

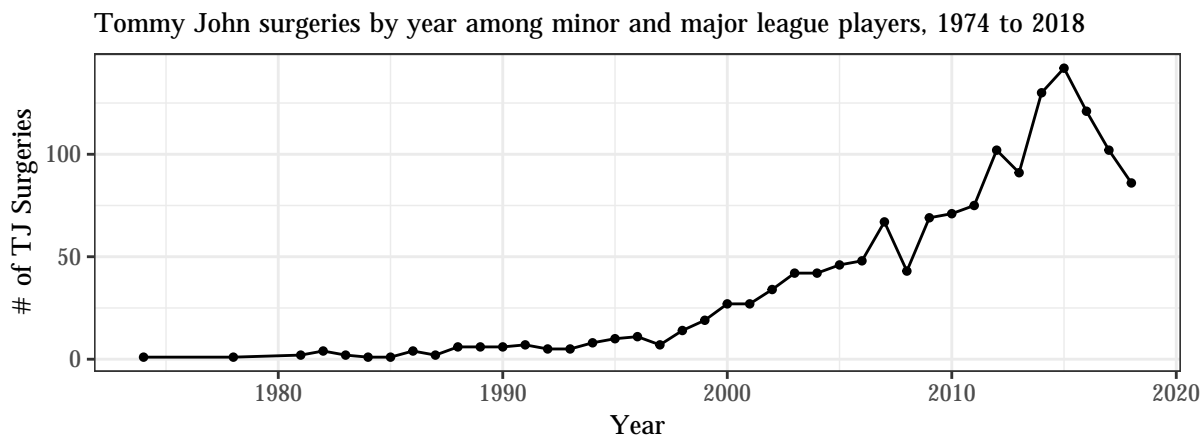
STA 640: Does Tommy John Surgery Make Pitchers Throw Harder?

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 upon recovery than they did before the injury. Consequently, many young pitchers, as young as 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 tendon 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 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

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. The relevant variables are detailed below.

Table 1: Description of variables

Variable	Type	Description
age	integer	age at time of TJ surgery
fastestPitch	character	four-seam fastball, two-seam fastball, or sinking fastball
handedness	character	throws lefty or righty
numPitches	integer	pitches thrown over the measure period
postVelocity	double	average fastball velocity (MPH) in time t+1
preVelocity	double	average fastball velocity (MPH) in time t
starter	logical	starter or reliever
weight	integer	weight according to MLB records

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.

Methodology and Results

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 the injury and surgery jointly as 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. Our covariates are all variables other than post-operative velocity.

We now explore two different methods of estimating the causal treatment effect, difference-in-differences and survivor average causal effect via principal stratification.

Method I: 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.

Under the difference-in-differences framework, our estimand is the average treatment effect for the treated:

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

We must first verify the validity of three key assumptions: (1) SUTVA, (2) overlap, and (3) parallel trend. One pitcher's velocity clearly 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 precisely to ensure that our control sample is representative of the treated group. Finally, we assess the parallel trend assumption which states that, in the absence of treatment, we would expect our treated and control groups 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	$\mathbf{Y}_{i,t}(\mathbf{0})$	$\mathbf{Y}_{i,t+1}(\mathbf{0}), Y_{i,t+1}(1)$
Treated	$\mathbf{Y}_{i,t}(\mathbf{0})$	$Y_{i,t+1}(0), \mathbf{Y}_{i,t+1}(\mathbf{1})$

We denote $\theta_1 = Y_{i,t+1}(1) \mid Z_i = 1$ and $\theta_0 = Y_{i,t+1}(0) \mid Z_i = 1$. While θ_1 is observed from the data, we need to estimate θ_0 . We take three different approaches to this problem. Under the regression method, we separately model $Y_{i,t}(0) \mid Z_i = 0$ and $Y_{i,t+1}(0) \mid Z_i = 0$. We can then estimate the effect from time t to $t+1$ on the treated group, and add this difference to $Y_{i,t}(0) \mid Z_i = 1$ to get an approximation for θ_0 .

$$(Y_{i,t}(0) \mid Z_i = 0) = \alpha + \beta X \quad (Y_{i,t+1}(0) \mid Z_i = 0) = \delta + \gamma X$$

Secondly, we can use inverse probability weights (IPW), which are defined for ATT as 1 for the treated group and $\hat{e}(X_i)/(1 - \hat{e}(X_i))$ for the control group, where $e(X)$ is an estimated propensity score. Thus, we first fit a propensity score model to the data via logistic regression:

$$Pr(Z_i = 1 \mid X_i) = \frac{\exp(\alpha + \beta X_i)}{1 + \exp(\alpha + \beta X_i)}$$

We can now define our three estimators, (1) regression, (2) IPW, and (3) double robust, which is an augmentation of IPW by regression.

$$\begin{aligned} (1) \quad \hat{\theta}_0^{reg} &= \frac{\sum_{i=1}^n Z_i Y_{i,t}}{\sum_{i=1}^n Z_i} + \frac{\sum_{i=1}^n Z_i (\hat{Y}_{i,t+1}(0) - \hat{Y}_{i,t}(0))}{\sum_{i=1}^n Z_i} \\ (2) \quad \hat{\theta}_0^{ipw} &= \frac{\sum_{i=1}^n Z_i Y_{i,t} w_i}{\sum_{i=1}^n Z_i} + \frac{\sum_{i=1}^n (1 - Z_i)(Y_{i,t+1} - Y_{i,t}) w_i}{\sum_{i=1}^n Z_i} \\ (3) \quad \hat{\theta}_0^{dr} &= \hat{\theta}_0^{ipw} + \frac{1}{\sum_{i=1}^n Z_i} \sum_{i=1}^n \frac{(Z_i - \hat{e}(X_i))(\hat{Y}_{i,t+1}(0) - \hat{Y}_{i,t}(0))}{1 - \hat{e}(X_i)} \end{aligned}$$

Point estimates for ATT and bootstrapped 95% confidence intervals are shown below.

Table 2: Average treatment effect for the treated

	Regression	IPW	Double Robust
$\theta_1 - \hat{\theta}_0$	0.2328	0.2736	0.274
95% Conf. Int.	(-0.2624, 0.717)	(-0.2036, 0.7666)	(-0.2066, 0.7623)

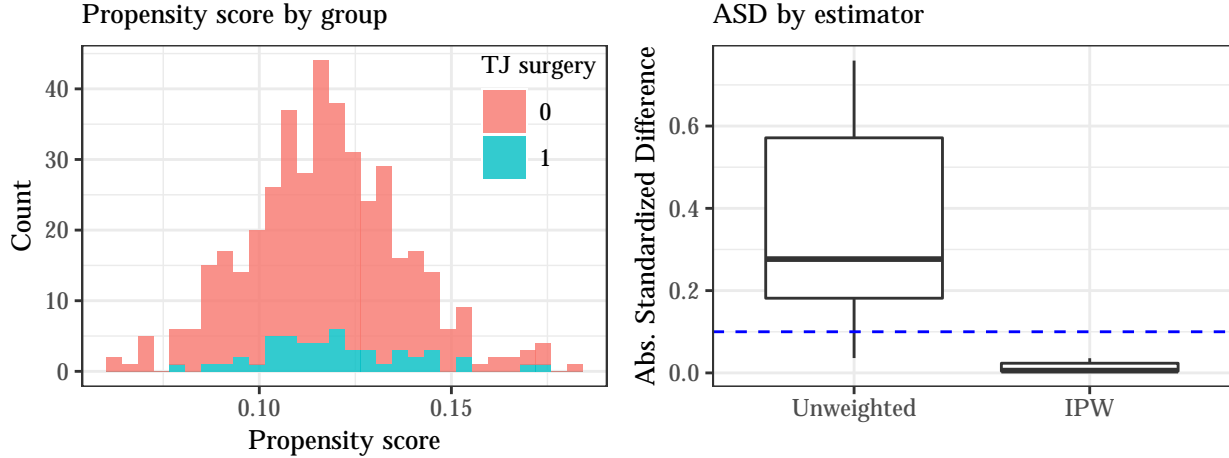
While all of the point estimates are slightly positive, 0 falls well within the bounds of the 95% confidence intervals. This suggests that we are unable to conclude there are any significant differences in fastball velocity before and after Tommy John surgery.

To assess the parallel trend assumption, we repeat this analysis over two pre-treatment periods. As the ratio of θ_1 to $\hat{\theta}_0$ is approximately 1, it is reasonable to assume that the trend is the same between the two groups.

Table 3: Assessing the parallel trend assumption

	Regression	IPW	Double Robust
$\theta_1 / \hat{\theta}_0$	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 below, and it is clear that overlap has been achieved. This is expected given that we had matched 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.



Method II: Survivor Average Causal Effect (SACE) via Principal Stratification

A fundamental issue with the DID approach is survivor bias; outcomes are lost for players who don't return to the MLB following surgery, either because they're still hurt, too old, or no longer competitive. To account for this reality, we borrow from the medical field which uses principal stratification for outcomes truncated by death. Units are stratified into four categories according to their counterfactual survival status based on treatment, where $S_i(1)$ and $S_i(0)$ denote whether they would survive treatment or non-treatment respectively. We define $G = \{S_i(1), S_i(0)\}$ as one of {"live-live" (LL), "live-die" (LD), "die-live" (DL), "die-die" (DD)}.

Table 4: Four possible survival types

	$S_i(0) = 1$	$S_i(0) = 0$
$S_i(1) = 1$	Live-live (will continue pitching regardless of injury / surgery)	Live-die (injury / surgery will save their career, would otherwise stop pitching)
$S_i(1) = 0$	Die-live (injury / surgery will end their career, would otherwise continue pitching)	Die-die (will not pitch again regardless of injury / surgery)

Additionally, we define a substitute variable A that we believe will provide information about a pitcher's true (unknown) stratum but that won't affect the outcome of interest. For our purposes, we take A to be a pitcher's average movement (e.g. horizontal / vertical break of the ball). Pitch movement is not associated with UCL injuries, but it is associated with survival in the sense that pitchers with better movement are more likely to stay in the major leagues than those with poor movement.

There are three assumptions that we must make to proceed:

Monotonicity: In our case, we use a slightly different version, reverse monotonicity. This is the assumption that there are no players in the LD 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.

Exclusion restriction: $A \perp Y \mid (Z, G)$, or as explained above, the idea that A is independent of the target variable given treatment, survival type, and covariates. We are looking at pre-treatment pitch movement, which is inherently independent of post-treatment velocity.

Substitution relevance: $P(A \mid G = LL) \neq P(A \mid G = LD)$, which means that the substitute variable is dependent on the survival type given covariates. That is to say, pitchers are more likely to survive in the MLB in general (regardless of injury) if they have better movement.

Taking these assumptions as plausible, we are interested in calculating the survivor average causal effect (SACE) on the LL subgroup, $SACE = \Delta_{LL} = \mathbb{E}[Y(1) - Y(0) \mid G = LL]$.

By the monotonicity assumption, none of the surviving players can have survival type LD, which means any treated unit must be in the LL stratum. As a result, we can model (1) their post-treatment velocity as a function of the covariates X and substitute variable A ; (2) the probability of continuing to play under non-treatment via logistic regression; and (3) the ratio of probabilities of surviving treatment and non-treatment.

$$(1) \quad \mathbb{E}[Y_i(1) \mid Z_i = 1, G_i = LL, X_i, A_i] = \alpha_0 + \alpha_1 X_i + \alpha_2 A_i$$

$$(2) \quad P(S_i(0) = 1 \mid X_i, A_i) = \frac{\exp(\beta_0 + \beta_1 X_i + \beta_2 A_i)}{1 + \exp(\beta_0 + \beta_1 X_i + \beta_2 A_i)} = P_\beta$$

$$(3) \quad \frac{P(S_i(1) = 1 \mid X_i, A_i)}{P(S_i(0) = 1 \mid X_i, A_i)} = \frac{\exp(\gamma_0 + \gamma_1 X_i + \gamma_2 A_i)}{1 + \exp(\gamma_0 + \gamma_1 X_i + \gamma_2 A_i)} = P_\gamma$$

We observe that for the control units, $P(S_i(0) = 1 \mid X_i, A_i) = P_\beta$ and $P(S_i(0) = 0 \mid X_i, A_i) = 1 - P_\beta$. Similarly, for the treated units, $P(S_i(1) = 1 \mid X_i, A_i) = P_\beta P_\gamma$ and $P(S_i(1) = 0 \mid X_i, A_i) = 1 - P_\beta P_\gamma$. In particular, as $P(S_i(1) = 1 \mid X_i, A_i)$ is the probability of survival given treatment, this tells us a unit's likelihood of being in the LL subgroup. With these four probabilities in hand, we can compute the joint log likelihood $\mathcal{L}(\beta, \gamma \mid X, Y)$ to find the maximum likelihood estimates of β and γ .

The last step is to model $\mathbb{E}[Y_i(0) \mid Z_i = 0, G, X, A] = \alpha_0 + \alpha_1 X_i + \alpha_2 G$. Note that because we are now regressing on G , we don't need to include A as a model variable because of the exclusion restriction. We use P_γ as a proxy for G (which is unobservable), setting $G = 1$ if a unit is more likely to be an LL and $G = 0$ for DL.

Finally, we can estimate $\mathbb{E}[Y(1) \mid G = LL]$ and $\mathbb{E}[Y(0) \mid G = LL]$, and consequently SACE:

$$\mathbb{E}[Y(1) \mid G = LL] = \frac{\mathbb{E}_X[\mathbb{E}[Y(1) \mid Z = 1, G = LL, X] * Pr(G = LL \mid X)]}{\mathbb{E}_X[Pr(G = LL \mid X)]}$$

$$\mathbb{E}[Y(0) \mid G = LL] = \frac{\mathbb{E}_X[\mathbb{E}[Y(0) \mid Z = 0, G = LL, X] * Pr(G = LL \mid X)]}{\mathbb{E}_X[Pr(G = LL \mid X)]}$$

$$SACE = \mathbb{E}[Y(1) - Y(0) \mid G = LL]$$

The point estimate and 95% confidence intervals are 0.2271 and (-0.2631, 0.7825), respectively. Similar to the DID method, we see that the point estimate is slightly positive but 0 falls squarely within the 95% confidence intervals.

Discussion

There does not appear to be any conclusive evidence that Tommy John surgery improves a pitcher's post-operative performance. Both difference-in-differences and SACE via principal stratification led to point estimates that were slightly positive, but with zero in the bounds of error. The results here are generally consistent with the opinions of doctors who perform Tommy John surgery, who tend to state that they are simply returning pitchers to their previous abilities rather than enhancing them. Nonetheless, more research could certainly be done. It would be interesting to add more covariates, such as service time in the MLB, which has been shown to be indicative of various injuries. Additionally, although a brief sensitivity analysis showed that the length of the buffer period did not seem to matter greatly, computational limitations prevented this from being expanded more thoroughly and could be an area of future inquiry.

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] Li, F. (2018, July). Difference-In-Differences for Causal Inference in Traffic Safety Research. Presented at JSM 2018, Vancouver, British Columbia.
- [5] 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>
- [6] 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>.
- [7] Sievert C. (2015). pitchRx: Tools for Harnessing 'MLBAM' 'Gameday' Data and Visualizing 'pitchfx'. R package version 1.8.2.
- [8] 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>