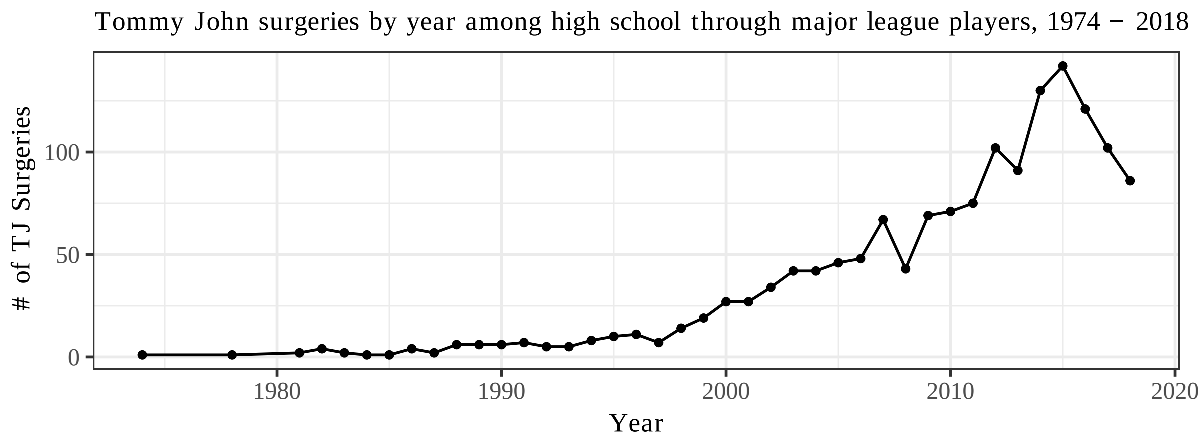
STA 640: Does Tommy John Surgery Make Pitchers Throw Harder?

Lisa Lebovici

December 12, 2018

Introduction

Since the first Tommy John surgery was performed in 1974, Major League Baseball has witnessed an explosion of pitchers undergoing the procedure. Seen as a way to extend players' careers, and the only effective solution for a severe elbow injury, the operation has become more and more common as pitchers begin throwing harder at younger ages. Indeed, the rate of surgeries among minor and major league players has increased at such a rapid pace that every year between 2012 and 2017 had at least as many surgeries as the entirety of the 1990s.



As the number of operations has grown, a myth has emerged that players who receive Tommy John surgery throw faster and harder upon recovery than they did before the injury. Consequently, many young pitchers, as young 15-19 years old, have preemptively sought out the surgery as a way to gain a competitive edge.

This paper will examine the validity of this claim through a potential outcomes framework. Employing both difference-in-differences and principal stratification, we attempt to answer the question as to whether or not pitchers increase their velocity following Tommy John surgery as compared to their pre-injury velocity.

Background

Tommy John surgery is named after the first pitcher to undergo the operation in 1974. It is performed to repair a tear in the elbow's ulnar collateral ligament (UCL), which absorbs a great deal of stress from a pitcher's throwing motion. The surgery works by reconstructing the ligament using a tendom from elsewhere in the patient's body (usually opposite elbow or knee) or from a cadaver. The recovery period is long, often lasting around 12-18 months. However, for pitchers with a UCL injury, there aren't really any other options currently that will give them a decent chance of continuing their careers other than Tommy John surgery.

Previous research has proved inconclusive as to whether or not the procedure improves players' post-surgery velocity. In a 2013 study, Erickson, Gupta, Harris et al. evaluated 148 pitchers between 1986 and 2012 on high-level performance metrics, as pitch-by-pitch data was not available at this time. They found that pitchers were statistically better in ERA, win-loss percentage, and walks/hits allowed, concluding that "performance declined before surgery and improved after surgery" [2]. On the other hand, Jiang and Leland (2014) did not find any significant decreases in the velocity of fastballs, changeups, and curveballs post-surgery when compared with a pair-matched control group without any known injuries [3]. Similar other studies have also found no difference or a slight decrease.

However, in many studies on this topic, the comparison that is being performed is velocity directly prior to the surgery (when a player is injured and naturally throwing slower) to velocity following recovery, which isn't necessarily apples-to-apples. This is usually due to a lack of available data far enough back in time to exclude the data directly before and after the surgery.

Data and Methodology

Pitch data was pulled from the PitchFx database, which is managed by the MLB and contains detailed information on every single pitch from 2008-present via high-speed cameras installed in all MLB stadiums. This data was combined with a database of all players that have had Tommy John surgery, as well as information on players' ages, height, weight, and handedness. When all of the disparate sources were put together, this left us with approximately 7.7M unique pitches.

Our outcome of interest, Y_i, is the average fastball velocity in MPH of a pitcher following surgery. The treatment, Z_i, is taken to be a UCL injury and the resulting surgery. We consider both the injury and surgery to be the treatment due to the fact that, as mentioned above, there aren't really any cases where players would get this particular injury without having the operation. Thus, the control group is all players who have not had a UCL injury and therefore never needed the surgery.

The estimand that we consider is the average treatment effect for the treated:

$$ATT = \mathbb{E}[Y_{i,t+1}(1) - Y_{i,t+1}(0) \mid Z_i = 1] = \theta_1 - \theta_0$$

Seven covariates were incorporated in the model and are described in the table below.

Table 1: Description of covariates

Covariate	Type	Description
age	integer	age at time of TJ surgery
weight	integer	weight according to MLB records
handedness	character	throws lefty or righty
fastestPitch	character	four-seam fastball, two-seam fastball, or sinking fastball
starter	logical	starter or reliever
numPitches	integer	pitches thrown over the measure period
preVelocity	double	average fastball velocity (MPH) pre-treatment

To determine the pre- and post-treatment velocity, we exclude both the year's worth of data directly preceding surgery and following their return to the major leagues, so as to hopefully compare a player's true pre-injury velocity with their post-injury velocity. We then take the year before and after those buffer periods as our measure period, from which we calculate average velocity. For the control units, we impose a hypothetical "surgery date" so that we can have comparable timeframes for pre- and post-operative performance.

We now explore two different methods for computing our estimand, ATT.

Method 1: Difference-in-differences

To use difference-in-differences, three assumptions must hold: (1) SUTVA, (2) overlap, and (3) parallel trend assumption. It seems reasonable to assume that one pitcher's velocity does not affect any other pitcher's velocity, so we can say that SUTVA holds. For overlap, we match every treated unit with 10 control units to ensure that our control sample is representative of the treated group. Finally, we assess the parallel trend assumption by performing a DID analysis over two pre-treatment periods to check whether, in the absence of treatment, we would expect our treated group and control group to follow the same trend over time.

With these assumptions met, we will take a parametric approach to difference-in-differences. The idea behind DID is that we can model both the before-after relationship as well as the treatment-control relationship:

	Before	After
Control	$\boldsymbol{Y_{i,t}(0)}, Y_{i,t}(1)$	$\boldsymbol{Y_{i,t+1}(0)}, Y_{i,t+1}(1)$
Treated	$\boldsymbol{Y_{i,t}(0)}, Y_{i,t}(1)$	$Y_{i,t+1}(0), \boldsymbol{Y_{i,t+1}(1)}$

We can denote $\theta_1 = Y_{i,t+1}(1)$ and $\theta_0 = Y_{i,t+1}(0)$ for the treated group. While θ_1 is observed from the data, we will need to estimate θ_0. We do this in the following way: we can separately model $Y_{i,t}(0) \mid X_i,\ Z_i = 0$ and $Y_{i,t+1}(0) \mid X_i,\ Z_i = 0$ using linear regression, with the outcome models defined below.

$$\begin{aligned} \text{velocity}_i &= \alpha_0 + \alpha_1 Z_i + \alpha_2 \text{time}_i + \alpha_3 Z_i * \text{time}_i + \alpha_4 \text{age}_i + \alpha_5 \text{weight}_i \\ &+ \alpha_6 \text{handedness}_i + \alpha_7 \text{fastestPitch}_i + \alpha_8 \text{starter}_i \end{aligned}$$

We can then estimate the effect from time t to time $t+1$ by applying these models to the treated group and calculating the difference; adding this difference to $Y_{i,t}(0)$ will give us an estimation for θ_0. Finally, taking $\theta_1 - \hat{\theta}_0$ will provide an estimate of ATT.

To estimate θ_0, we take three different approaches: regression, inverse probability weighting (IPW), and a double robust estimator [5]. We first define the regression estimator as follows:

$$\hat{\theta_0}^{reg} = \frac{\sum_{i=1}^n Z_i Y_{i,t}}{\sum_{i=1}^n Z_i} + \frac{\sum_{i=1}^n Z_i(\hat{Y}_{i,t+1}(0) - \hat{Y}_{i,t}(0))}{\sum_{i=1}^n Z_i}$$

Next, we turn to the IPW estimator. Note that inverse probability weights are computed as 1 for the treated group and $\hat{e}(X_i)/(1 - \hat{e}(X_i))$, where $e(X)$ is an estimated propensity score. Using logistic regression, a propensity score model with main effects was fit to the data:

$$\begin{aligned} logit(Pr(Z_i = 1 \mid X_i)) &= \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{weight}_i + \beta_3 \text{handedness}_i + \beta_4 \text{fastestPitch}_i \\ &+ \beta_5 \text{starter}_i + \beta_6 \text{numPitches}_i + \beta_7 \text{preVelocity}_i \end{aligned}$$

This allows us to then define the corresponding IPW estimator:

$$\hat{\theta_0}^{ipw} = \frac{\sum_{i=1}^n Z_i Y_{i,t} w_i}{\sum_{i=1}^n Z_i} + \frac{\sum_{i=1}^n (1 - Z_i)(Y_{i,t+1} - Y_{i,t}) w_i}{\sum_{i=1}^n Z_i}$$

Finally, we can compute the double robust DID estimator by augmenting the IPW estimator with regression:

$$\hat{\theta_0}^{dr} = \hat{\theta_0}^{ipw} + \frac{1}{\sum_{i=1}^n Z_i} \sum_{i=1}^n \frac{(Z_i - \hat{e}(X_i))(\hat{Y}_{i,t+1}(0) - \hat{Y}_{i,t}(0))}{1 - \hat{e}(X_i)}$$

Method 2: Survivor Average Causal Effect via Principal Stratification

A fundamental issue of the DID approach is survivor bias; outcomes are lost for players who don't return to the Major Leagues following the surgery, either because they're still hurt, too old, or simply not competitive enough anymore. To account for this reality, we can borrow from the medical field which uses principal stratification for outcomes truncated by death. We stratify units into four categories according to their counterfactual survival status based on treatment, as shown below.

Table 2: Four possible survival types

	$S_i(0) = 1$	$S_i(0) = 0$
$S_i(1) = 1$	Pitch-pitch (PP) (will return to pitching regardless of surgery)	Pitch - don't pitch (PD) (will only pitch if they get injured)
$S_i(1) = 0$	Don't pitch - pitch (DP) (will only pitch if they don't get injured)	Don't pitch - don't pitch (DD) (will not pitch again regardless of surgery)

We can then define a substitute variable A to enable an estimate of the Survivor Average Causal Effect (SACE), using equations provided by Ding et al (2011).

There are three assumptions that we must make to proceed:

(1) monotonicity: there are no players in the pitch - don't pitch category. In other words, there are no pitchers whose careers were saved by getting a UCL injury and having surgery, but who would have stopped playing in the absence of an injury.

(2) exclusion restriction: the substitute variable is independent of the outcome given treatment, survival type, and covariates (the effect of movement on pitch velocity is already captured, so this is satisfied).

(3) substitution relevance: the substitute variable is dependent on the survival type given covariates. That is to say, pitchers who would continue playing regardless of injury would be expected to have better pitch movement than those who won't continue playing.

Given these assumptions to be plausible, we can now proceed with the SACE estimand, defined accordingly:

$$SACE = \underbrace{\mathbb{E}[Y_i \mid G_i = PP,\ Z_i = 1,\ X_i]}_{\text{directly from data}} - \underbrace{\mathbb{E}[Y_i \mid G_i = PP,\ Z_i = 0,\ X_i]}_{\text{identifiable from assumptions}}$$

Results

Method 1: Difference-in-differences

After combining all of our data sets and filtering for players who had enough data pre- and post-surgery, we had a total of 1,282 units (49 treated and 1,233 control). We used the R package, `MatchIt`, to pair each treated unit with 10 control units using nearest neighbor matching with logistic regression and allowing for replacement. This gave us a final sample of 418 units, 49 treated and 369 control.

After fitting the propensity score model and outcome models defined in the previous section, we can compute the three different estimators for $\hat{\theta}_0$, and finally ATT. Point estimates are shown below, as well as 95% confidence intervals which were calculated via 1,000 bootstrapped samples.

Table 3: Average treatment effect for the treated

	Regression	IPW	Double Robust
$\theta_1 - \hat{\theta}_0$	0.2455	0.2736	0.2735
95% Conf. Int.	(-0.2574, 0.7548)	(-0.2036, 0.7666)	(-0.203, 0.7664)

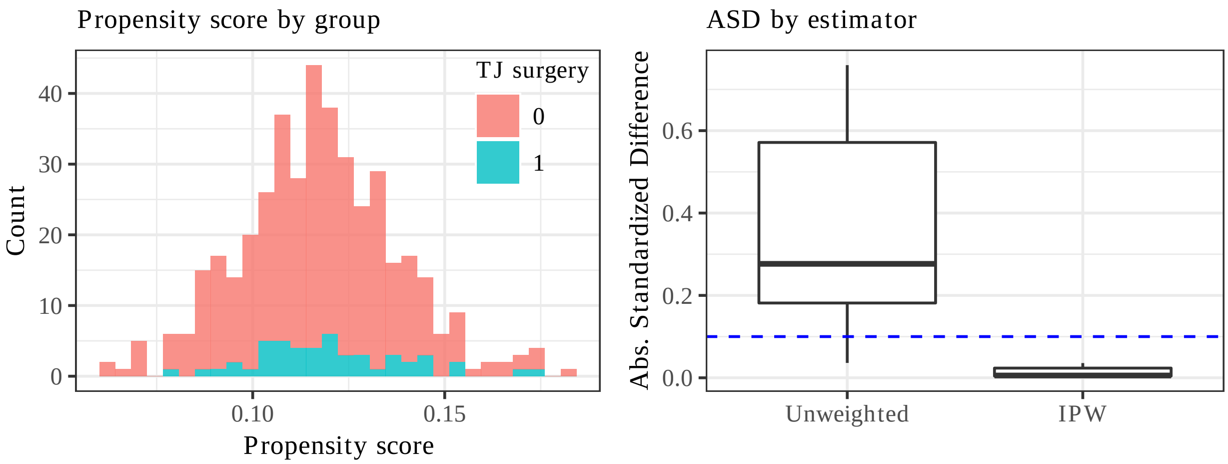
While all of the point estimates are slightly positive, 0 falls well within the bounds of the 95% confidence intervals. This suggests that we are unable to conclude there are any significant differences in pre- and post-fastball velocity before and after Tommy John surgery.

To assess the parallel trend assumption, we repeat this analysis using two pre-treatment groups. If the ratio of θ_1 to $\hat{\theta}_0$ is approximately 1, this indicates that the trend is plausibly the same between the two groups.

Table 4: Assessing the parallel trend assumption

	Regression	IPW	Double Robust
θ_1 / $\hat{\theta}_0$	1.005	1.0058	1.0058
95% Conf. Int.	(0.9998, 1.0105)	(0.9994, 1.012)	(0.9994, 1.0118)

Finally, we ensure that the data is balanced and that we have overlap between the groups' covariates. The distribution of propensity scores for each of the treated and control groups is shown before, and it is clear that overlap has been achieved. This is expected given that we had matched control units to the treated units to ensure that they had similar populations. Additionally, we look at the absolute standardized differences (ASD) between the regression and IPW estimators, and see that IPW has drastically improved the balance as compared to the unweighted estimator.



Method 2: Survivor Average Causal Effect via Principal Stratification

SACE: estimand and confidence intervals

Discussion

References

[1] Ding et al., (2011). Identifiability and Estimation of Causal Effects by Principal Stratification With Outcomes Truncated by Death. Journal of the American Statistical Association 106(496) DOI: 10.1198/jasa.2011.tm10265

[2] 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. 2013;42(3): 536-542.

[3] Jiang, J. J., & Martin Leland, J. (2014). Analysis of Pitching Velocity in Major League Baseball Players Before and After Ulnar Collateral Ligament Reconstruction. The American Journal of Sports Medicine, 42(4), 880–885. https://doi.org/10.1177/0363546513519072

[4] Keller RA, Marshall NE, Guest J-M, Okoroha KR, Jung EK, Moutzouros V. Major League Baseball pitch velocity and pitch type associated with risk of ulnar collateral ligament injury. J Shoulder Elbow Surg. 2016;25(4):671–675.

[5] Li, F. (2018, July). Difference-In-Differences for Causal Inference in Traffic Safety Research. Presented at JSM 2018, Vancouver, British Columbia.

[6] Roegele, J. (2012, November 1). Tommy John Surgery List (@MLBPlayerAnalys). Retrieved December 10, 2018, from https://docs.google.com/spreadsheets/d/1gQujXQQGOVNaiuwSN680Hq-FDVsCwvN-3AazykOBON0/edit#gid=0

[7] Sievert C. (2014). Taming PITCHf/x Data with {pitchRx} and {XML2R} The R Journal, 6(1). URL http://journal.r-project.org/archive/2014-1/sievert.pdf.

[8] Sievert C. (2015). pitchRx: Tools for Harnessing 'MLBAM' 'Gameday' Data and Visualizing 'pitchfx'. R package version 1.8.2.

[9] Wang L, Zhou X, Richardson RS, (2017). Identification and estimation of causal effects with outcomes truncated by death. Biometrika, Volume 104, Issue 3, 1 September 2017, Pages 597–612,https://doi.org/10.1093/biomet/asx034