Investigating Tommy John Surgery With Machine Learning

Nick Seelert and Preston Engstrom

Abstract—**In this study we review the literature surrounding ulnar collateral ligament injury leading to Tommy John surgery. A vast literature review was conducted to outline the issue at hand and possible risk factors involved. A large data collection effort was conducted which resulted in a massive database of interest but no valuable results. We discuss where we went wrong and possible future work.**

I. Introduction

IN 2015 the first ever prevalance study was conducted within all 30 Major League Baseball (MLB) organizations and reported 25% of MLB pitchers had received TJS [1]. Estimates previously had suggested the rate was much lower at 10% [2]. The rapid rise in TJS occurance could be an increase in the number of injuries or simply be related to better understanding of the issue and injury data being more readily available allowing for more accurate measure [3]. Regardless of the causes behind the numbers, the current risk is large with very little known about the risk factors. This makes it difficult for team doctors and pitchers to take the necessary preventative measures to avoid injury. At a minimum the injury is season ending, but it has the potential to alter pitchers career paths and have a financial impact on their teams.

It was reported that Major League Baseball's annual revenue for 2015 was $9.5 billion [4]. Some of the top grossing teams can afford to pay large multi-year contracts exceed $200 million making players expensive assets to their teams. The teams investment can only be realized when the player is healthy and playing. It has been reported that elbow related injuries account for an average of 4451 days lost per MLB season [1]. As such teams wish to mitigate their risk by investing in players with low risk exposure to long term injuries as well as winning performance measures.

Given the high prevalence of TJS observed along with the high monetary value of baseball players, it is of both medical and financial interest to investigate the risk factors associated with players requiring Tommy John surgery. By identifying risk factors it can help to develop preventative measures as well as reduce the risk exposure to teams regarding a player's health.

In this report we present an overview of the ulnar collateral ligament (UCL) and the damage that leads to Tommy John surgery. The problem of UCL injury and Tommy John surgery is then reviewed through previous published work in medical and academic journals. Based on this information present our own analysis of he problem using logistic regression and artifiial networks. Our results are discussed and compared with previous literature. We conclude with suggested future directions for furthering the understanding of the risks in baseball that lead to TJS.

II. Background

A. Tommy John Surgery

The surgery was first performed by Dr. Frank Jobe in 1974 on Tommy John due to irreversible damage done to the ulnar collateral ligament (UCL) in his pitching arm. As such UCL reconstruction surgery has become known as Tommy John surgery. The surgery replaces the damaged UCL in the pitching arm elbow of with a tendon from another body part. Originally the replacement tendon was taken from the forearm, however more recently doctors have favoured harvesting from the hamstring.

The degree to which requiring TJS affects a pitchers ability to return to play varies. However, several studies have reported between 67% to 83% of Major League Baseball (MLB) pitchers who underwent TJS returned to play at the same level [2] [5] [6]. While the prospects for return are promising, it has been reported that it can take upwards of 16 months to return [6]. This is concerning to both the player and their team, especially if the are under a large contract.

It is widely believed by athletes and coaches that performance can actually improveme as a result of TJS [1]. However, increases in performance after TJS can be related to strength gains during rehabilitation or mechanical improvements [5]. The improvements may not even be real as the lower stats prior to the injury could be a result of impaired performance due to smaller injuries leading up to a full UCL tear. [6]. As such, there are no data to support prophylactic UCL reconstruction to improve performance or prevent injuries [2]

B. Ulnar Collateral Ligament

The UCL is a thick triangular ligament where each side of the triangle is a separate ligament (Figure 1). The entire structure is located in the middle of the elbow toward the body. Of the three sub-ligaments, the anterior oblique ligament is the most important in terms of stress and forces. The anterior oblique ligament is commonly what is referred to as the UCL, as such we will also simply refer to the UCL.

Pitching generates high valgus and extension forces across the elbow 2 during late cocking and acceleration phases. Stress is highest on the UCL when the elbow is flexed 90°and the shoulder is rotated away from the body 3 [8]. At this point the anterior bundle of UCL is the primary restraint against valgus stress at the elbow. During this motion it is thought the vagus load approaches the maximum tensile strength of the UCL. It

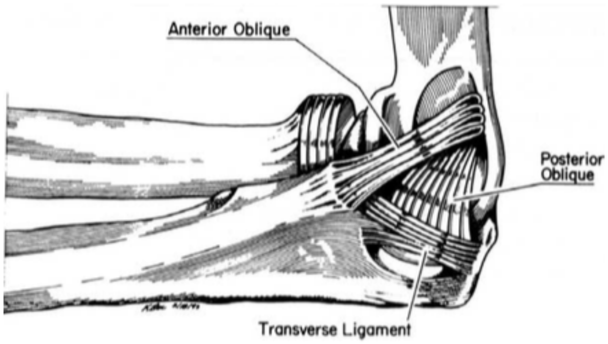

Fig. 1. Three sub ligaments make up the UCL. [3]

is therefore possible that a pitchers throwing elbow gradually leads to failure as a result of repeatedly being exposed to near-failure stresses [5]

Biomechanical studies have shown that moments of force on the shoulder and elbow as well as valgus torque are higher for fastballs than other pitches [9]. This would suggest that pitchers that throw more fastballs could have an increased risk to UCL damage and thus increase chance of requiring TJS.

Given such a high prevlance of of TJS in professional baseball it is frustrating to all parties involved that diagnosing an UCL injury is difficult and likely not made until well after damage is done. One study suggested that diagnosis of UCL damage not made until an average of 6.4 months after symptoms began [5]. One reason is can be so difficult to diagnose is that partial tearing of the UCL is likely indistinguishable from normal elbow laxity on examination. However, a player questionnaire reported that 96% of players acknolwedged pain during the late cocking and acceleration phases of throwing [5]. As such proper diagnosis may require a more qualitative analysis by team doctors with players.

C. Risk Factors

Many risk factors have been speculated including mechanics, pitch type, fatigue, overuse, velocity, and medical issues [9] [8]. However, proper investigation of the issue has been difficult. It has only been recently that the prevelance of TJS

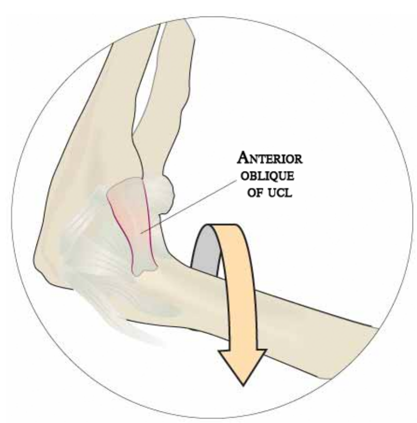

Fig. 2. As arm rotates high valgus stress is generated on the UCL. [7]

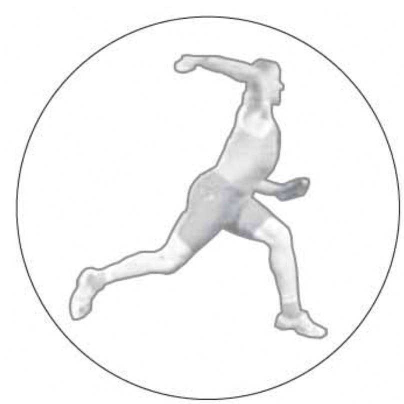

Fig. 3. Greatest valgus stress occurs during the late cocking phase of the pitchers throw. [7]

has been investigated and risk factors are only beginning to be uncovered [1]. The ability to determine risk factors is also limited by studies being retrospective in nature. This limits the ability of researchers to determine changes in factors such as mechanics.

D. Identifying Players at Risk

1) Common Statistics: Several studies have made an attempt to find a relationship between declines in common pitcher statistics and the player requiring TJS. Comparing the pitchers statistics after returning from surgery hows shown a performance decrease, however the significance varies widely [6]. Also, several case-controlled studies failed to find a statistical difference between controls and injury pitchers [10] [2].

It has been argued that common statistics like earned runs average (ERA), batting average against (BAA), and walks plus hits divided by innings pitched (WHIP) are not the best measures to consider when trying to investigate risk factors or predictors for TJS [11]. This is due to the fact that these statistics are also affected by the ballpark the game is played in, the skill of the rest of the defense, luck, and several other factors. Factors more specific a pitchers performance would include fielding-independent pitching (FIP), pitch velocity, pitch type, movement, and other variables tracked for each pitch thrown by a pitcher.

2) Pitch Specific Measures: Every stadium in MLB has been equipped with the Sportvision PITCHf/X system. This system consists of 2 cameras located in the stands above home plate and first base. The purpose of these cameras is to track the flight of every pitch thrown and determine its position, velocity and acceleration [12]. The system is paired with an MLB employee that records further information abouth the pitch such as strike, top and bottom of strike zone, etc.

Studies focusing directly on PITCHf/X data has also been varied. This could be largely due to the fact the classification algorithm used by the PITCHf/X system is constantly being tweaked so comparing data over multiple years can be counfounded by variation inherent in the system [12]. In an attempt to overcome this several groups including the website http://brooksbaseball.net reviews all PITCHf/X data and manually reclassifies errors confirmed by their research.

While there can be a lot of noise present in the PITCHf/X data it is still useful to analyze. A study of 38 pitchers who had undergone TJS reported a small but significant decrease in the velocity of fastballs post injury [10]. However, a more recent study a larger sample size (83) failed to find a significant decrease in fastball velocity [9]. Both studies were case-controlled and reported no significant difference in velocity between groups. However, it was reported that pitchers requiring TJS pitched a significantly higher percentage (7%) of fastballs than controls. This suggests that the number of fastballs thrown is a risk factor but not the velocity of the pitch itself. [9] Another study found decreases were observered in percentage of fastballs thrown and pitches in strike zone up to 2 years after surgery [6]. This may be related to changes in mechanics post surgery.

3) Injury History: A difficult set of factors to investigate in regards to players being at risk of UCL injury is previous injuries to the throwing arm, specifically the elbow. Difficulties are related to the available data on player injuries. Traditionally teams have preferred to hold exact details regarding their player's injuries, therefore data reported publically may not be accurate. However, in more recent years more data surrounding player injuries has become available and is making it easier to investigate.

One study looked at players being placed on the disabled list (DL), a public record of players removed from the teams active roster due to injury. The authors reported before surgery 74% of players were placed on DL because of an injury to their throwing arm and of these 58% were elbow related. After surgery 57% of players were placed on DL because of an injury to their throwing arm, of which 26% were an elbow injury [6]. This study was case-controlled and no significant difference was found between groups for total number of times placed on the DL for injuries related to the throwing arm. However, a higher number of of players requiring TJS went to DL specifically for elbow related injuries [6].

III. METHODOLOGY

A. Data Collection

Data was collected and combined from several sources available online. Standard baseball statistics such as hits, walks, strike outs, etc. were collected from the Baseball Databank which is a publically maintained dataset of historical baseball data. Data was cloned from the Chadwick Baseball Databank GitHub repository [13]. PITCHf/X and pitch count data was scraped from http://brooksbaseball.net using the XML package in RStudio 0.99.893. Players who underwent TJS between 2010-2015 were identified in the Baseball Injury Consultants private database. Access to this database was graciously provided to us by RockFence LLC. Data was obtained from this database using MySQLWorkbench 6.3.6 build 511 CE. Players were referenced by MLB media ID for PITCHf/X and Baseball IC data and by a Baseball Bureau reference ID for stats data obtained in the Baseball Databank. In order to cross reference the different ID types the Chadwick Baseball Bureau Register was used [14]. All data collectoin was conducted on a Macbook Pro with OS X El Capitain 10.11.4.

To construct the data set first a list of potential TJS case pitchers were obtained. This was done by identifying all pitchers labelled in the Baseball IC database as having undergone UCL Reconstruction. The year of surgery was considered as the index year for that pitcher. If a pitcher had multiple TJS only the first index year was selected. For all TJS cases PITCHf/X and statistical data was obtained for the index year and one year prior. Fielding Independent Pitching (FIP) calculations were made using the equation in Figure 4. The FIPs constants were collected from FanGraphs [15]. FIP is an attempt to give a measure similar to ERA but with more emphasis on plays that the pitcher has direct control over. This attempts to give a performance measure for the pitcher while removing the effect a luck and defense has on any individual pitcher [16].

$$FIP = \frac{13 * HR + 3 * (BB + HBP) - 2 * K}{IP} + FIPc$$

Fig. 4. FIP equation: HR is home runs, BB is walks, HBP is batters hit by pitch, K is strike outs, IP is innings pitched, and FIPc is the FIP constant. [16]

B. Input Selection

Inputs were selected based on what was used in previous literature as well as taking into account critiques and improvements [11]. From the statistical data, we selected home runs (HR), walks (BB), batters hit by pitch (HBP), strike outs (K), and innings pitched (IP). As stated above, these statistical variables were used to calculate the FIP score for each pitcher in an attempt to get a comparable performance statistic between pitchers. From the PITCHf/X data we selected average release speed (mph), max release speed (maxmph), horizontal movement (pfx_x), vertical movement (pfx_z), horizontal pitch location (hloc), vertical pitch location (vloc), and grooved pitches (bway) which are pitches straight down the center of home plate.

These variables were selected for each pitch type the player had used in the index year and year prior. Pitch counts were also collected for each pitch type a player had. The number of pitch types each pitcher has, therefore the number of variables with values could vary. To account for this we replaced any NA in pitchF/x and pitch count data with a 0 value. In total there were 156 input variables per player, however as previously stated many of these could be 0 due to pitchers not throwing all pitch types. One pitch type, screwballs, was only found in some control pitchers and was therefore dropped since no TJS pitcher had thrown any in their respective years.

C. Control Selection

Control pitchers were seleced with regard to TJS pitchers. For each index year between 2010-2015 groups were created where each control pitcher had data for that year and the year prior. For selection in analysis the TJS pitchers were first selected and then a control pitcher was selected from their respective bins based on TJS index year. We did no control for pitchers age or throwing arm, however birth year was included as an input.

IV. Experiments

Data was split into testing and training data with a 80/20 hold-out. Equal numbers of TJS and control pitchers were used in the initial dataset before splitting. To analyze the data we conducted a logistic regression using the glm function in RStudio 0.99.893. We also ran a random forest on our data using the Random Forest package 4.6-12. Our choice for random forest modelling was based on previous work done by RockFence LLC. They had limited results so we wished to see if we could replicate their work on our larger dataset.

Our original goal of running the data through an artificial neural network could not be done due to the curse of dimensionality with 156 input variables and not enough available data. In an attempt to reduce the dimensionality we conducted a principal compentent analysis using the prcomp function, however no real advantage was found and we decided to abandon conducting a neural network experiment. As such further investigation needs to be done limit the number of variables used in order to investigate this problem with a neural network.

Given no useful results could be found given such a large number of variables we decided to reduce the dimensions to solely focus on FIP score and 2 pitch types, Cutter and Fastballs. This smaller number of variables was run through the logistic regression and random forest in an attempt to obtain useful results.

V. Analysis

The average age of pitchers was 30 years old and average FIP scores were similar for index year and year prior as seen in Table I. None of these variables showed a significant difference in logistic regression.

TABLE I
Average age and FIP score for pitchers

	Means	SD (+/-)
Age	30	3.8
FIPx	4.37	1.9
FIPy	4.23	1.68

Table II shows a confusion matrix on the prediction of our random forest. An accuracy of 52.38% was reported but this was mostly due to the model failing to predict anything as a TJS. The 95% confidence intervals reported for this accuracy was 29.78% to 74.29%. This large spread also shows the model was essentially unable to properly classify the data.

Given that our data set was unable to provide any valid results we decided to run a random forest analysis on a small dataset previously collected and provided by RockFence LLC. The exact details of how the data were collected were not provided by the company, however verbal confirmation on previous results were provided. As per our conversation with their representative they advised that logistic regression also showed no significant results worth reporting but a random tree show some factors of interest albeit very minimal. As such we attempted to run a random forest on the same data set so we had some valuable data to present in this report.

TABLE II
Confusion matrix for random forest conducted on our dataset

	Reference	
Prediction	0	1
0	8	1
1	9	3

TABLE III
Confusion matrix for random forest conducted on RockFence LLC provided dataset

	Reference	
Prediction	0	1
0	124	5
1	46	16

As seen in table III their model was also a poor predictor of TJS pitchers. This models accuracy was reported as 73.30% with 95% confidence intervals between 66.43% and 79.43%. The high number of correct classifications could be a result of more non TJS pitchers being in the dataset and thus classifying the majority of the pitchers as non TJS will appear to have a high accuracy.

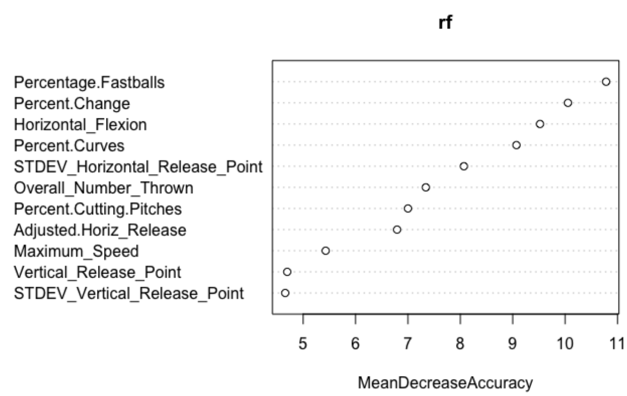

Fig. 5. Mean decrease accuracy showing how variables affect the accuracy of the model

The general idea of mean decrease accuracy as shown in Figure 5 is to permute the values of each feature and measure how much the permutation decreases the accuracy of the model. Therefore, for unimportant variables, the permutation should have little to no effect on model accuracy, while permuting important variables should significantly decrease it.

Figure 6 shows the mean decrease impurity which uses the R function varImpPlot to measure the mean decrease GINI impurity.Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset.

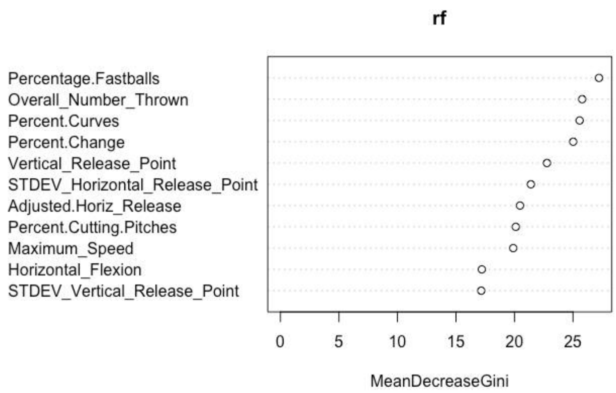

Fig. 6. Mean decrease impurity showing how often elements can affect the model if chosen randomly

The RockFence LLC supplied data corresponds with other previously reported cases that the number of fastballs thrown is an important factor in predicting TJS surgery.

VI. Discussion

Our failure to find results echos much the previous work on the topic of TJS. It remains to be an elusive issue that will continue to require the application of various machine learning algorithms to slowly reveal risk factors. However, our main issue from the start was curse of dimensionality. We tried to collect and analyze too much data which created confusion and no results. Given the large dataset we have already built, our data collection and cleaning processes will be made much easier for deeper and future analysis.

A limiting factor for pitch specific data is that the PITCHf/X system was only installed in MLB stadiums starting in 2008. As such, the most valuable data pertaining specifically to pitchers performance is very limited. On top of this problem is the constant updating and tweaking to the PITCHf/X system itself can create errors and variance within the data. As such it is reasonable to believe that we are still several years away from collecting enough data to make reliable predictions based on this data.

Given the small data set it another alternative would be to work on trends within the season leading up to the injury. Trends may be identified related to fatigue and overuse that may elucitate identifiably risk factors. This would be an extremely time consuming process as data would need to be pulled for every pitcher for every game and then analyzed. While no single game is likely to be of value, trends within season may be more valuable than season means as we and others have attempted to look at.

An overarching theme to the difficulty in assessing this problem is the lack of data. As mentioned previously, teams tend to withold injury specifics regarding their players and therefore proper classification may be difficult. It is also an issue that there is no data on these pitchers regarding spring training, off season practice, and even practice data. If the most important risk factors related to UCL injury leading to TJS are overuse and fatigue these are key areas that need to be investigated to get a proper assessment of the issue.

VII. Conclusion

Given the lack of results from our dataset it is clear that we attempted to analyze too much data. Due to the magnitude of the dataset we collected the majority of our time was spent trying to clean and combine a massive dataset for analysis. While in theory this appeared to be a reasonable goal it appeared to be unnecessary. The much smaller dataset provided by RockFence LLC also failed to give a good model but was able to reiterate other findings regarding the importance of fastballs.

Future work should take this cue and focus on aspects of a pitchers performance that involves only high speed pitches. It may not be the velocity itself that is off importance but the frequency with which the these high speed pitches are thrown. Pitches of interest would include fastballs, cutters, and change ups. While cutters and change ups have slower velocities due to spin on the ball they may have similar issues regarding the stresses placed on the pitchers elbow as these pitchers start out fast and slow as they approach the plate.

A principal compenent analysis could be run on data only on these pitches to continue to reduce the number of components for analysis. This is likely necessary given the small number of pitchers available for sample selection. A neural network would be of extreme interest and was our initial goal when we set out to complete this project. Unfortunately the data collection and cleansing stages took up the majority of our time and left us unable to reduce the data small enough that it would have made sense to build a proper sized network as we were dealing with small sample size and large variable number.

Overall this is an interesting problem to undertake and is only now being addressed. The most recent work was only published in April 2016 so in the coming years we will hopefully have a better dataset and more accurate predictions of risk factors involved.

References

[1] S. A. Conte, G. S. Fleisig, J. S. Dines, K. E. Wilk, K. T. Aune, N. Patterson-Flynn, and N. ElAttrache, "Prevalence of Ulnar Collateral Ligament Surgery in Professional Baseball Players," *The American Journal of Sports Medicine*, vol. 43, no. 7, pp. 1764–1769, 2015.

[2] B. J. Erickson, A. K. Gupta, J. D. Harris, C. Bush-Joseph, B. R. Bach, G. D. Abrams, A. M. San Juan, B. J. Cole, and A. a. Romeo, "Rate of return to pitching and performance after Tommy John surgery in Major League Baseball pitchers." *American Journal of Sports Medicine*, vol. 42, no. 3, pp. 536–543, 2014.

[3] M. Safran, C. S. Ahmad, and N. S. Elattrache, "Ulnar collateral ligament of the elbow." *Arthroscopy : the journal of arthroscopic & related surgery*, vol. 21, no. 11, pp. 1381–1395, 2005.

[4] Forbes. (2015) MLB Sees Record Revenues For 2015 up $500 million and approaching $9.5 billion. [Online]. Available: http://www.forbes.com/sites/maurybrown/2015/12/04/mlb-sees-record-revenues-for-2015-up-500-million-and-approaching-9-5-billion/print/

[5] E. L. Cain, J. R. Andrews, J. R. Dugas, K. E. Wilk, C. S. McMichael, J. C. Walter, R. S. Riley, and S. T. Arthur, "Outcome of ulnar collateral ligament reconstruction of the elbow in 1281 athletes: Results in 743 athletes with minimum 2-year follow-up." *The American journal of sports medicine*, vol. 38, no. 12, pp. 2426–2434, 2010.

[6] E. C. Makhni, R. W. Lee, Z. S. Morrow, A. P. Gualtieri, P. Gorroochurn, and C. S. Ahmad, "Performance, Return to Competition, and Reinjury After Tommy John Surgery in Major League Baseball Pitchers: A Review of 147 Cases." *The American journal of sports medicine*, vol. 42, pp. 1323–1332, 2014.

[7] N. Y. Times. (2013) Why the Ulnar Collateral Ligament Tears. [Online]. Available: http://www.nytimes.com/interactive/2013/09/18/sports/baseball/why-the-ulnar-collateral-ligament-tears.html

[8] G. S. Fleisig, C. E. Leddon, W. A. Laughlin, M. G. Ciccotti, B. R. Mandelbaum, K. T. Aune, R. F. Escamilla, T. D. MacLeod, and J. R. Andrews, "Biomechanical Performance of Baseball Pitchers With a History of Ulnar Collateral Ligament Reconstruction," *The American Journal of Sports Medicine*, vol. 43, no. 5, pp. 1045–1050, 2015.

[9] R. A. Keller, N. E. Marshall, J.-M. Guest, K. R. Okoroha, E. K. Jung, and V. Moutzouros, "Major League Baseball pitch velocity and pitch type associated with risk of ulnar collateral ligament injury," *Journal of Shoulder and Elbow Surgery*, vol. 25, no. 4, pp. 671–675, 2016.

[10] J. J. Jiang and J. M. Leland, "Analysis of pitching velocity in major league baseball players before and after ulnar collateral ligament reconstruction." *The American journal of sports medicine*, vol. 42, no. 4, pp. 880–5, 2014.

[11] A. D. Gray and J. L. Cook, "Return to Pitching After Tommy John Surgery: Letter to the Editor," *The American Journal of Sports Medicine*, vol. 42, no. 12, pp. NP54–NP55, 2014.

[12] M. Fast, "What the Heck is PITCHf/x?" *The Hardball Times Baseball Annual 2010*, pp. 1–6, 2010.

[13] Chadwick. (2016) Baseball Databank. [Online]. Available: https://github.com/chadwickbureau/baseballdatabank

[14] ——. (2016) Baseball Databank. [Online]. Available: https://github.com/chadwickbureau/register

[15] FanGraphs. (2016) wOBA and FIP Constants. [Online]. Available: http://www.fangraphs.com/guts.aspx?type=cn

[16] ——. (2016) FIP. [Online]. Available: http://www.fangraphs.com/library/pitching/fip/