6.4	98.6 ± 10.0	100.0 ± 9.7

[a]Values are expressed as mean ± SD.

TABLE 3
Descriptive Data for Starting and Relief Pitchers in Each Group During the Index Year[a]

	Starting Pitchers (n = 74)		Relief Pitchers (n = 134)	
Variable	Surgery	Control	Surgery	Control
FIP[b]	3.97 ± 0.74	4.15 ± 1.02	3.96 ± 1.04	4.12 ± 1.06
Innings pitched	153.02 ± 52.85	155.99 ± 53.81	58.48 ± 33.87	59.33 ± 32.17
No. of pitches in repertoire	5.16 ± 0.83	5.32 ± 1.08	4.42 ± 1.02	4.94 ± 1.01
Days between games[c]	6.09 ± 0.64	6.85 ± 1.96	3.85 ± 1.51	4.58 ± 2.31
Total pitches	2417 ± 806	2469 ± 798	941 ± 533	962 ± 514
Pitches per game[c]	93.86 ± 5.79	92.57 ± 6.84	25.86 ± 21.57	28.41 ± 19.95
Pitches per inning[c]	15.95 ± 1.00	16.07 ± 1.24	16.23 ± 1.39	16.42 ± 1.40
Pitch speed, m/s[c]	39.54 ± 1.12	39.20 ± 1.26	40.30 ± 1.38	39.91 ± 1.26
Spin rate, revs/s[c]	29.29 ± 2.49	28.71 ± 2.89	29.67 ± 4.10	28.79 ± 3.56
Vertical release location[c,d]	0.97 ± 0.05	0.98 ± 0.04	0.96 ± 0.08	0.96 ± 0.07
Horizontal release location[c,d]	0.34 ± 0.09	0.32 ± 0.10	0.34 ± 0.11	0.32 ± 0.12

[a]Values are expressed as mean ± SD. FIP, fielding-independent pitching; revs, revolutions.
[b]A lower FIP is indicative of better performance.
[c]Within-pitcher mean during the index year.
[d]Release locations normalized to standing height.

TABLE 4
Significant Contributors to the Binary Logistic Regression Model[a]

Variable	β Coefficient	Odds Ratio (95% CI)	*P* Value
Height	−0.061	0.941 (0.897-0.988)	.013
Mean days between games	−0.378	0.685 (0.542-0.865)	.002
No. of pitches in repertoire	−0.398	0.672 (0.492-0.917)	.012
Mean pitches per game	0.020	1.020 (1.007-1.033)	.003
Mean pitch speed	0.323	1.381 (1.103-1.729)	.005
Horizontal release location[b]	−3.588	0.028 (0.001-0.642)	.025

[a]The dependent binary variable was surgical reconstruction of the ulnar collateral ligament (1 = surgery; 0 = control). Nagelkerke R^2 = 0.199.
[b]Normalized to standing height.

location on the pitching rubber, these results present 2 possibilities: either the control pitchers (1) stood wider (to the throwing arm side) on the pitching rubber, or (2) released the ball further from the body. While there are no empirical data relating to the former, if the latter is true, then these data are at odds with the findings of Aguinaldo and Chambers,[1] who reported that sidearm pitchers (ie, those with more pronounced horizontal release locations) generate significantly greater elbow valgus torques than overhead pitchers. According to our data, arm slots closer to the midline of the body (ie, closer to "overhead") were more injurious. In light of this inconsistency, there is a need for further scientific examination of the "arm slot," as it relates to injury. This line of inquiry should also seek to examine why taller MLB pitchers in this study exhibited lower odds of undergoing UCL reconstruction surgery.

Our data support previous suggestions that greater pitch speed is a precursor to UCL injuries.[2,40,44] The OR revealed that an MLB pitcher was 38% more likely to undergo UCL reconstruction surgery for every 1-m/s increase in his mean pitch speed (admittedly, an unlikely

TABLE 5
Cumulative Confusion Matrix
Across the 5 Folds of Cross-validation
for the Naïve Bayes Classifier[a]

		Predicted Class	
		Surgery	No Surgery
Actual class	Surgery	80	35
	No surgery	24	69

Sensitivity = 70%, specificity = 74%, accuracy = 72%, precision = 77%, false omission rate = 34%

[a]Number of times (across the 5 cross-validation procedures) that each predictor was featured in the optimized model: height = 4; mass = 3; age = 1; fielding-independent pitching = 2; days between games = 4; pitches in repertoire = 4; pitches per inning = 0; pitches per game = 3; mean pitch speed = 3; mean spin rate = 3; horizontal release location = 4; vertical release location = 4.

TABLE 6
Cumulative Confusion Matrix
Across the 5 Folds of Cross-validation
for the Support Vector Machine Classifier[a]

		Predicted Class	
		Surgery	No Surgery
Actual class	Surgery	78	27
	No surgery	26	77

Sensitivity = 74%, specificity = 75%, accuracy = 75%, precision = 75%, false omission rate = 26%

[a]Number of times (across the 5 cross-validation procedures) that each predictor was featured in the optimized model: height = 2; mass = 3; age = 1; fielding-independent pitching = 3; days between games = 5; pitches in repertoire = 3; pitches per inning = 3; pitches per game = 2; mean pitch speed = 4; mean spin rate = 3; horizontal release location = 3; vertical release location = 2.

increase at the professional level). The most logical explanation for this finding is that elbow loading is reportedly greater in faster pitches.[11,21,39] With performance in mind, this relation creates a predicament for MLB stakeholders, as pitch speed is also positively associated with effectiveness in MLB.[51] Assuming that MLB clubs (and pitchers) are unwilling to compromise performance levels, the reality may be that pitch speed is best classified as a nonmodifiable risk factor in the current landscape.

Consistent with our hypothesis, mean pitches per game during the index year was a significant predictor of UCL reconstruction surgery. This endorses some current concepts relating to pitch counts in individual games, which is arguably the most scrutinized factor in UCL injuries.[40,45] According to the OR, unit (ie, 1 pitch) increases in mean pitches per game over a 12-month period increased a pitcher's odds of undergoing UCL reconstruction surgery by 2%. Although less influential than the abovementioned risk factors, this has obvious implications for those devising pitching strategies or administering training doses. Considering our findings above, future research could probe whether the risk posed by increasing pitch counts can be offset by affording an MLB pitcher a longer recovery period before his next game. This would help to inform the application of these findings—and, potentially, policy in the sport—specifically, whether teams should restrict pitch counts or expand their rotations. To that point, the ORs revealed that a pitcher would have needed to throw an average of 16 fewer pitches per game to achieve the same reduction in surgery odds that an additional day between games would have provided.

The mandatory exclusion of position (starting or relief pitcher) from the regression model was a corollary of its expectedly high correlation ($r = 0.81$, $P < .001$) with mean pitches per game. With that in mind, and despite approximately two-thirds of the surgery pitchers in this study being relief pitchers (they are also more prevalent across the league), starting MLB pitchers may be more susceptible to UCL injuries for the simple reason that they throw more pitches per game. This notion is consistent with that proposed by both Keller et al[28] and Olsen et al[40] (notwithstanding that the latter study concerned adolescents). More concentrated examination of the position factor is required, but the combination of this preliminary evidence and the differences in Table 3 (innings pitched, playing schedules, and pitch count) presents the possibility that the injury mechanism is position dependent (see secondary data in Appendix Tables A1-A4, available online at http://ajsm.sagepub.com/supplemental). As such, future researchers and sports medicine staff might refrain from treating starting and relief pitchers homogeneously in the context of UCL injuries.

Finally, it is important to acknowledge that the logistic regression model explained 19.9% of the variance in UCL reconstruction surgery. Accordingly, there exist factors beyond the scope of this study that relate to UCL injuries at the MLB level. Nevertheless, empirically identifying significant predictors of UCL reconstruction and quantifying their relation thereto are positive developments in this space. So too was the development of machine learning models that predicted future UCL surgeries with an accuracy of 72% to 75% using this selection of rudimentary metrics, most of which could be captured at the recreational/youth level. This is a promising revelation that highlights the potential value of both capturing data on a large scale and artificial intelligence in sports medicine.[55] It is expected that supplementing the current predictor features with other intuitive (and currently inaccessible) features such as the pitcher's injury history, practice/conditioning regimens, physical capacities, pitching mechanics, musculoskeletal screening results, and playing history would improve the predictive accuracy of the model and, looking forward, could potentially provide MLB sports medicine personnel with a valuable injury screening tool.

Limiting to this study, only MLB pitchers were investigated. Devoid of empirical confirmation, we would, therefore, discourage the extrapolation of these findings to other (minor league, collegiate, high school, or youth) populations. The current study only appraised public

databases, precluding the inclusion of other factors that relate to injuries (eg, medical records, practice workloads, kinematics and kinetics, and playing history), which future work should incorporate. Equally limiting is that some pitchers in the control group could ultimately undergo UCL reconstruction surgery or potentially already had during periods of their career in which their injuries were not documented in the public domain (eg, youth). This lack of access to medical history also prevented the authors from developing more stringent delimitation criteria (eg, selecting controls who have never reported elbow injuries or have undergone surgery of any kind on their throwing limb), which more controlled studies may be able to overcome. The use of third-party data sources is also a limitation in the sense that they may be subject to inaccuracies and/or inconsistencies (although cross-checks were conducted where possible). Lastly, this study investigated an index period of 1 year (so constrained by the availability of PITCHf/x data), which could be extended in future studies.

CONCLUSION

This study identified significant predictors of UCL reconstruction surgery in a cohort of MLB pitchers. According to the regression model, fewer days between consecutive games, a smaller repertoire of pitch variations, a less pronounced horizontal release location, a higher mean pitch count per game, a smaller stature, and a higher mean pitch speed significantly increased the likelihood of UCL reconstruction at the MLB level. These data imply that recovery after pitching bouts may be just as, if not more, critical to injury prevention as pitch counts. This finding also appears to advocate the appraisal of 6-man rotations in MLB, with respect to UCL reconstruction rates. The revelation that 2 less modifiable factors at the MLB level, pitching repertoire and release location, may be related to UCL reconstruction has implications for how the skill of pitching is taught at the youth level. Pitching with greater speed also presented an injury risk and, in the context of previous research, may indicate a performance-injury trade-off. Finally, the initial accuracy of machine learning models using this selection of rudimentary predictor variables highlights the potential applications of artificial intelligence in identifying at-risk pitchers. However, stakeholders and researchers must petition the uniform collection of large-scale performance and physical capacity data across all levels of baseball for this to become an applicable reality.

REFERENCES

1. Aguinaldo AL, Chambers H. Correlation of throwing mechanics with elbow valgus load in adult baseball pitchers. *Am J Sports Med*. 2009;37(10):2043-2048.
2. Androutsopoulos I, Koutsias J, Chandrinos KV, Spyropoulos CD. An experimental comparison of naive Bayesian and keyword-based anti-spam filtering with personal e-mail messages. In: *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: Association for Computing Machinery; 2000:160-167.
3. Bahamonde R. Review of the biomechanical function of the elbow joint during tennis strokes. *Int Sport Med J*. 2005;6(2):42-63.
4. Bushnell BD, Anz AW, Noonan TJ, Torry MR, Hawkins RJ. Association of maximum pitch velocity and elbow injury in professional baseball pitchers. *Am J Sports Med*. 2010;38(4):728-732.
5. Cain EL, Andrews JR, Dugas JR, et al. Outcome of ulnar collateral ligament reconstruction of the elbow in 1281 athletes results in 743 athletes with minimum 2-year follow-up. *Am J Sports Med*. 2010;38(12):2426-2434.
6. Conte SA, Fleisig GS, Dines JS, et al. Prevalence of ulnar collateral ligament surgery in professional baseball players. *Am J Sports Med*. 2015;43(7):1764-1769.
7. Crotin RL, Bhan S, Karakolis T, Ramsey DK. Fastball velocity trends in short-season minor league baseball. *J Strength Cond Res*. 2013;27(8):2206-2212.
8. Dines JS, Frank JB, Akerman M, Yocum LA. Glenohumeral internal rotation deficits in baseball players with ulnar collateral ligament insufficiency. *Am J Sports Med*. 2009;37(3):566-570.
9. Dodson CC, Thomas A, Dines JS, Nho SJ, Williams RJ, Altchek DW. Medial ulnar collateral ligament reconstruction of the elbow in throwing athletes. *Am J Sports Med*. 2006;34(12):1926-1932.
10. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. 2013;36(1):27-46.
11. Dun S, Loftice J, Fleisig GS, Kingsley D, Andrews JR. A biomechanical comparison of youth baseball pitches: is the curveball potentially harmful? *Am J Sports Med*. 2008;36(4):686-692.
12. Erickson BJ, Gupta AK, Harris JD, et al. Rate of return to pitching and performance after Tommy John surgery in Major League Baseball pitchers. *Am J Sports Med*. 2014;42(3):536-543.
13. Erickson BJ, Harris JD, Tetreault M, Bush-Joseph C, Cohen M, Romeo AA. Is Tommy John surgery performed more frequently in Major League Baseball pitchers from warm weather areas? *Orthop J Sports Med*. 2014;2(10):2325967114553916.
14. Escamilla RF, Fleisig GS, Barrentine SW, Zheng N, Andrews JR. Kinematic comparisons of throwing different types of baseball pitches. *J Appl Biomech*. 1998;14(1):1-23.
15. Eygendaal D, Rahussen FTG, Diercks R. Biomechanics of the elbow joint in tennis players and relation to pathology. *Br J Sports Med*. 2007;41(11):820-823.
16. Fangraphs. Fangraphs baseball. Available at: http://www.fangraphs.com/. Accessed May 1, 2015.
17. Fast M. What the heck is PITCHf/x. *The Hardball Times Annual*. 2010:153-158.
18. Fleisig GS, Andrews JR, Cutter GR, et al. Risk of serious injury for young baseball pitchers: a 10-year prospective study. *Am J Sports Med*. 2011;39(2):253-257.
19. Fleisig GS, Andrews JR, Dillman CJ, Escamilla RF. Kinetics of baseball pitching with implications about injury mechanisms. *Am J Sports Med*. 1995;23(2):233-239.
20. Fleisig GS, Barrentine SW, Zheng N, Escamilla RF, Andrews JR. Kinematic and kinetic comparison of baseball pitching among various levels of development. *J Biomech*. 1999;32(12):1371-1375.
21. Fleisig GS, Kingsley DS, Loftice JW, et al. Kinetic comparison among the fastball, curveball, change-up, and slider in collegiate baseball pitchers. *Am J Sports Med*. 2006;34(3):423-430.
22. Fleisig GS, Weber A, Hassell N, Andrews JR. Prevention of elbow injuries in youth baseball pitchers. *Curr Sport Med Rep*. 2009;8(5):250-254.
23. Fortenbaugh D, Fleisig GS, Andrews JR. Baseball pitching biomechanics in relation to injury risk and performance. *Sports Health*. 2009;1(4):314-320.
24. Gibson BW, Webner D, Huffman GR, Sennett BJ. Ulnar collateral ligament reconstruction in Major League Baseball pitchers. *Am J Sports Med*. 2007;35(4):575-581.
25. Grantham WJ, Byram IR, Meadows MC, Ahmad CS. The impact of fatigue on the kinematics of collegiate baseball pitchers. *Orthop J Sports Med*. 2014;2(6):2325967114537032.

26. Gray AD, Cook JL. Return to pitching after Tommy John surgery: letter to the editor. *Am J Sports Med*. 2014;42(12):NP54.
27. Jiang JJ, Leland JM. Analysis of pitching velocity in Major League Baseball players before and after ulnar collateral ligament reconstruction. *Am J Sports Med*. 2014;42(4):880-885.
28. Keller RA, Steffes MJ, Zhuo D, Bey MJ, Moutzouros V. The effects of medial ulnar collateral ligament reconstruction on Major League pitching performance. *J Shoulder Elbow Surg*. 2014;23(11):1591-1598.
29. Kerut EK, Kerut DG, Fleisig GS, Andrews JR. Prevention of arm injury in youth baseball pitchers. *J La State Med Soc*. 2008;160(2):95-98.
30. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med*. 2001;23(1):89-109.
31. Limpisvasti O, ElAttrache NS, Jobe FW. Understanding shoulder and elbow injuries in baseball. *J Am Acad Orthop Surg*. 2007;15(3):139-147.
32. Lyman S, Fleisig GS, Andrews JR, Osinski ED. Effect of pitch type, pitch count, and pitching mechanics on risk of elbow and shoulder pain in youth baseball pitchers. *Am J Sports Med*. 2002;30(4):463-468.
33. Lyman S, Fleisig GS, Waterbor JW, et al. Longitudinal study of elbow and shoulder pain in youth baseball pitchers. *Med Sci Sport Exerc*. 2001;33(11):1803-1810.
34. Makhni EC, Lee RW, Morrow ZS, Gualtieri AP, Gorroochurn P, Ahmad CS. Performance, return to competition, and reinjury after Tommy John surgery in Major League Baseball pitchers: a review of 147 cases. *Am J Sports Med*. 2014;42(6):1323-1332.
35. Marchi M, Albert J. *Analyzing Baseball Data With R*. Boca Raton, Florida: CRC Press; 2013.
36. Mazzone L, Rosenthal J. *Pitch Like a Pro: A Guide for Young Pitchers and Their Coaches, Little League Through High School*. New York: St Martin's Press; 1999.
37. MLB Advanced Media. MLBAM database. Available at: http://gd2.mlb.com/components/game/mlb/. Accessed May 1, 2015.
38. Murray TA, Cook TD, Werner SL, Schlegel TF, Hawkins RJ. The effects of extended play on professional baseball pitchers. *Am J Sports Med*. 2001;29(2):137-142.
39. Nissen CW, Westwell M, Õunpuu S, Patel M, Solomito M, Tate J. A biomechanical comparison of the fastball and curveball in adolescent baseball pitchers. *Am J Sports Med*. 2009;37(8):1492-1498.
40. Olsen SJ, Fleisig GS, Dun S, Loftice J, Andrews JR. Risk factors for shoulder and elbow injuries in adolescent baseball pitchers. *Am J Sports Med*. 2006;34(6):905-912.
41. Osbahr DC, Cain EL, Raines BT, Fortenbaugh D, Dugas JR, Andrews JR. Long-term outcomes after ulnar collateral ligament reconstruction in competitive baseball players: minimum 10-year follow-up. *Am J Sports Med*. 2014;42(6):1333-1342.
42. Paletta GA, Wright RW. The modified docking procedure for elbow ulnar collateral ligament reconstruction: 2-year follow-up in elite throwers. *Am J Sports Med*. 2006;34(10):1594-1598.
43. Pattekari SA, Parveen A. Prediction system for heart disease using Naïve Bayes. *International Journal of Advanced Computer and Mathematical Sciences*. 2012;3(3):290-294.
44. Petty DH, Andrews JR, Fleisig GS, Cain EL. Ulnar collateral ligament reconstruction in high school baseball players clinical results and injury risk factors. *Am J Sports Med*. 2004;32(5):1158-1164.
45. Popchak A, Burnett T, Weber N, Boninger M. Factors related to injury in youth and adolescent baseball pitching, with an eye toward prevention. *Am J Phys Med Rehabil*. 2015;94(5):395-409.
46. Roos KG, Marshall SW. Definition and usage of the term ''overuse injury'' in the US high school and collegiate sport epidemiology literature: a systematic review. *Sports Med*. 2014;44(3):405-421.
47. Urbin M, Fleisig GS, Abebe A, Andrews JR. Associations between timing in the baseball pitch and shoulder kinetics, elbow kinetics, and ball speed. *Am J Sports Med*. 2012;41(2):336-342.
48. Wei J, Jian-qi Z, Xiang Z. Face recognition method based on support vector machine and particle swarm optimization. *Expert Syst Appl*. 2011;38(4):4390-4393.
49. Wei W, Visweswaran S, Cooper GF. The application of naive Bayes model averaging to predict Alzheimer's disease from genome-wide data. *J Am Med Inform Assoc*. 2011;18(4):370-375.
50. Weinstein-Gould J. Keeping the hitter off balance: mixed strategies in baseball. *J Quant Anal Sports*. 2009;5(2):1-20.
51. Whiteside D, Martini DN, Zernicke RF, Goulet GC. Ball speed and release consistency predict pitching success in Major League Baseball [published online December 15, 2015]. *J Strength Cond Res*. doi:10.1519/JSC.0000000000001296
52. Whiteside D, Martini DN, Zernicke RF, Goulet GC. Changes in a starting pitcher's performance characteristics across the duration of a Major League Baseball game. *Int J Sports Physiol Perform*. 2016;11(2):247-254.
53. Xue Y, Yap CW, Sun LZ, Cao ZW, Wang JF, Chen YZ. Prediction of P-glycoprotein substrates by a support vector machine approach. *J Chem Inf Model*. 2004;44(4):1497-1505.
54. Yukutake T, Yamada M, Aoyama T. A survey examining the correlations between Japanese Little League baseball coaches' knowledge of and compliance with pitch count recommendations and player elbow pain. *Sports Health*. 2013;5(3):239-243.
55. Zelič I, Kononenko I, Lavrač N, Vuga V. Induction of decision trees and Bayesian classification applied to diagnosis of sport injuries. *J Med Syst*. 1997;21(6):429-444.
56. Zhou XB, Chen C, Li ZC, Zou XY. Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes. *J Theor Biol*. 2007;248(3):546-551.
57. Zimmerman J. Disabled list data. Available at: http://www.baseball-heatmaps.com/disabled-list-data/. Accessed May 14, 2015.