The association between elbow injury risk and counter movement jump performance in professional baseball pitchers.

John Mayberry; Bryce Patterson; Scott Mullen; Scott Murayama; Phil Wagner

The incidence rate of Ulnar Collateral Ligament (UCL) sprains has been rising in recent years amongst Major League Baseball (MLB) pitchers. Determining valid screening procedures which allow strength and conditioning coaches to identify pitchers at increased risk of such injuries is therefore of critical importance. This present study seeks to validate the use of countermovement jump (CMJ) tests as a tool in this pursuit. A multinomial logistic model is fit to four years worth of CMJ and arm injury data from pitchers in a single MLB organization. It is shown that elbow injury risk is highly dependent on ground reaction force data obtained during CMJ. In particular, pitchers who rely too much or too little on impulse momentum compared to force production during CMJ are three times as likely to sustain elbow injuries compared to players with more balanced profiles. In contrast, shoulder injury risks appear to be roughly independent of CMJ profiles. This study offers the first predictive link between neuromuscular performance during CMJ and injury risk in professional baseball pitchers.

INTRODUCTION

Major league Baseball (MLB) players have experienced a significant increase in injury rates over the last three decades (7). In particular, there have been considerable increases in the rates of Ulnar Collateral Ligament (UCL) injuries from 33 pitchers requiring UCL reconstruction in the 1990s, compared to 101 UCL reconstructions between 2000 to 2009 (22). Despite considerable research efforts to identify modifiable risk factors for UCL injuries (6,22), injury rates have continued to increase since 2010 and the situation is currently being described as a UCL injury "epidemic" (3,22). Ulnar Collateral Ligament injuries are most common in pitchers (19), occurring with repetitive high valgus torque across the elbow joint during the throwing motion. The majority of this stress is concentrated on the anterior band of the ulnar collateral ligament (8).

Previous studies have shown that the risk of UCL injury may be increased by high acute or chronic pitch volumes (12), pitch speed (2), throwing mechanics (5) and fatigue (9). One posited explanation for these results is that repetitive pitching motions cause an accumulation of micro traumas to the elbow and shoulders leading to increasing risk of ligament injuries (1). The micro traumas occur due to the UCL approaching or exceeding cadaveric failure loads with certain arm positons common to throwing motions in baseball (8,16). Preventing ligament failure requires counteracting the valgus forces placed on the UCL during throwing with other biomechanical factors such as efficient operation of the kinetic chain (5) which transfers force towards the distal segments of the body during throwing movements (10). Deficiency in full activation of the kinetic chain has been shown to be associated with elbow injury risk during throwing (10). For example, internal rotation of the shoulder causes a protective varus force against the valgus force at the
elbow (5). A deficiency in internal rotation ability of the shoulder has been associated (5) with increased risk of elbow injury.

The neuromuscular system functions to ensure proper activation, coordination and transfer of forces through the kinetic chain. Consequently, assessing neuromuscular performance may indicate deficiencies in force transfer through the kinetic chain and subsequent injury risk. Neuromuscular performance is commonly assessed through investigating force production and performance during a counter movement jump (CMJ) assessment. The purpose of this study is to evaluate whether there is a predictive connection between CMJ performance and elbow injury incidence in professional baseball pitchers.

CMJ is among the most reliable and valid forms of jump tests for predicting athletic performance (13,15) and lifting ability (4). Since jumping higher is not the goal of all athletes, recent work has begun to focus on ground reaction force (GRF) profiles of athletes during CMJ. For example, the study (17) showed that a "best three of six CMJ scan" which averages an athlete’s Eccentric Rate of Vertical Force Development (ERFD), Average Vertical Concentric Force (AVCF), Concentric Vertical Impulse (CVI), and Counter Movement Jump Height (JH) across the three maximum height CMJs in a sequence of six is a reliable form of jump testing. This study seeks to extend this work by showing that such CMJ scanning provides a valid method for elbow injury prediction in professional baseball pitchers. While the primary focus is injuries to the elbow, shoulder regions are also included in the analysis for comparative purposes.

**METHODS**

**Experimental Approaches to the Problem**

This study retroactively examines four years (January 2013 - January 2017) of ground reaction force and injury data from pitchers in a single professional baseball organization (both major and minor leagues). Athletes in the organization routinely performed a CMJ scan on a force plate throughout the four year period and this scan data was recorded and stored in a privately operated software management system (Sparta Performance Science, Menlo Park, CA). Injury data was obtained from athletic trainers and coaches at the organization as well as publicly available disabled list data for MLB players. Injuries were classified based on the OSICS 10 code system. For the purposes of analysis, injury types were further grouped together into categories based on location: hip/groin, trunk/abdominal, spine, thigh, shoulder, elbow, lower arm (including forearm, wrist, and hand), lower leg (including knee, ankle, and foot), and other (including head, neck, chest, and upper arm). For this study, only elbow, and shoulder injuries were analyzed because of the small sample sizes ($n < 20$) obtained in the other categories. Injury data was then merged with scan data by averaging all scans of an athlete occurring prior to, but in the same season as an injury. In cases where an athlete obtained multiple injuries in the same season, only scans occurring in between injuries were averaged for subsequent injuries. A control group was formed by averaging seasonal scans for players who did not obtain any
injuries. Principal Component Analysis was performed to identify uncorrelated combinations of scan variables and multinomial logistic regression was used to determine the correlations between the resulting components and injury risk. A Receiver Operator Characteristic (ROC) Curve was then formed to assess the model fit and develop rules for injury risk alerting procedures.

**Subjects**

Subjects for this study were all professional baseball pitchers from a single MLB organization. Overall, a combined 499 player-seasons were included in our analysis from a total of 270 different pitchers. Athlete information was de-identified from both scan and injury data using a randomly generated athlete id. The CMJ data collection process was completed free of injuries and was conducted as part of the athletes' routine testing. Participants were all over the age of 18 and provided consent prior to testing, data collection, and the publication of results as part of their agreement with Sparta Performance Science; as such, ethics approval for this study was not sought. The mean age of players included in the database (as measured from the start of the season) was 23 years old with a standard deviation of 3.782 years. No other identifiable demographic information of participants was reported in this study.

Healthy players completed a mean number of 3.607 (± 2.677 SD) CMJ scans per season while injured players completed a mean of 2.576 (± 2.677 SD) CMJ scans per season prior to their injury. Table 1 summarizes the number of player-seasons in each injury group.

**Table 1. Distribution of injuries within our sample. Counts are in terms of player-seasons.**

<table>
<thead>
<tr>
<th>Injury Location</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Injury</td>
<td>308</td>
</tr>
<tr>
<td>Elbow</td>
<td>53</td>
</tr>
<tr>
<td>Shoulder</td>
<td>65</td>
</tr>
</tbody>
</table>

**Procedures**

Participants performed a series of six CMJs on a commercially available piezoelectric force plate with a sampling frequency of 1000 Hz (9260AA6, Kistler Instruments, Winterthur, Switzerland). Three Force-time variables (ERFD, AVCF, CVI) were extracted from GRF data via numerical integration (11) during both the eccentric and concentric phase of the jump - see (17) for additional details of variable computations and definitions. 30 s were allotted in between successive jumps and measurements from the three jumps with maximal vertical height were averaged to obtain an overall score for ERFD, AVCF, and CVI during the trial. The height (in m) of the three highest jumps was also averaged for the variable JH.

Injury data was merged with scan data by averaging all player scans throughout a season during the period prior to an injury (or for healthy players, throughout the entire season). There were 19 players who received multiple injuries in a season and for these players, only scans occurring in between injuries were averaged for repeat injuries. For example, if
a player had six scans and the first injury took place after the third scan, then the first three scans would be averaged for the first injury and the latter three for the second.

**Statistical Analysis**

All scan variable (ERFD, AVCF, CVI, JH) were first scaled to z-scores for normalization purposes. Only z-scores are reported below due to propriety nature of scan data (Sparta Performance Science). A Principal Component Analysis (PCA) was performed to identify uncorrelated combinations (principal components) of scan variables. The first three principal components accounted for a combined 99% (47%, 34%, and 18%, respectively) of variation in scan variable scores and hence, were the only ones used in subsequent analysis. A multinomial logistic regression models was then fit using injury type (None, Elbow, or Shoulder) as a dependent variables, (standardized) player age as a covariate, and the three principal components as predictors. It was found that a model which included interactions between components performed significantly better than a model including only age (LR = 29.236, df=12, p=0.004) or a model including age and only linear predictors (LR=24.030, df=6, p<0.001). Furthermore, dropping the third principal component did not significantly detract from the model fit (LR=11.180, df=6, p=0.083) and hence, only the first two components were included in the model analyzed in results below. Wald z-tests were used to determine specific components which significantly contributed (p-value < 0.05) to the model fit. All statistical analysis was performed using R: A language and environment for statistical computing (20). Multinomial models were implemented using the nnet package (21).

A plot of false positive vs true positive rates (ROC curve) for different threshold injury probabilities was used to develop early alert procedures for elbow injury detection. Threshold probabilities were chosen at increments of 0.01, starting at 0.01 and ending at 0.5. For a given threshold, all scans with elbow injury probabilities below the threshold were predicted as healthy (test = negative) while all scans with elbow injury risks above the threshold were predicted as injured (test = positive). Each point on the curve shows the true positive rate (fraction of elbow injury scans which tested positive) vs the false positive rate (fraction of healthy scans which tested positive) for the labeled threshold. The point on the curve farthest from the line $y = x$ was used to determine a baseline alert level. The baseline as well as a more conservative and a more liberal cutoff were all assessed for predictive performance using Fisher’s Exact Test.

**RESULTS**

Figure 1 and Table 2 show correlations between all scan variables and the first three principal components. Notice that PC1 has a strong positive correlation with AVCF, ERFD, and JH while PC2 has a strong positive correlation with CVI and JH. Therefore, the relative magnitudes of PC1 and PC2 determine how an athlete relies on force production (ERFD/AVCF) vs impulse momentum (CVI), respectively, during a jump. The signs of PC1 and PC2 determine whether the athlete is actually successful at transferring force and impulse, respectively, to jump height with positive signs indicating higher jumps.
Table 2. Correlations between scan variables and principal components.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ERFD</th>
<th>AVCF</th>
<th>CVI</th>
<th>VJ</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERFD</td>
<td>1.000</td>
<td>0.369</td>
<td>-0.215</td>
<td>0.201</td>
<td>0.644</td>
<td>-0.106</td>
<td>-0.758</td>
</tr>
<tr>
<td>AVCF</td>
<td>0.369</td>
<td>1.000</td>
<td>-0.504</td>
<td>0.605</td>
<td>0.945</td>
<td>-0.020</td>
<td>0.319</td>
</tr>
<tr>
<td>CVI</td>
<td>-0.215</td>
<td>-0.504</td>
<td>1.000</td>
<td>0.361</td>
<td>-0.447</td>
<td>0.866</td>
<td>-0.217</td>
</tr>
<tr>
<td>VJ</td>
<td>0.201</td>
<td>0.605</td>
<td>0.361</td>
<td>1.000</td>
<td>0.612</td>
<td>0.774</td>
<td>0.147</td>
</tr>
</tbody>
</table>

Figure 1. Scatterplots of scaled AVCF vs PC1 colored by jump height (A) and scaled CVI vs PC2 colored by jump height (B)

Table 3. Coefficients, SEs, and Wald’s z-test p-values for the multinomial model by injury type (E=Elbow, S=Shoulder)

<table>
<thead>
<tr>
<th>Term</th>
<th>E.Cof</th>
<th>E.SE</th>
<th>E.p</th>
<th>S.Cof</th>
<th>S.SE</th>
<th>S.p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.933</td>
<td>0.171</td>
<td>0.000</td>
<td>-1.561</td>
<td>0.142</td>
<td>0.000</td>
</tr>
<tr>
<td>z(Age)</td>
<td>0.513</td>
<td>0.144</td>
<td>0.000</td>
<td>0.320</td>
<td>0.132</td>
<td>0.015</td>
</tr>
<tr>
<td>PC1</td>
<td>0.000</td>
<td>0.121</td>
<td>0.997</td>
<td>0.039</td>
<td>0.108</td>
<td>0.717</td>
</tr>
<tr>
<td>PC2</td>
<td>0.157</td>
<td>0.139</td>
<td>0.259</td>
<td>0.049</td>
<td>0.119</td>
<td>0.680</td>
</tr>
<tr>
<td>PC1*PC2</td>
<td>-0.312</td>
<td>0.085</td>
<td>0.000</td>
<td>-0.035</td>
<td>0.087</td>
<td>0.687</td>
</tr>
</tbody>
</table>

Table 3 summarizes coefficients from the fitted multinomial model. Both elbow and shoulder injury risks increased significantly with age with the odds of obtaining an injury more than doubling (predicted increase of $e^{0.513+0.320} \approx 2.300$) for every one standard deviation (about 3 years) increase in age. There was a significantly negative interaction effect between PC1 and PC2 for elbow injuries, but not for shoulder injuries. In particular, the model predicted a 73% ((≈ $100e^{-0.312}$)) decrease in elbow injury odds for each one unit increase in $PC1 \times PC2$. This effect is illustrated in Figure 2 below. Athletes in the lowest tertile of the product $PC1 \times PC2$ were more than three times as likely to sustain elbow injuries than athletes in the highest tertile (probability of injury for Low = 0.172 vs 0.056 for the High group) whereas shoulder injury risks remained roughly constant through all three tertiles.
Figure 2. Empirical injury probabilities for athletes in the lowest, middle, and highest tertile of the product of PC1 and PC2.

A more detailed illustration of the injury model is presented in Figure 3 where the predicted probabilities of injuries are plotted against PC1 and PC2. For elbow injuries, there is a sharp increase in injury risk in the upper left and lower right corners of the heat map due to the strong interactive effect between PC1 and PC2 whereas shoulder injury rates are roughly independent of PC1 and PC2.

Figure 3. Heat Map of predicted injury probabilities based on the model from Table 3. The x- and y-axis limits were chosen to cover the middle 95% of PC1 and PC2 values in the sample.
Figure 4 translates model results for Elbow injuries from principal components back to the measured CMJ variables of ERFD, AVCF, CVI, and VJ. The two corners of high intensity in the heat map correspond to two different types of athletes:

1. (High PC1, Low PC2) those who jump with low impulse momentum (CVI) coupled with high rates of force production (AVCF/ERFD) and
2. (Low PC1, High PC2) those who jump with high impulse momentum coupled with low rates of force production.

Both types of athletes are assigned high injury risk under the model although the increase in injury rates is slightly steeper for the first type.

Figure 4. Sample signatures of four athletes in the sample who obtained elbow injuries and were assigned high risk of elbow injury by the model (prob > 0.35). Elbow 1a and Elbow 1b correspond to Type 1 athletes (low CVI, high ERFD/AVCF) while Elbow 2a and 2b correspond to Type 2 athletes (high CVI, low ERFD/AVCF).

The ROC curve for various injury risk alert procedures is shown in Figure 5. Based on the curve, it is possible to identify three different alert thresholds:

1. p = 0.11 (Baseline)
2. p = 0.09 (Conservative)
3. p = 0.13 (Liberal)

Table 4 further summarizes the classifications for the baseline threshold. Overall, this threshold yielded a false negative rate of approximately 20% with a false positive rate of 41% and performed significantly better than chance guessing (Fisher Exact odds ratio = 5.490, p < 0.001). The more conservative threshold led to an additional reduction of false negatives to 9% albeit at an increased false positive rate of almost 60%. It is interesting to
note that the classifications were significantly better for the earlier years (2013-2014) than the later years (2015-2016) - see Discussion for hypothesized reasons.

**Figure 5.** ROC curve showing the relationship between false and true positive rates for different classification procedures based on elbow injury probabilities obtained from the model. Each point is labeled with the threshold \( p \) used for classifying positive test results with a positive result meaning that the probability of elbow injury was greater than \( p \).

**Table 4.** True Positives (TP), False Negatives (FN), False Positives (FP) and True Negatives (TN) for the baseline model (threshold \( p = 0.11 \)) broken down by time period.

<table>
<thead>
<tr>
<th>Period</th>
<th>TP</th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013-2014</td>
<td>31</td>
<td>4</td>
<td>80</td>
<td>92</td>
</tr>
<tr>
<td>2015-2016</td>
<td>11</td>
<td>7</td>
<td>74</td>
<td>127</td>
</tr>
<tr>
<td>Combined</td>
<td>42</td>
<td>11</td>
<td>126</td>
<td>182</td>
</tr>
</tbody>
</table>

Since the original impetus for the present study was specifically UCL injuries, Figure 6 further breaks down the relationship between principal components for elbow injuries sustained to the UCL (n=15) and other regions of the elbow (n=31 which includes fractures, neuritis, tendinopathy, strains, inflammation, impingement, and loose body syndrome. Note that seven elbow injuries were excluded from this comparison due to insufficient information about the injury type). We see after separating out other injury types, the same strongly negative correlation between the principle components remains for UCL injuries. There was no evidence of a relationship between elbow injury type and the slope of the regression lines between PC1 and PC2 as shown in Figure 5B (\( F_{4,42} = 1.044, p = 0.361 \)). Similar results (or rather lack thereof) held for other breakdowns of injury locations (eg, Strains vs Sprains; Acute vs. Chronic) suggesting that while the general region of an injury can be predicted from jump information, specific mechanisms may be more challenging to pinpoint.
DISCUSSION

The model presented above shows that critical force-time measurements taken from ground reaction force data during CMJ can be used to predict elbow injury risk in professional baseball pitchers. In particular, players who rely too much (or too little) on CVI during CMJ relative to their average rates of force production (EFRD/AVCF) are more at risk of sustaining injuries to the elbow region including critical UCL tears. In contrast, shoulder injury rates were less variable across different jump profiles and hence, appear to be less related to neuromuscular performance during CMJ.

Theories as to why certain types of CMJ scan profiles predict higher elbow injury risks are multi-fold. Baseball pitchers in the first high risk group (low CVI and high ERFD/AVCF) are producing and applying a high amount of force over a short period of time. The consequence of such sharp spikes in vertical force production may be additional tightness in distal joints (such as the elbow) during throwing motions. This tightness may in turn increase the likelihood of acute traumas such as ligament tears or even lead to chronic elbow soreness. Pitchers in the second high risk group (high CVI, low ERFD/AVCF) have the opposite problem; they are producing and applying a low amount of force over a long period of time. They are relying too much on impulse-momentum to compensate for deficiencies in eccentric and concentric strength during throwing. Such prolonged force application in weaker athletes could wear down joints and ligaments over time, again increasing the likelihood of both acute and chronic injuries. Pitchers with more balanced
CMJ scan profiles are more efficient at absorbing the stresses of repetitive throwing motions by distributing them across different neuromuscular mechanisms and hence less likely to obtain elbow injuries.

Further work is warranted to tests these hypotheses and confirm the validity of CMJ testing as a tool for injury prediction. First, the present study focused only on pitchers in a single organization at the highest level of competitiveness. Testing against other levels of play, such as college or elite high school, as well as against larger samples of pitchers at the professional level is needed to fully validate the hypothesized predictive power of CMJ measurements. One concern with our results is the increased rate of false negatives over the latter two years of testing. It is true that the overall number of elbow injuries was much lower in 2015-2016 than in 2013-2014 and this may be due to improvements in strength and conditioning procedures which resulted from the adoption of force-plate jump testing in 2013. As more organizations adopt force-plate testing as an evaluation tool, we hope to answer this question more definitely.

Second, injury mechanism (sprain, strain, or chronic neuritis, for example) was not distinguishable in the present study, likely because the sample sizes for the individual injury types were too small to detect significant differences. While we can say that elbow injuries are predictable from CMJ data, predicting the type of injury would also be desirable. Third, our model ignores the severity of an injury. This was due to the unreliability of "time missed" data in our sample. The use of time missed as a continuous response could lend more power to injury predictions. Finally, our model ignores other injury types such as core, knee, and other lower leg injuries because of small sample sizes for the pitchers in our study. In the future, we would like to investigate not if CMJ measurements can be used to predict such injuries, but also how they correlate with arm injuries.

The challenge of "small data" makes injury prediction a difficult pursuit. CMJ measurements have previously been shown to constitute a reliable set of metrics for athlete conditioning and this present study gives the first (to our knowledge) study of their validity as a tool for injury prediction. In fact, using the more conservative threshold for injury prediction yielded elbow injury detection rates of just over 90%. This increased sensitivity did correspond with a decrease in specificity to under 50% though which highlights the need to develop and investigate low-risk prescriptive measures for individuals with positive test results. One possibility comes from a recent study (14) which investigated the impact of different workout plans on the force-plate variables discussed herein. For example, athletes with low CVI and high AVCF/ERFD (see Elbow 2a and 2b in Figure 4) could experience decreased injury risk through exercise plans focused on improving mobility and sustained force transfer over time (ie. split squats) while athletes with low ARFD/EVCF and high CVI (see Elbow 1a and 1b in Figure 4) could benefit from prescribed plans emphasizing deadlifts and (Olympic style) front/back squats. While "resting" a player based on a test with high false positive rates could carry heavy consequences at a competitive level of play, workout adjustments like those mentioned above are much less risky prescriptions. Overall, this combined reliability, validity, and actionability of CMJ metrics establishes them as a powerful tools for tracking athlete wellness.
REFERENCES


