Distribution Agreement

Michael Leeder April 14, 2020

Valuating MLB Pitchers with Respect to Mechanic Predisposition to Injury

by

Michael Leeder

Evan Saltzman
Adviser

Economics

Evan Saltzman

Adviser

Kelli Lanier

Committee Member

Cassandra Casias

Committee Member

2020

Valuing MLB Pitchers with Respect to Mechanic Predisposition to Injury

By

Michael Leeder

Evan Saltzman

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2020

Abstract

Valuing MLB Pitchers with Respect to Mechanic Predisposition to Injury
By Michael Leeder

In this study, I examined the mechanical feature of knee collapse amongst a cohort of randomly selected Major League pitchers and related it to monetary value. Further, this paper examines the budgetary strategy that teams could employ given new information regarding knee collapse. While the knee collapse's relationship with the amount of time a given player is hurt proved to be inconclusive, the relationship between knee collapse and overall production pointed to a higher value for players who did not exhibit the mechanical feature. As it happened, teams were already paying a premium for players with "clean" mechanics, however that premium was much too high. The need for future study is present, as certain variables could be quantified more precisely in the future.

Valuing MLB Pitchers with Respect to Mechanic Predisposition to Injury

By

Michael Leeder

Evan Saltzman

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2020

Acknowledgements

Many thanks to Dr. Saltzman, who took a lot of time out of his very busy schedule to meet and help me apply what I am passionate about to the field of economics. I would not have been able to accomplish this without him. Also, I want to extend thanks to my committee members, Dr. Lanier and Dr. Casias, for lending me their time for proofreading and listening to my presentation. Finally, I'd like to thank Josh Heenan, for giving me the idea to pursue this nature of research and for always being available to discuss the project.

Table of Contents

i. Introduction

Major League Baseball, and baseball in general, is currently facing a new and continuously developing epidemic. Every year, pitchers are going down with torn Ulnar Collateral Ligaments (UCLs), an elbow injury that usually results in season-ending Tommy John surgery and puts the thrower at a serious risk of never being the same thrower again. Although UCL tears are the most well-known and catastrophic injury, there are many other major injuries that sideline throwers for large amounts of time. Many researchers have tried to find the answer for the epidemic, developing studies and private training facilities across the country that advertise elite training, therapy, and preventative work to maximize one's chances of making it to the Major Leagues. With these training facilities in existence, teenagers are managing to reach higher and higher velocities, as well as facing larger and larger workloads, possibly increasing their risk exposure to torn UCLs and other major injuries. Major League organizations are now faced with an extremely intriguing game where they must try to project arms years into the future and weigh the possibility that they get hurt against the amount of money they plan on paying the prospect.

The purpose of this paper is to measure the risk of a pitcher suffering a UCL tear or another major injury, and how baseball teams can exploit information on risk to make optimal contract offers to pitchers. Pursuing this question is important, as a reliable model could shape a front office's strategy for acquiring free agents as well as their drafting strategy.

Pursuing this purpose will be a two-fold approach. First, it must be established what could be possible risk factors for throwers suffering major injury. The approach to answering this question is to examine the pitching motion from top to bottom, evaluating different mechanical risk factors. In collaboration with Dr. Josh Heenan, who currently serves as the President of Advanced Therapy and Performance, an Omaha, Nebraska and Stamford, Connecticut based training and research facility, I have established several different mechanical queues that could point to major injury risk, including a premature back-knee collapse or hip flexor acceleration. In order to establish whether this mechanical issue is particularly significant in relation to major injury, I construct a random sample of current and former Major League pitchers from the year 2000 until present. I examine whether these pitchers exhibit a knee collapse in their motion and test the effect of knee collapse on performance.

Since the project is two-fold, it is most appropriate to establish hypotheses for each model being pursued. There will undoubtedly be several variables at play in each conclusion, but the belief is that there will be a strategy developed that MLB teams may use in the acquisition of professional level pitching.

In the injury risk model, the hypothesis is that the back-knee collapse and premature hip flexor acceleration will be a significant variable and predictor to time on the injured list, specifically due to major injury. This is due to the mechanical inefficiency produced by the premature movement presents a danger to the rest of the body during the pitching motion.

The question as to whether this is an indicator of acute risk or chronic risk is unanswerable by this study. To teams, the question as to how the injury happens is important for player development, but not necessarily in player acquisition from free agency. Therefore, clubs should not be as concerned as to how to prevent the injury, only whether there is an increased risk on a possible asset. This is because no matter how the pitcher gets injured, any missed time will result in not only lost money in guaranteed contracts, but also possible lost revenue of ticket, concession, and memorabilia sales.

In the valuation model, the hypothesis is that pitchers who are at less risk to injury will be worth considerably more than those pitchers who are at a high risk of injury for any given year. The implications here are very interesting, and once the data is established it will present a dilemma to each club based on their willingness to take on risk. The way this study will show valuation is by the statistic called Wins Above Replacement (WAR), which measures how much more valuable a player is than the average replacement level player. These 'replacement level players' are based on the next AAA player that would be used to fill the position. AAA is the Minor League's highest level, right below the MLB. MLB teams each control several Minor League affiliates and have the freedom to move players up and down throughout the organizational structure. One of the main reasons for movement is injury.

To further expand on the necessity of acquiring players at an efficient price, each club is faced with an annual budget. Some clubs are able to afford a much higher budget than others because of the lack of a salary cap in Major League Baseball. It will be interesting to note exactly how much risk each club will be taking on when acquiring each pitcher and what each team's risk profile could look like if the model were to be applied.

The results of the injury model are inconclusive, while the results of the valuation model provide interesting insight into how MLB teams may value this mechanical inefficiency. Pitchers in this data set that exhibit the mechanical inefficiency do not seem to be more likely to become injured, however they seem to be less productive. In this paper, I will highlight related literature, explain the data and methodology behind the study, and provide the regressions and results.

ii. Related Literature

Although this type of project seems to not have been tackled, there are several different works that help provide the framework for this paper. The most important question that seems to be prevalent in past research is the idea of workload. Workload is discussed in the research of Petty and Andrews et al. about the success rate of UCL reconstruction in a younger cohort. Dr. James Andrews is considered to be one of the cutting-edge orthopedic surgeons surrounding overhead throwing injuries, and was the dedicated student of Dr. Frank Jobe, who performed the first Tommy John surgery. Andrews's group found that workload is a possible risk factor towards needing UCL construction, especially among the elite cohort of younger arms (Petty, 2004).

Velocity is a very well-known variable, perhaps the most well-known amongst casual baseball observers. Pitchers are throwing harder and harder in today's game, and this could theoretically put more stress on a pitcher's arm than it can handle, causing the UCL to be at a greater risk of injury. In the work Keller et al. however, the study found that velocity did not seem to be a significant factor in pitchers requiring UCL reconstruction. They did find, however, that the more fastballs a given pitcher threw, the greater risk they have to need UCL reconstruction. The threshold that they found to be significant was approximately 48%, meaning that once a pitcher was using more than 48% fastballs, they were more at risk to need UCL construction (Keller, 2015).

Mechanics themselves are an interesting line of study for different projects, although none have been identified that discuss the back-knee acceleration in relation to UCL risk. Werner et al.'s work surrounding elbow stress, including the discussion of how much stress that the overhead throwing motion places on the elbow is very important, as it provides a baseline that overhead throwing is inherently stressful on the elbow joint. They do briefly discuss knee angle at release, but this is limited to front knee angle (angle relative to home plate). This variable was found to be insignificant. In this study of the knee, Werner's work is important because it establishes that elbow valgus (stress) does in fact result in enough force to tear the UCL, and also that it doesn't seem to be the front knee that affects stress (Werner, 2002).

As far as risk allocation with budgets is concerned, it is thought that risk allocation in the MLB at the moment is done through long-term contracts. This is because the longer a team has a player under control, the more of a chance that player has to be healthy. This has little to do with their actual mechanics or risk to injury. It is only as an explanation as to how teams are protecting themselves in a market that is becoming more expensive by the year.

iii. Data and Methodology

The data set of all pitchers that have pitched in major league games dating back to the year 2000 were grouped together, separated into starters and relievers, and then randomly sampled. In order to make sure that the entire body of work of pitchers who meet the threshold is taken into account, the entirety of a pitcher's career amongst the measurable data set is being counted as long as the pitcher meets the threshold for at least one season. For example, if pitcher A threw 55 innings of work and made less than 10 starts in 2005 only, but pitched from 2001-2006 in the major leagues, all years from 2001-2006 would be included in the sample.

The cutoff being used to cut down the large sample size of pitchers is the threshold of 100 innings and at least 10 games started in the same season to be considered a starter. For relievers, the threshold is less than 10 games started and more than 50 innings pitched in the same season. These thresholds were used because they are good indicators of throwing a 'complete' season. The need for a complete season is important in this study, especially in the valuation portion, because it provides a baseline of what a pitcher is capable of given a full body of work.

Once the overall population is broken down into that smaller subset, 550 pitchers will be randomly selected. Those 550 pitchers will have their entire career in the subset for testing and their videos and pictures will be examined for the mechanical flaw.

This is only one of a myriad of variables that go into a pitcher's throwing motion. This study will consider many different variables, including differing interpretations of the variable of overall health and workload. As mentioned in the previous section, days on the Injured List will be primarily used as the dependent variable when discussing health, as it is the most efficient way available to measure how long a player is injured in-season.

To best treat the sample in order to create an econometric model that may be able to show the injury risk amongst pitchers, knee collapse will be used as a binary variable. The methodology for coding this variable into the data set was to manually watch video from every pitcher that qualifies for the sample size and mark their mechanics as 0 or 1. Any knee collapse that was considered 'minimal' still qualified as 0, anything that was deemed to be more than minimal was coded 1. The collapse needed to be fairly obvious to the eye in order to be coded as 1. To achieve the most data points in a limited time frame it is assumed that pitchers' mechanics stay relatively the same at the elite level.

Height and weight will be taken into account for some of the pitchers in the sample. Because weight is ever changing, it is impossible to know what a given pitcher's weight is on any particular day, or when the sample is being taken. Despite being an important variable, any regression using height and weight will be kept separate from the main model due to the lack of data points.

After examining the risk factors that could go into a pitcher's likelihood to become injured with major injury, the economics side of the research question must be examined, which is how can Major League organizations use this information in order to make sound contract offers to pitchers. Pitchers receive the most lucrative contract offers. If MLB teams are at an increased risk to lose tens of millions of dollars, it will make them reconsider the monetary value they are willing to pay.

In order to properly measure for risk throughout the data set once the increase or decreased risk of pitchers getting hurt is established, the coefficients of each of the variables in OLS regression will be compared. The first of which, and what was focused on as the dependent variable for the valuation portion of the study, is Wins Above Replacement, otherwise known as WAR. This statistic determines how much better a player is than what's considered 'replacement' level, more commonly known as either a minor-leaguer or the next readily available free agent. WAR differs among positions, as expected offensive and defensive production at each different position varies. As a reference to typical WAR numbers, the average WAR of the sample is 1.00, with the 25^{th} percentile being -0.10 and the 75^{th} percentile at 1.80. As a more personnel-based example, this past year's Cy Young Award winners (given to the best pitcher in each league), Justin Verlander and Jacob DeGrom, had WAR ratings of 7.4 and 7.6, respectively.

For pitchers, WAR is usually determined by R/9 (runs per nine innings) or FIP (fielding independent pitching). Since I am using Baseball-Reference's WAR statistic, it is calculated based on R/9. This particular measurement of WAR is known in the sabermetric community as bWAR (Baseball Reference, 2019).

The measurement of WAR, the amount of wins a player is worth above a replacement level minor league player, is extremely prevalent. This is because it allows teams to know exactly how much better a given player is than the cheapest option. In the scope of this study, it will allow the teams to know exactly what premium is efficient to pay for pitchers who are more efficient mechanically. It could also give information as to what discount is worth the risk when dealing with a player who is valuable but may have an increased risk to injury.

Finally, in both the injury model and the valuation model it is important to note that each team presents a unique situation given resources, equipment, personnel level, and training regimen. Therefore, in the context of this study, which team a given pitcher throws for will be accounted for.

iv. Injury Risk Model

The final random sample generated included 550 pitchers. After matching up these pitchers to their overall careers and putting it in pitcher-year format, the data set included 5,683 pitcher-year combinations. Of these 5,683 pitcher-year combinations, back knee collapse was exhibited in approximately 19% of the sample.

The econometric model for injury risk utilizes the sample population along the following equation, where each beta is representative of a variable discussed under the previous section. For this specific injury model, this equation was used:

Equation 1

$$ILDays = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta' {}_3 Team + \varepsilon_i$$

Holding teams in control is a natural conclusion, as a player's environment could be linked to whether that player is hurt more often. This model is limited, albeit, in the sense that there could be more variables at play than the knee collapse. However, as mentioned in the previous section, baseball biometric data is ever-changing. Players' weights vary across the course of a season of play and could fluctuate as much as 10 or 15 pounds, depending on their environment, travel schedule, health status, and more. The OLS analysis of the model is shown below.

Model 1: Injured List Days	
(Intercept)	7.77
	(6.17)
Age	0.07
	(0.17)
Knee Collapse	-3.95^{*}
	(1.77)
R^2	0.02
Adj. R^2	0.01
Num. obs.	3159
RMSE	39.06

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

Based on the OLS model, there is no significant evidence that knee collapse leads to more time on the Injured List for any given player during any given year. If anything, the model suggests that there is a slight negative relationship between knee collapse and time on the Injured List.

This suggestion is interesting to examine further and is a great topic for future research. To expand on this analysis, different control variables were added where appropriate, even though it reduced the overall sample size. To begin with, the differences in the coefficients of height, weight, and an interaction variable between the two were compared to Model 1. This created a second model with the specifications:

Equation 2:

$$ILDays = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team + \beta_4 Height + \beta_5 Weight + \beta_6 Height * Weight + \varepsilon_i$$

In this model, the sample size is reduced to 1,217 pitcher-year combinations, which is hardly enough to draw an overall conclusion for the entire injury prediction picture. However, the model does cause a significant change in the coefficient of knee collapse, shown below.

Model 2: Injured List Days with Height/Weight	
(Intercept)	-250.21
	(321.58)
Age	0.03
	(0.29)
Knee Collapse	-9.46^{**}
	(3.16)
Height	3.51
	(4.34)
Weight	0.98
	(1.53)
Height*Weight	-0.01
	(0.02)
R^2	0.03
Adj. R^2	0.00
Num. obs.	1254
RMSE	38.24

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

The new coefficient of knee collapse suggests that a presence of a collapse can be related with almost 10 days less on the Injured List than those who do not exhibit the knee collapse. This is contrary to the initial hypothesis that knee collapse leads to pitchers becoming injured more often. Several different factors may contribute to the negative and statistically significant coefficient that is shown. The most glaring is the fact that the binary variable does not necessarily consider the degree in which a pitcher is collapsing their knee.

Moving on in discussing knee collapse in isolation, data was included from Major League Baseball's StatCast system (MLB, 2015). StatCast was installed in all MLB parks beginning in 2015, so this isolates the sample down to only pitchers who were active in 2015. Using the StatCast data, average fastball velocity was selected for the eligible entries in the sample. Fastball velocity could be a significant variable in the sense that the faster a pitcher throws, the more pressure the movement causes in the body. The econometric model used for this is shown below.

Equation 3:

$$ILDays = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta_3' Team + \beta_4 AvgFastballVelocity + \varepsilon_i$$

This model includes 125 unique pitchers, which is a very small sample, however limiting the model to these specifications reduces the coefficient of knee collapse on Injured List days. However, it is not significantly significant, and average fastball velocity itself does not seem to have a significant effect on time on the Injured List.

Model 3: Injured List Days with Pitch Speeds	
(Intercept)	128.76
	(320.99)
Age	-2.02
	(1.99)
Knee Collapse	-3.59
	(13.37)
Average FB Velocity	-0.69
	(3.10)
R^2	0.49
Adj. R^2	0.31
Num. obs.	125
RMSE	50.15

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

Lastly, the dependent variable was adjusted to account for the severity of a given injury. In this instance, any stint on the Injured List that lasted 15 or more days was counted as a serious injury, while any stay on the Injured List shorter than 15 days was classified as a minor injury. It was then tested what effect these variables had on the coefficient for knee collapse. The complete econometric model and results are shown below.

Equation 4:

$$SeriousInjury = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team + \varepsilon_i$$

Model 4: Serious Injury	
(Intercept)	0.04
	(0.07)
Age	0.01^{**}
	(0.00)
Knee Collapse	-0.03
	(0.02)
R^2	0.02
Adj. R^2	0.01
Num. obs.	3159
RMSE	0.42

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

As shown, it can be concluded that there is no statistically significant relationship between knee collapse and the likelihood of sustaining a serious injury among the sample population.

I then returned to the dependent variable of Injured List days and ran Poisson and zero inflation analysis using the same regressors from Equation 1 in attempt to find any possible correlation. There was still nothing of statistical significance among the coefficients enough to warrant a separate conclusion.

Model 5: Poisson	
(Intercept)	2.11^{***}
	(0.05)
Age	0.01^{***}
	(0.00)
Knee Collapse	-0.31^{***}
	(0.01)
AIC	137585.24
BIC	137791.21
Log Likelihood	-68758.62
Deviance	131369.18
Num. obs.	3159

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

Model 6: Zero Inflation	
Count model: (Intercept)	4.15***
	(0.38)
Count model: Age	-0.03**
	(0.01)
Count model: Knee Collapse	-0.19
	(0.11)
Count model: Log(theta)	-0.75***
	(0.07)
Zero model: (Intercept)	2.73***
	(0.41)
Zero model: Age	-0.07***
	(0.01)
Zero model: Knee Collapse	0.28*
	(0.11)
AIC	16162.84
Log Likelihood	-8012.42
Num. obs.	3159

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

In general, the conclusion around the first step in this study is that the knee collapse present in a pitcher's mechanics has no significant effect on injury. As discussed earlier in the section, it is difficult to assume that this conclusion applies to all pitchers across all periods of time due to constraints that presented themselves during this particular study. If given unlimited time and resources, I believe that a future study should include the exact degree of collapse, as well as real-time height, weight and pitch usage data. This will allow the model to accurately consider all of these important minute factors when it comes to pitcher well-being.

v. Valuation Model

Although the injury risk model did not provide overtly conclusive results regarding pitcher well-being, the hypothesis that there is a premium on pitchers who do not exhibit knee collapse still required testing. The first econometric model used tested WAR values against the same regressors shown earlier, as well as including Injured List days as an independent variable.

Equation 5

$$WAR = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team + \beta_4 InjuredListDays + \varepsilon_i$$

Model 7: WAR	
(Intercept)	2.24^{***}
	(0.26)
Age	-0.04^{***}
	(0.01)
Knee Collapse	-0.21^{**}
	(0.08)
Injured List Days	-0.00^{***}
	(0.00)
R^2	0.05
Adj. R^2	0.04
Num. obs.	3156
RMSE	1.66

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

As shown, the WAR of a given pitcher in the sample is about one fifth less if they exhibit the knee collapse. This coefficient was statistically significant within the sample. A result such as this does have a lot of meaning for MLB teams. A fifth less of a win is quite substantial, especially for the sample, where the average WAR was 1.00.

The next step is to establish why knee collapse has a statistically significant effect on WAR, if this effect is just limited to WAR, and whether adding different variables causes more of a decrease, an increase, or a change in statistical significance. To begin, I decided to move away from WAR as an entire unit and more towards individual statistics that have either direct or indirect effects on WAR. The first dependent variable examined was innings pitched.

Equation 6:

$$IP = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team + B_4 ILDays - \varepsilon_i$$

Model 8: Innings Pitched	
(Intercept)	126.74^{***}
	(9.88)
Age	-1.49^{***}
	(0.27)
Knee Collapse	-10.92^{***}
	(2.84)
Injured List Days	-0.17^{***}
	(0.03)
R^2	0.03
Adj. R^2	0.02
Num. obs.	3159
RMSE	62.53

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

This model suggests that pitchers who exhibit the knee collapse tend to throw about 11 less innings than those who do not exhibit the knee collapse. This is equivalent to about one and a half starts less for starters, and approximately 5 appearances for bullpen arms.

Finally, I sought to investigate further why knee collapse leads to a lower WAR and workload. The first dependent variable tested in isolation was walks, as free bases can lead to both more runs (lower WAR) and less innings pitched (higher pitch counts).

Equation 7:

$$Walks = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team - \varepsilon_i$$

Model 9: Walks	
(Intercept)	86.64^{***}
	(6.41)
Age	-0.86^{***}
	(0.11)
Knee Collapse	2.15
	(1.28)
R^2	0.13
Adj. R^2	0.11
Num. obs.	1205
RMSE	15.87

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

The results point to a statistically insignificant increase in walks, but only approximately two more walks for any given pitcher. Although this is an increase, it is not necessarily enough in practice to result in such a massive swing of overall value.

Next, I tested whether it was the lack of strikeouts that caused a drop in productivity. Strikeouts are a pitcher's best weapon, as strikeouts help the pitcher avoid any outcome other than an out. Since many modern hitters swing for primarily power outcomes, the inability to strike out enough batters to offset the risk of extra base hits is an extremely dangerous scenario from a production standpoint. The econometric model was adjusted to examine strikeouts as the dependent variable.

Equation 8:

$$Strikeouts = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team - \varepsilon_i$$

Model 10: Strikeouts	
(Intercept)	200.50^{***}
	(16.85)
Age	-1.97^{***}
	(0.29)
Knee Collapse	-5.56
	(3.37)
R^2	0.12
Adj. R^2	0.10
Num. obs.	1205
RMSE	41.68

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

Again, the model shows that a pitcher does strikeout less batters in a given year, but not at a statistically significant value. At a rate of just over five less strikeouts in a given year, it is unlikely that those random five at-bats resulted in significant change to production.

vi. Salaries and Monetary Implications

Looking at the results from an MLB club's standpoint, the decrease in valuation of pitchers who exhibit the mechanical difference versus those who do not is incredibly valuable. Pitchers who are worth a fifth less of a win should be purchased at a discount. This is reflected within the data set when the variables are regressed onto salary.

Equation 9:

$$Salary = \beta_0 + \beta_1 KneeCollapse + \beta_2 Age + \beta'_3 Team + \beta_4 WAR - \varepsilon_i$$

Model 11: Salaries	
(Intercept)	-9442253.33^{***}
	(961630.08)
Knee Collapse	-408935.47^{*}
	(165605.79)
Age	369215.10^{***}
	(15119.79)
WAR	803722.25^{***}
	(36129.70)
R^2	0.22
Adj. R^2	0.22
Num. obs.	4676
RMSE	4346203.32

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Standard errors are reported in parenthesis.

Team fixed effects are included, but not shown.

While teams already pay a premium for pitchers who do not exhibit this mechanical inefficiency, the question remains of whether this is an appropriate premium, and how this should affect the different teams' acquisition of pitchers. When you compare the coefficients of WAR and knee collapse to salary, it is discovered that teams are overpaying for the increased production. This is massive for 'small-market' teams. Because baseball has a lack of a salary cap, meaning that teams are able to pay as much, or as little, as they wish for each player, there is a large gap between what the teams with the largest budgets are willing to pay versus the teams with the smallest budgets.

For teams with small budgets, if they were to use this model to identify pitchers whose WAR is higher than approximately 0.2 than their next best starter and exhibit knee collapse, it is possible to acquire players cheaper than their 'real' price. For teams that have larger budgets, it is also possible to save money using this method, but the risk associated with a decrease in production may not be worth the saved funds.

vii. Overall Conclusion and Further Discussion

Looking at the analysis as a whole, it is difficult to draw a specific conclusion regarding a player's overall health with respect to knee collapse. Every model showed either insignificant coefficients or resulted in a slightly negative significant coefficient. To say that this means that there is no relationship between knee collapse and health is a stretch, however, due to the opportunities for future study.

Future study could draw more conclusive outcomes if there is a method for measuring the exact angle of knee collapse. If this mechanical inefficiency is truly an inefficiency, then it follows that it causes more damage with a higher degree of inefficiency. Another variable that could be included in a future study with more data available would be real-time height and weight. An organization could accomplish this internally, however sample size and the duration of study would be a concern.

In all, the need for future study is extremely important on pitching injuries. As pitchers continue to throw harder and push the boundaries of what the human body is capable of, especially at an early age, the more exposure to injury they will be. It is paramount for organizations, coaches, and players themselves to take the proper precautions so that they best protect themselves from injury.

Bibliography

Baseball Reference., “Pitcher WAR Calculations and Details.” https://www.baseball-reference.com/about/war_explained_pitch.shtml

Baseball Reference. “MLB Value.” Data Set. https://www.baseball-reference.com/leagues/MLB/2019-value-pitching.shtml

Krautmann, Anthony C., and Margaret Oppenheimer. “Contract Length and the Return to Performance in Major League Baseball.” Journal of Sports Economics 3, no. 1 (February 2002): 6–17. doi:10.1177/152700250200300100[2].

Keller, Robert A., Marhsall, Nathan E., Guest, John-Michael, Okoroha, Kelechi R., Jung, Edward K., Moutzouros, Vasilios. “Major League Baseball pitch velocity and pitch type associated with risk of ulnar collateral ligament injury”. Journal of Shoulder and Elbow Surgery. https://www.researchgate.net/profile/Kelechi_Okoroha/publication/298723498_Major_League_Baseball_pitch_velocity_and_pitch_type_associated_with_risk_of_ulnar_collateral_ligament_injury/links/5b0f479faca2725783f41002/Major-League-Baseball-pitch-velocity-and-pitch-type-associated-with-risk-of-ulnar-collateral-ligament-injury.pdf

Major League Baseball., Baseball Savant. “Custom Leaderboards for StatCast Data.” Data Set. https://baseballsavant.mlb.com/leaderboard

Maxcy, Joel. "Motivating Long-Term Employment Contracts: Risk Management in Major League Baseball." *Managerial and Decision Economics* 25, no. 2 (2004): 109-20. http://www.jstor.org.proxy.library.emory.edu/stable/25151282

Petty, Damon H., James R. Andrews, Glenn S. Fleisig, and E. Lyle Cain. “Ulnar Collateral Ligament Reconstruction in High School Baseball Players: Clinical Results and Injury Risk Factors.” The American Journal of Sports Medicine 32, no. 5 (July 1, 2004): 1158–64. https://doi.org/10.1177/0363546503262166.

Rhoads, Derek. “Pitching Injury Dashboard.” Data Set. https://jagfantasysports.com/injury_dashboards/

Werner, Sherry L., Murray, Tricia A., Hawkins, Richard J., Gill, Thomas J. “Relationship between throwing mechanics and elbow valgus in professional baseball pitchers”. Standford University Biomechanics, Journal of Shoulder and Elbow Surgery., 2002. https://biomechanics.stanford.edu/me337/projects/werner02.pdf