Using Quasi-Experimental Methods to Evaluate Pitcher Performance Post-Tommy John Surgery

Atul Venkatesh [1] Michael Herron [2]

[1]Program in Quantitative Social Science, Dartmouth College [2]Professor of Quantitative Social Science, Dartmouth College

Abstract

While recovery from Tommy John surgery spans over a year, it can take even longer for a pitcher to return to full strength. This research uses both Difference-in-Differences and the Synthetic Control Method to analyze a pitcher's performance after Tommy John surgery.

Setting the Scene

In a fast-paced league, missing an extended period of time would certainly damage a player's development and shorten their career. However, novel technology and recovery methods increase the likelihood to return to form. Enter Carlos Rodón, who underwent Tommy John surgery in 2019 while with the Chicago White Sox. Immediately following surgery, Rodón went from a middle-of-the-rotation pitcher to an ace-caliber all-star. When reflecting on the impact of Tommy John surgery on his success, Rodón said the following:

> *"I don't think [Tommy John Surgery] stalled my career, I think it helped my career 'cause I got to work on some things, like mechanically on things throughout my delivery that got me to where I am now".*

Just how much did Rodón improve? Table A displays his statistics before and after surgery.

	ERA	WHIP	FIP	SO9
2017	4.15	1.370	4.69	9.9
2018	4.18	1.260	4.95	6.7
Tommy John Surgery Recovery				
2021	2.37	0.957	2.65	12.6
2022	2.88	1.028	2.25	12.0

Table: Carlos Rodón's statistics before and after surgery

While there is significant improvement in every statistic after surgery, we still have not separated surgery-related effects from natural progression and league trends. How would Carlos Rodón have performed if he never needed Tommy John surgery? More broadly, how can we isolate the treatment effect of undergoing Tommy John surgery? In traditional experimental design, the counterfactual is measured through random assignment. However, Tommy John surgery cannot be randomly assigned. As a result, we employ **quasi-experimental methods** to establish a treatment effect.

Data and Methods

- **Data**: Tommy John Surgery data from @MLBPlayerAnalys on X. Box score data from baseballr.
- **Treated pitchers**: 49 pitchers who underwent Tommy John surgery since 2010, each with 3 qualifying seasons pre-surgery and post-surgery (>10 starts or >20 appearances). 2020 is excluded entirely due to COVID and opt-outs.
- **Control pitchers**: Customized for each treated pitcher, known as a **donor pool**. Consists of all pitchers who were fully healthy across the treated pitcher's pre- and post-treatment windows. Donor pools range from 50 to 100 pitchers.
- **Performance metrics**: Fielding Independent Pitching (FIP) and Strikeouts Per Nine (SO9)
- **Causal methods**:
  1. **Differences-in-Differences (DiD)**: ATE calculated as (post-pre for treated pitcher) minus (average post-pre for donor pool) for both FIP and SO9. Quantifies how surprising a treated pitcher's deviation post-surgery is relative to league-wide variance over the same window. DiD is a useful benchmark, but it compares the treated pitcher to an entirely different set of pitchers and cannot answer the counterfactual: how would the pitcher have performed had they not needed surgery?
  2. **Synthetic Control Method (SCM)**: SCM constructs a 'synthetic' healthy version of a treated pitcher by taking a weighted linear combination of donor pitchers whose pre-treatment FIP or SO9 trajectories the treated pitcher's own pre-surgery performance. Each pitcher receives a distinct synthetic control for FIP and SO9. Post-treatment, the gap between the actual pitcher's performance and their synthetic counterpart serves as the treatment effect. In theory, the synthetic control tells us how the pitcher would have performed had they not needed surgery. No additional covariates were added. Statistical significance is determined through *placebo testing* where each donor has the SCM run on them and the results are compared to the treated player.

***A Note on Scope**: This analysis does not try to establish strict causality between Tommy John surgery and changes in performance. Rather, the goal is to identify trends and provide a more rigorous empirical framework for evaluating pitchers that goes beyond simple before-and-after comparisons. Both DiD and SCM contextualize individual pitcher trajectories against the baselines but do not make definitive causal claims.*

Result 1: Carlos Rodón's post-surgery trends

The below figure shows the donor pitchers that replicate rodón's strikeouts per nine performance. The sum of each donor pitcher's SO9 multiplied by their weight should closely resemble Rodón's SO9 in each of his three seasons prior to surgery.

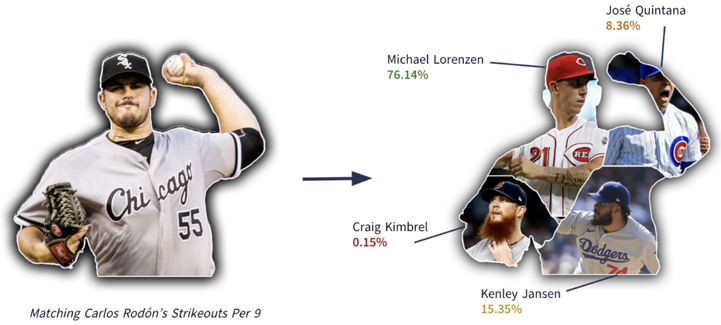

Figure: Rodón and his synthetic controls

First, we compare Rodón's variation to the donor pool's variation using DiD.

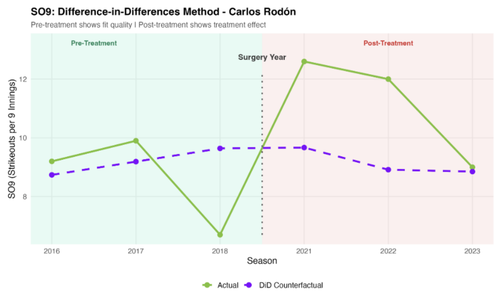

Figure: Rodón's year-by-year trajectory compared to the donor pool

The average SO9 among the donor pool (DiD counterfactual) varies minimally by year. Rodón's strikeouts per nine goes from around league average and well below league average in his pre-treatment years to well above average in his first two years post-treatment. This tells us that Rodón's treatment-induced improvement is significant in the broader context of normal year-to-year variation for a healthy pitcher.

The next step is to compare Rodón to his healthy synthetic control for both FIP and SO9.

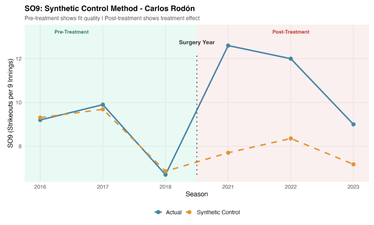

Figure: Rodón SO9 Synthetic Control

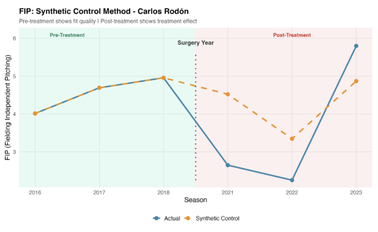

Figure: Rodón FIP Synthetic Control

Despite a down 2023 campaign, Rodón experiences a statistically significant increase in strikeouts per nine. Across all 51 of Rodón's donors, no pitcher observed a greater change in strikeouts per 9 than Rodón himself. Rodón's FIP is a bit of a different story. His post-treatment performance, while lower than his synthetic control, is not as significant partly due to a down 2023 three years post-treatment. With that being said, there is still a clear improvement compared to pre-surgery.

Result 2: Other pitchers' post-surgery trends

This section will highlight a few other pitchers who either significantly benefited or were significantly harmed by surgery.

The following figure illustrates two players, Tyler Glasnow and David Robertson, who both seemed to come back from surgery stronger than ever.

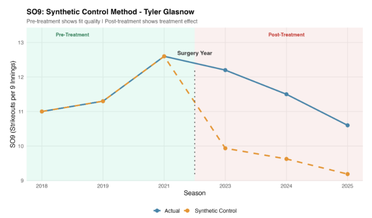

Figure: Glasnow SO9 Synthetic Control

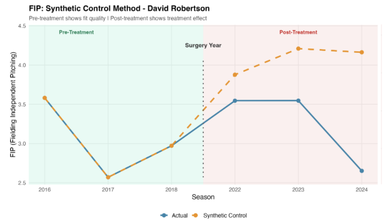

Figure: Robertson FIP Synthetic Control

A naive approach might claim that Glasnow's decrease in SO9 meant a negative treatment effect. However, his synthetic control has a much larger decrease. Robertson performed slightly better than his control in year 1 and year 2 post-treatment before having his best FIP year since 2017. His treatment effect was higher than 87 percent of the donors ($p = 0.13$).

The next two figures display Noah Syndergaard and Justin Verlander's post-surgery trends. Both players experienced negative effects following surgery.

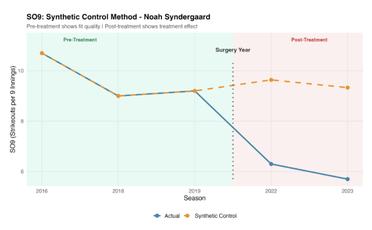

Figure: Syndergaard SO9 Synthetic Control

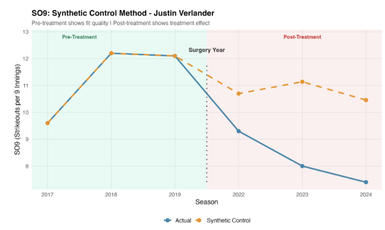

Figure: Verlander SO9 Synthetic Control

Based on the results, it is reasonable to conclude that Tommy John surgery significantly hindered Syndergaard's performance ($p = 0.064$). While Verlander's FIP has continued to stay consistent, his SO9 dipped significantly post-treatment.

Result 3: Cumulative findings

While Tommy John surgery varies by individual, the aggregated findings reveal a distinct multi-year trend. The difference for treated pitchers relative to their synthetic counterfactual is quantified below.

	Mean SO9	Median SO9	Mean FIP	Median FIP
Year 1	-0.879	-0.931	0.366	0.455
Year 2	-0.988	-1.025	0.579	0.528
Year 3	-0.149	-0.201	-0.006	0.020

Table: Three-Year Post-Surgery Statistical Differentials - Compared to synthetic control

As we can see, patience is key with Tommy John surgery recovery. Not only does it take 1-2 seasons before a player can begin pitching, but our data shows that it takes another 2-3 seasons for the pitcher to regain form and reach their healthy versions.

Limitations and Future work

- Pitcher performance is volatile and majority of statistics are dependent on outside factors.
- Survivorship Bias - While all selected players returned to MLB play, the initial dataset contained nearly 400 pitchers. Donor pool pitchers might be dealing with minor injuries.
- Future work could forecast performance post-surgery or assess the financial implications for teams.

References

Ahmad (2025) UCL injuries, *Dr. Ahmad Sports Med*; **Beckley et al. (2018)** Japan's Economic Miracle, *JEAS*; **Berra (2012)** Force of habit, *ESPN*; **Capili & Anastasi (2024)** Quasi-experimental designs, *AJN*; **Corcoran (2019)** Biomechanics of pitching, *ND*; **Gupta et al. (2014)** Performance post-TJS, *OJSM*; **Heath (2024)** Defensive shifts, *SAGB*; **Kennedy-Shaffer (2024)** Shift ban analysis, *arXiv*; **Petti & Gilani (2021)** *baseballr* package; **Ready (2025)** TJS fix, *Yale Med*; **Data:** Surgery records via @MLBPlayerAnalys (X).