A Quality Monitoring Approach to Evaluating Re-Injury Likelihood to MLB Pitchers

Overview

Introduction & Motivation

Data & Methodology

Analysis & Findings

Conclusion & Future Directions

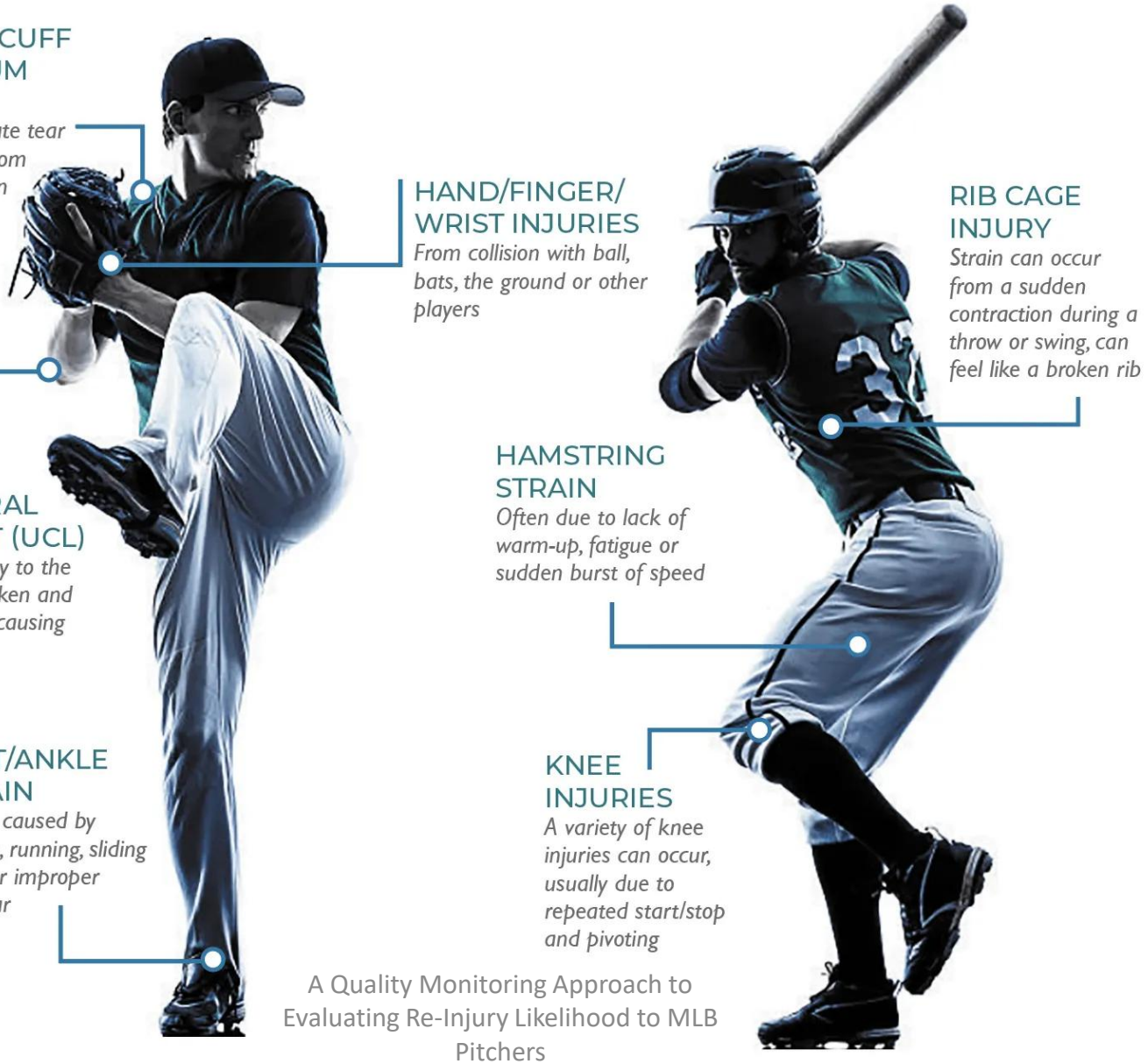
Why This Study Matters?

The issue of recurring injuries and its impact on teams and players.

Introduction to MLB Pitcher Injuries

- Major League Baseball (MLB).

TOP BASEBALL INJURIES



ROTATOR CUFF OR LABRUM TEAR

Mild to moderate tear due to stress from throwing motion or poor form

HAND/FINGER/ WRIST INJURIES

From collision with ball, bats, the ground or other players

RIB CAGE INJURY

Strain can occur from a sudden contraction during a throw or swing, can feel like a broken rib

ULNAR COLLATERAL LIGAMENT (UCL)

Provides stability to the elbow, can weaken and tear over time causing pain

HAMSTRING STRAIN

Often due to lack of warm-up, fatigue or sudden burst of speed

FOOT/ANKLE SPRAIN

Can be caused by overuse, running, sliding bases or improper footwear

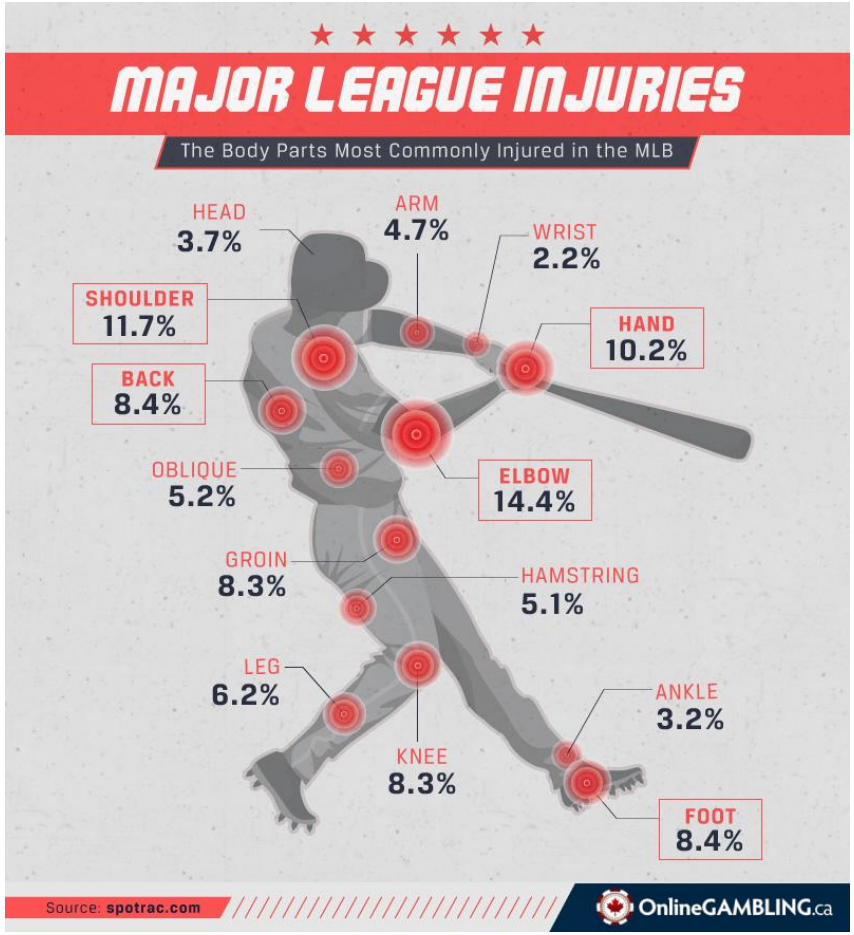
KNEE INJURIES

A variety of knee injuries can occur, usually due to repeated start/stop and pivoting

MLB Injury Overview

- Types of injuries considered in the study.

- Upper extremity injuries, particularly elbow injuries, to MLB pitchers
- Injuries to the ulnar collateral ligament (UCL) that require Tommy John surgery
- Re-injuries to the same arm after a pitcher has undergone Tommy John surgery



Background Work

Specifically considering injuries to MLB (baseball pitchers in general), most studies focus on;

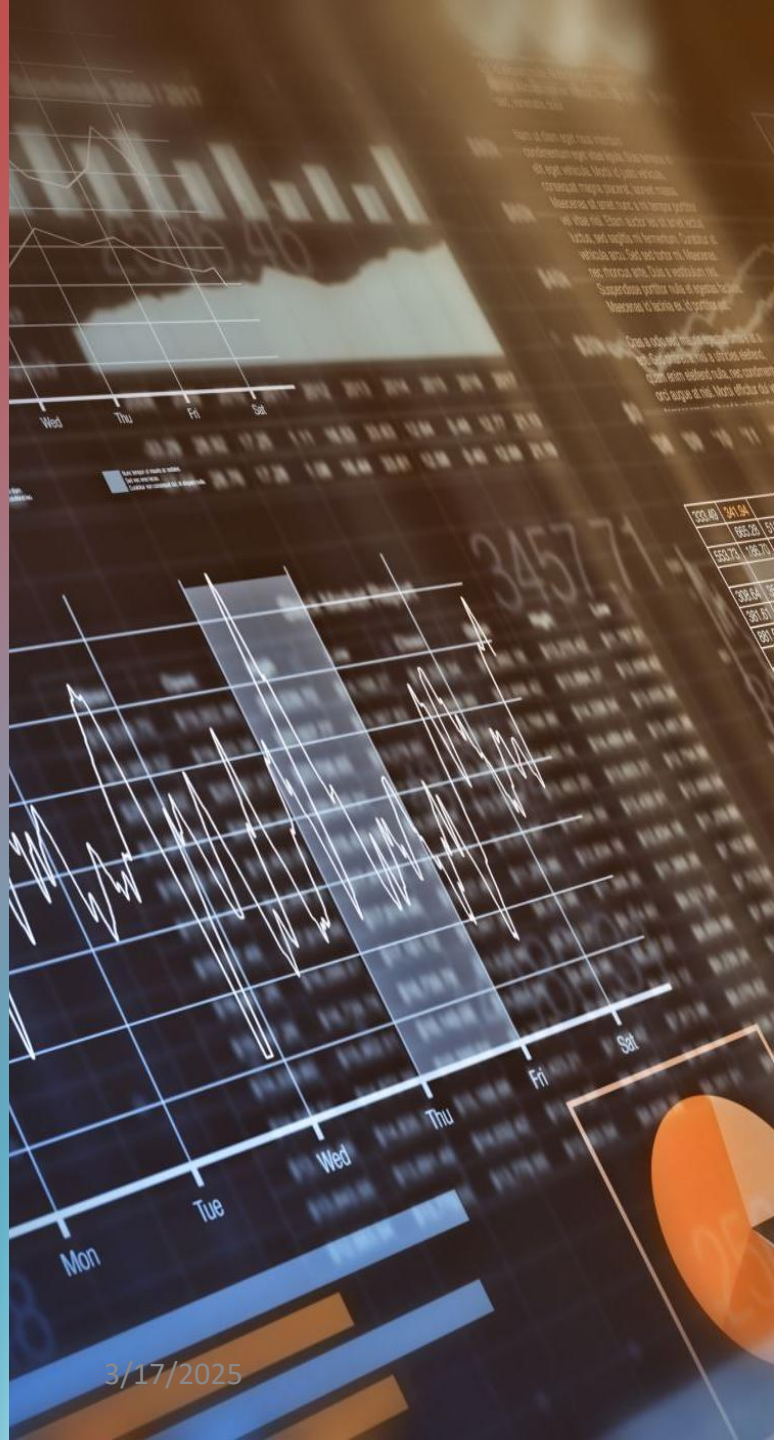
1. Frequency of injury,
2. Performance post-injury,
3. Injury detection, and
4. Strategies for injury prevention.

Previous Work

- Pitchers sustain more frequent and more severe injuries compared to other baseball players, with longer recovery times (Posner et al., 2011).
- Recovery time, measured as the average days on the IL, is also longer for pitchers than for fielders (Posner et al., 2011) & (Ciccotti et al., 2017).
- Pitchers who undergo Tommy John surgery have a high likelihood of requiring a second stint on the injured list, and often show declines in performance metrics after returning. (Ciccotti et al., 2017).

Previous Work

- In a longitudinal study from 1999 to 2011, found that 57% of pitchers who underwent Tommy John surgery needed a second IL stint for injuries to the same arm. (Makhini et al., 2014)
- Pitchers who require a second surgery to repair a reinjured UCL have very low rates of returning to their prior workload (Makhini et al., 2014) & (Brown and Do, 2022)



Current Challenges

Monitoring methods rely on *retrospective analysis* rather than *real-time tracking*.

This suggests that traditional approaches:

- **Analyze past injuries** instead of detecting risks in real-time.
- **May not be proactive** in preventing injuries.
- **Could delay intervention**, making it difficult to prevent re-injury before it occurs.

Study Objectives

- Developing a control charting method for detecting re-injury.

- The study proposes a control charting technique to **monitor pitcher performance in real-time** as an indicator of potential re-injury.

- The technique can classify pitchers into **re-injured and non-re-injured** groups with reasonable accuracy using sensitivity and specificity metrics.

- The technique can also provide an estimate of how many games it takes before a pitcher's performance is deemed "out-of-control" and potentially **indicative of lingering injury**.

Key Research Questions

How well does this method predict re-injury risk?

Expected Impact

- Benefits for teams, trainers, and players.

- The proposed quality monitoring approach could help MLB teams with early detection of pitcher injuries, which could prevent further injury and improve team performance.

- The approach aims to identify when a pitcher's performance is indicative of a lingering injury, which could help trainers and players by allowing for earlier intervention and rehabilitation.

- The study discusses the practical implications and limitations of the proposed quality monitoring approach, which could be useful for teams, trainers, and players in understanding how to best implement and utilize the method.

Data Sources

MLB injury data for the 2021-2023 seasons were obtained from FanGraphs' Injury Report (FanGraphs, n.d.)

Where and how data was collected (MLB records, Fangraphs, Lahman database).

Lahman database in R, all analyses were performed using the R statistical software version 4.3.3 (R Core Team, 2024).

Preprocessing the Data

- Cleaning steps, filtering for pitchers only.
  - *The original = 3,057 observations were filtered to include only those pitchers who;*

    (1) played for one team the whole season,

    (2) had at most two stints on the IL,

    (3) returned from injury prior to August of the given year, and

    (4) had at least five game appearances post-injury.

- For the 2021 - 2023 regular seasons, 219, 192, and 147 pitchers, respectively, met these criteria, for a total analyzable **sample size of = 558.**

Key Variables

- **<u>Innings Pitched (IP)</u>** - the total number of innings a pitcher has thrown during a game or season. *One inning consists of three outs.*

» $\mathrm{IP} = \frac{\mathrm{IPouts}}{3}$

  - Note ; In the Lahman dataset, the number of outs made by the pitcher is recorded in **IPouts**

Key Variables

- WHIP (**Walks plus Hits per Inning Pitched**): a baseball statistic that measures how many base runners a pitcher allows per inning.

» $\text{WHIP} = \frac{H + BB}{IP}$

where,
- **H**: Hits allowed
- **BB**: Walks allowed
- **IP**: Innings pitched

- *Participant;* Only pitchers with *15+ innings pitched* are considered to ensure *a reliable sample size*.

  - **Significance**: A ***low WHIP*** indicates a pitcher who ***is effective*** at limiting base runners, while a **high WHIP** suggests the opposite.

Why WHIP?

- Justification for using WHIP as the performance metric.

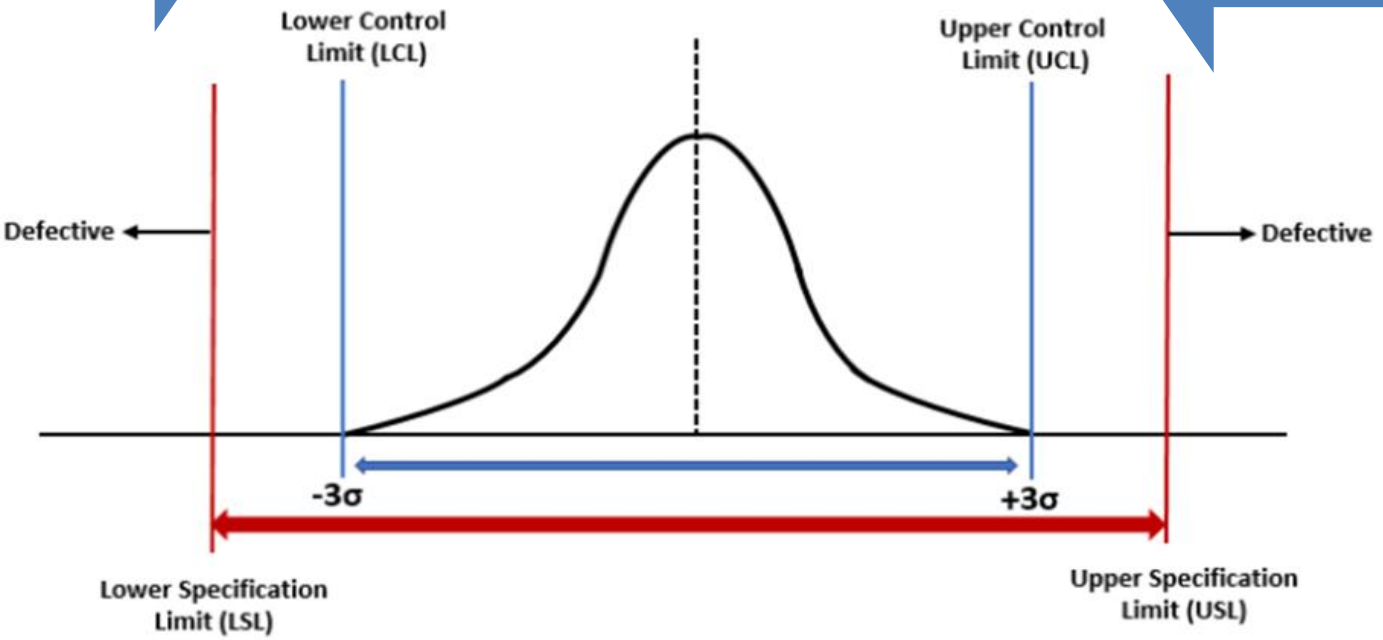
➢ 1) Increases in a pitcher's WHIP can indicate the pitcher is not meeting the expected standard for MLB pitchers, making it a relevant metric to **monitor for signs of declining performance** that could signal a re-injury.

➢ 2) The league-wide **mean and standard deviation of WHIP can be calculated on a season-by-season basis**, providing a relevant benchmark to compare individual pitcher WHIP values against, allowing the proposed monitoring approach to be implemented in a practical, real-world setting.

Process Capability Analysis

What is Process Capability Analysis?

- Refers to a process' ability (or inability) to regularly adhere to a given standard



Process Capability Index

- The most common PCI is called the "Cp" index
  - $C_p = \frac{\text{USL} - \text{LSL}}{6\text{s}}$

  where, USL and LSL; ***upper and lower specification limits***

- If ***$Cp \geq 1$***, this indicates that at least 99.73% of the observed values from the process are expected to fall within USL and LSL, which implies that the process is ***capable***.

- **If *$Cp < 1$***, this indicates that a potentially *non-negligible proportion* of observed values will *fail to meet the standard*, which implies that the process may be ***incapable***.

Process Capability Index

- For Instance, if a pitcher's WHIP gets close to 0;
  - the pitcher is almost certainly performing at a high level
- And since only upward, positive increases in WHIP are cause for concern
  - so, only a one-sided PCI is needed in such a case

$$C_{pu} = \frac{\text{USL} - \bar{x}}{3s}$$

Note; The domain of C_{pu}, that indicates capability, or the lack thereof is identical to C_p.

CUSUM Control Charts

- The CUSUM chart focuses on the cumulative sum of deviations from the target value;

  - it considers deviances ($\overline{x}_t$ - $\overline{\overline{x}}$), rather than $\overline{x}_t$ itself

  - then, the cumulative sum of the deviances (adjusted for positive and negative deviances) is plotted and compared to a decision interval (h).

Note; why not the Shewhart $\overline{X}$ -Chart; because CUSUM is efficient at detecting shifts of smaller magnitude (< 2σ).

CUSUM Control Charts

- For positive deviances, which would monitor upward shifts in the process mean, the CUSUM statistic is defined as:

$$C_t^+ = \max\left(0, C_{t-1}^+ + \left(\overline{x}_t - \overline{\overline{x}} - \mathrm{k}\right)\right)$$

where, $C_0^+ = 0$

and

k denotes a positive "*slack value*" employed to improve sensitivity to small shifts.

Note; If $C_t^+ > h$, this indicates the process mean may have shifted in an upward direction away from $\overline{\overline{x}}$.

Proposed Capability Monitoring Technique

- In the context of MLB pitchers, Calculating the league-wide WHIP on a season-to-season basis ($i = 2021, 2022, 2023$) for pitchers with at least 15 innings pitched;

  - ✓ we obtain, the mean ($\bar{\bar{x}}_i$) and,
  - ✓ standard deviation ($\hat{\sigma}_i$) of the WHIP-process

- we carefully defined USL for WHIP,

$$USL_i = \bar{\bar{x}}_i + 3\hat{\sigma}_i.$$

Note; this set the statistical thresholds for abnormal performance.

Proposed Capability Monitoring Technique

- These individual WHIP values were compared to their respective USL_i values in a manner similar to C_{pu}:

$$W_{ti} = \frac{(USL_i - WHIP_{ti})}{\hat{\sigma}_i}$$

- ✓ If $\boldsymbol{W_{ti} > 0}$, this indicates the pitcher is likely performing at or above their expected level of play for that specific game (i.e., a ***capable*** performance)

- ✓ If $\boldsymbol{W_{ti} \leq 0}$**,** this would indicate the opposite (i.e., an ***incapable*** performance).

Note; W_{ti} can be refer to as the standardized difference between USL_i and $WHIP_t$

Proposed Capability Monitoring Technique

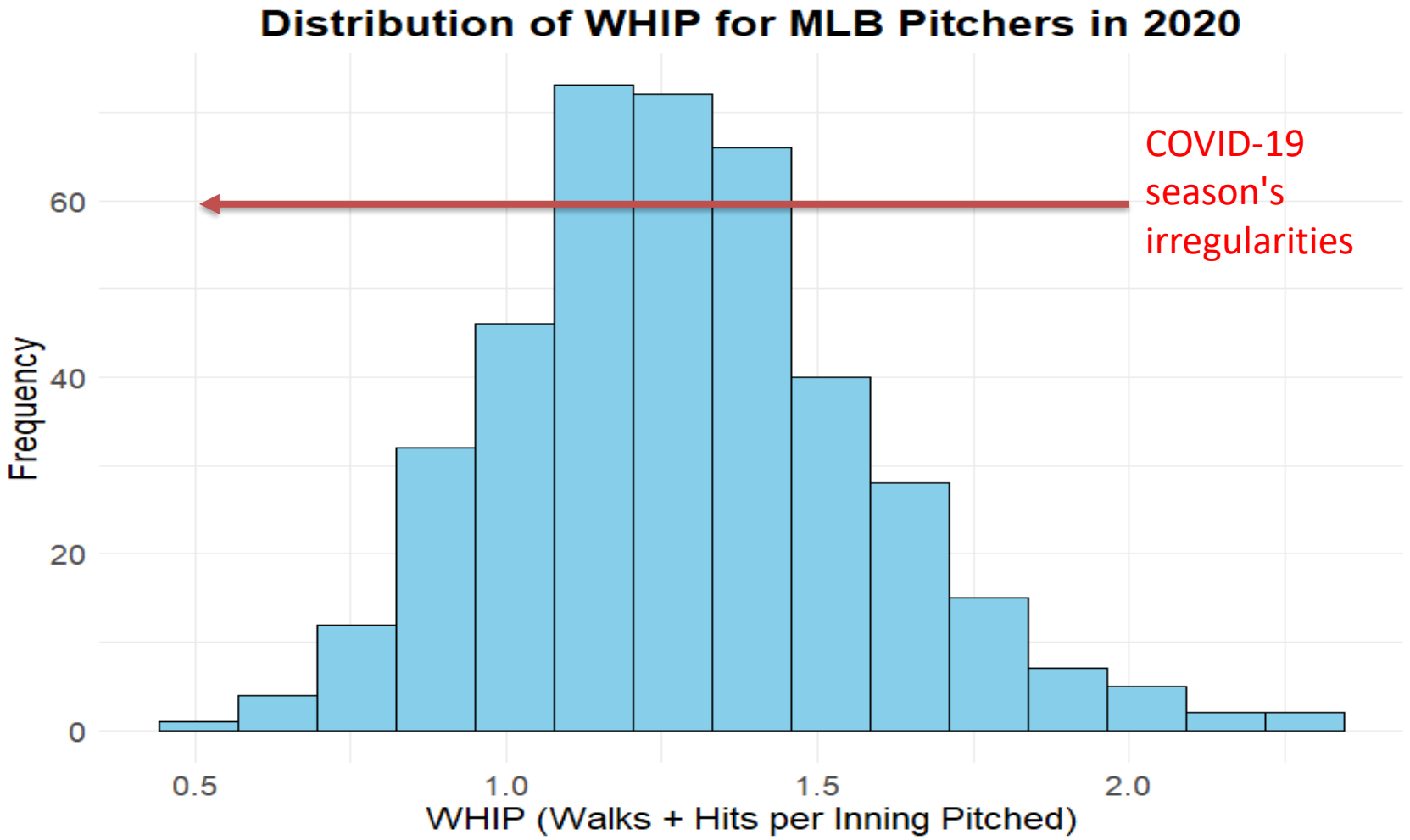
- Noting the expected value of W_t is 3, the CUSUM statistic becomes:

$$C_t^+ = \max(0, C_{t-1}^+ + (W_t - 3 - k))$$

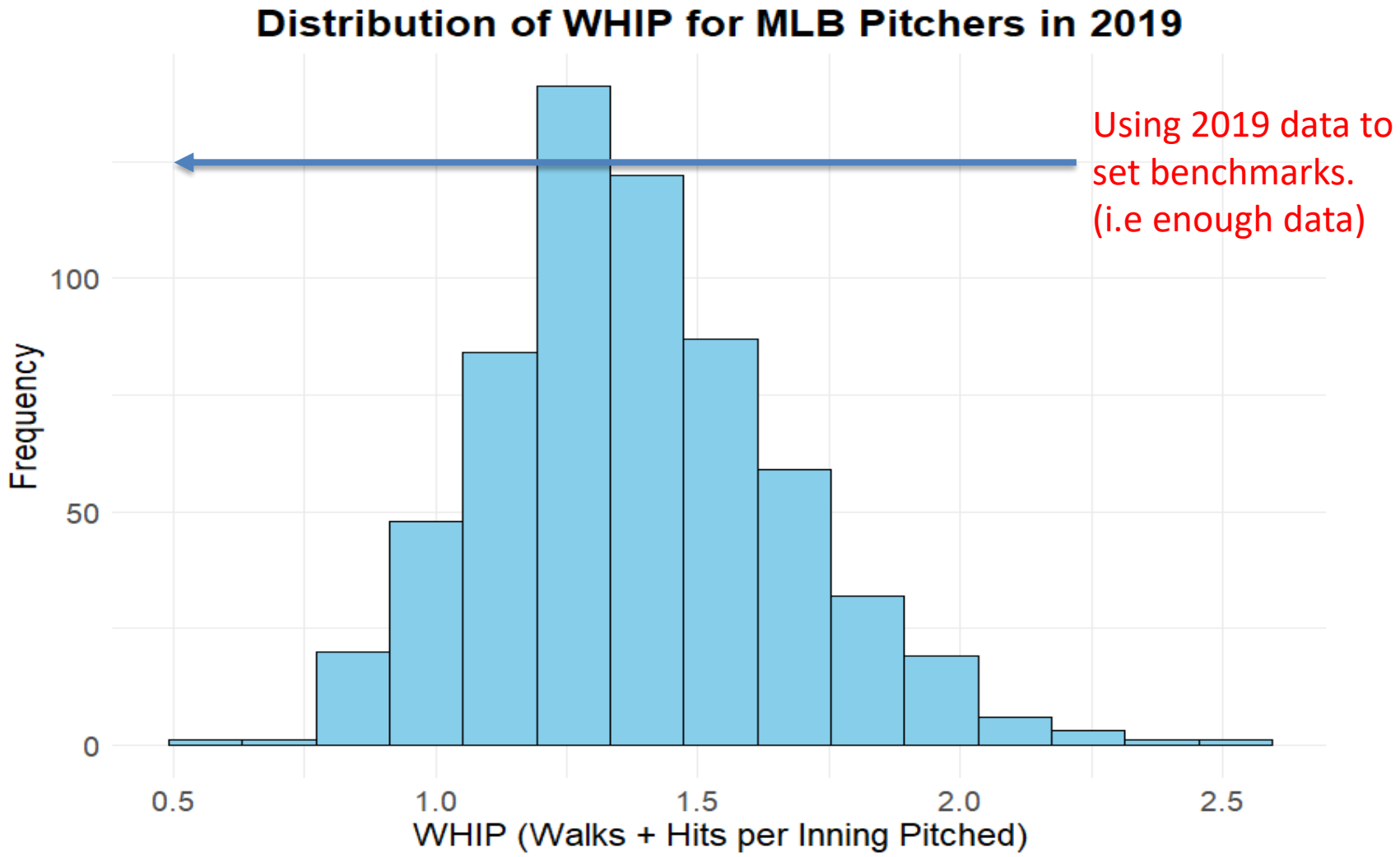
  - k is typically chosen to be the absolute value of the midpoint between the expected value of the monitored variable (3 in this case) and an out-of-control value (USL_i in this case)

- As before, if $\boldsymbol{C_t^+ > h}$ (h usually being defined at 4σ or 5σ), this indicates the pitcher has been performing substantially **below standard** due to inflated WHIP values.

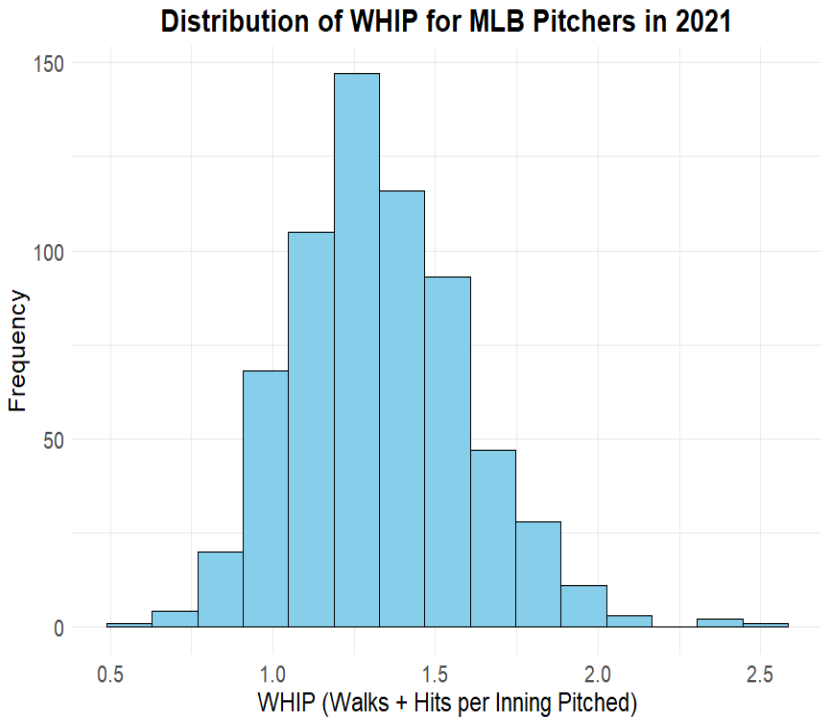
Estimating Performance Baselines

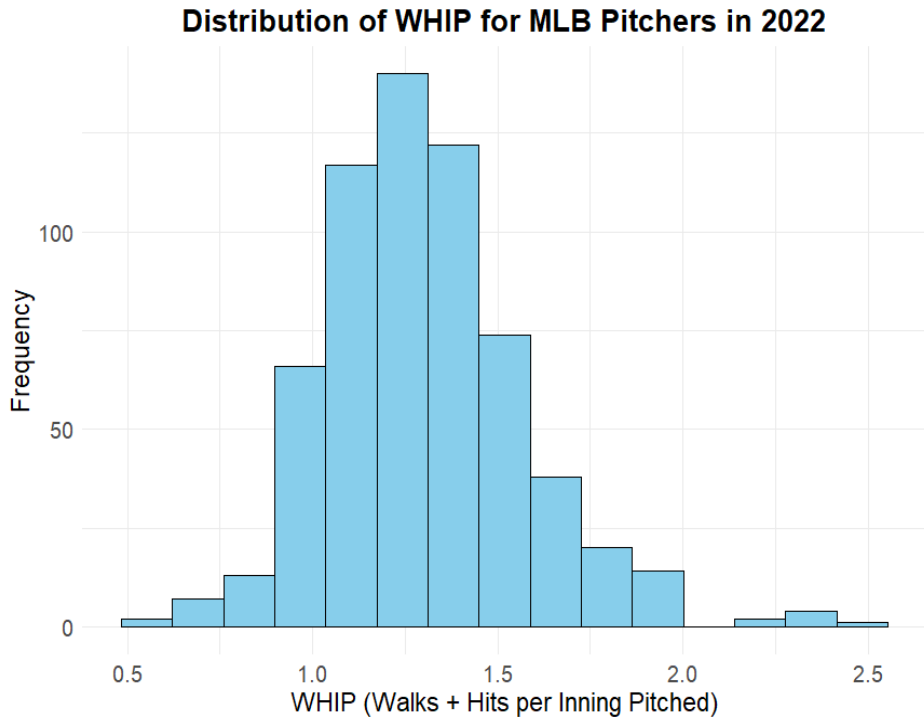


Estimating Performance Baselines



Why Not Use 2020 Data?





Key Metrics for Evaluation

How often the model misidentify injuries?

- Sensitivity (True Positive Rate) = $\frac{TP}{TP+FN}$

  where non-injured players are incorrectly identified as having a re-injury.

- Specificity (False Positive Rate) = $\frac{TN}{TN+FP}$

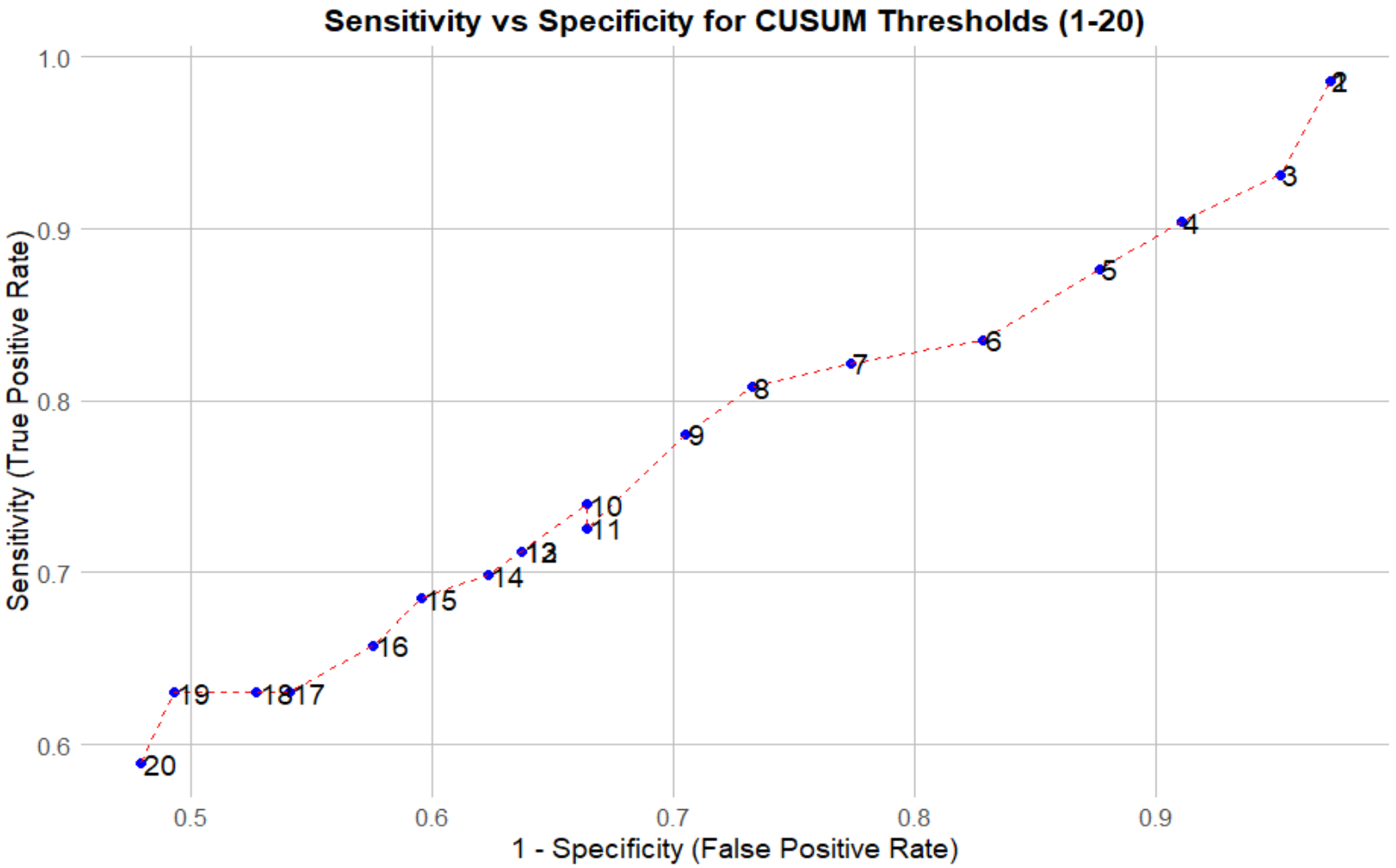
  where these thresholds (h) catch most re-injuries but also include more false alarms

Estimated Sensitivity and Specificity Values for Various Decision Interval Values

Season	2021		2022		2023	
Mean (SD) WHIP	1.28 (0.29)		1.33 (0.27)		1.30 (0.27)	
h	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
1	98.63%	2.74%	94.74%	0.86%	100.00%	2.63%
2	98.63%	2.74%	94.74%	3.45%	100.00%	4.39%
3	93.15%	6.16%	93.42%	5.17%	100.00%	10.53%
4	90.41%	10.27%	86.84%	9.48%	96.97%	15.79%
5	87.67%	12.33%	84.21%	15.52%	93.94%	21.93%
6	82.19%	19.86%	75.00%	21.55%	87.88%	23.68%
7	80.82%	23.29%	72.37%	24.14%	84.85%	27.19%
8	79.45%	26.71%	68.42%	28.45%	84.85%	28.95%
9	**73.97%**	**33.56%**	**68.42%**	**31.90%**	**84.85%**	**32.46%**
10	**72.60%**	**34.25%**	**59.21%**	**32.76%**	**84.85%**	**35.09%**
11	**72.60%**	**35.62%**	**56.58%**	**34.48%**	**81.82%**	**38.60%**
12	71.23%	36.99%	52.63%	36.21%	78.79%	39.47%
13	69.86%	37.67%	48.68%	37.93%	75.76%	41.23%
14	68.49%	38.36%	47.37%	42.24%	75.76%	42.11%
15	65.75%	41.78%	46.05%	44.83%	66.67%	46.49%
16	63.01%	45.21%	42.11%	47.41%	63.64%	47.37%
17	63.01%	46.58%	40.79%	50.00%	63.64%	48.25%
18	63.01%	49.32%	36.84%	51.72%	63.64%	50.00%
19	58.90%	52.05%	36.84%	51.72%	57.58%	50.00%
20	57.53%	53.42%	35.53%	54.31%	51.52%	52.63%

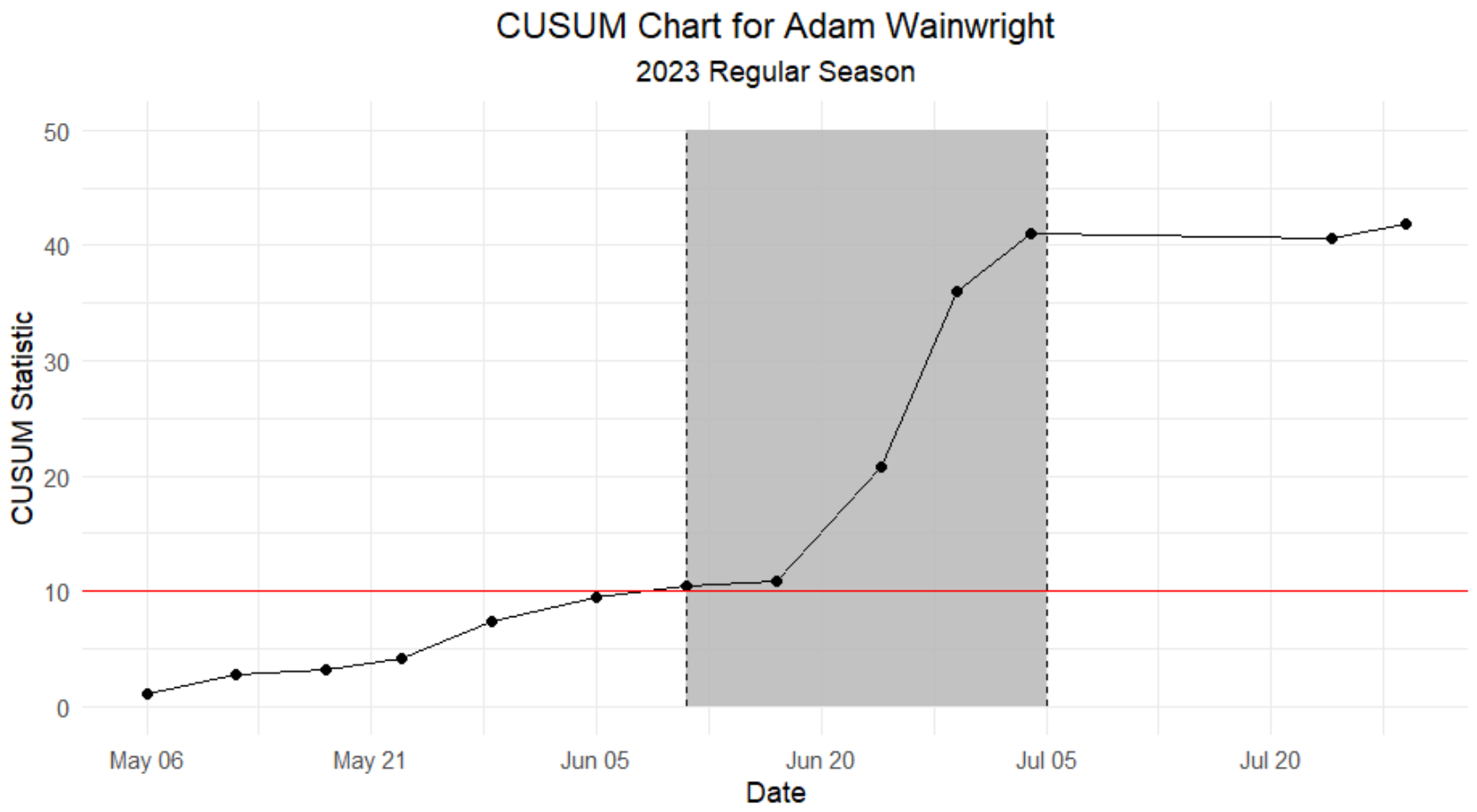
CUSUM Thresholds for Detecting Injuries

How different h-values affect detection rates.



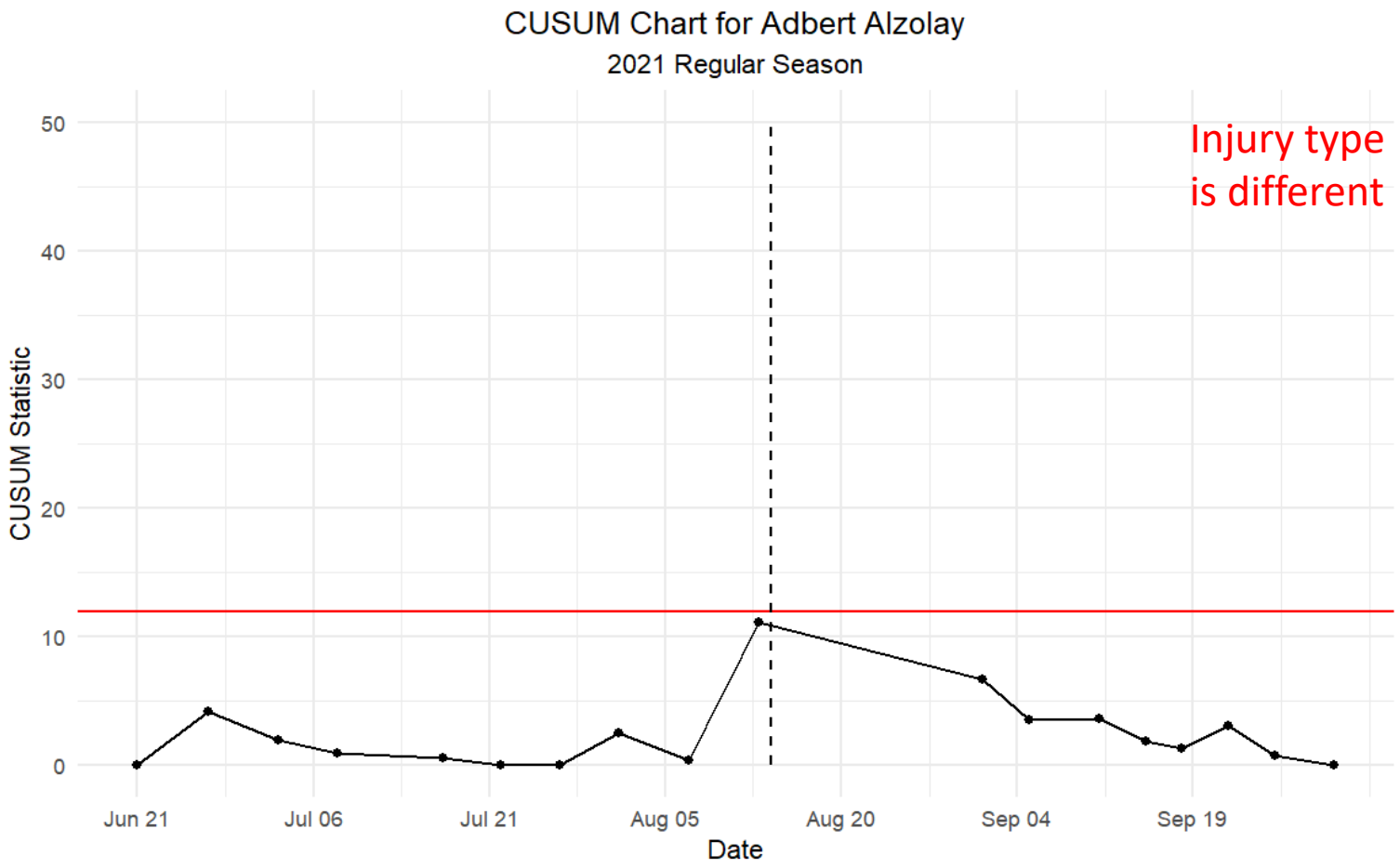
Case Study

A pitcher flagged by CUSUM before re-injury.



Case Study 2

A pitcher Never flagged by CUSUM After re-injury.



How Reliable is this Approach?

- There is **a tradeoff between sensitivity and specificity**, where increasing the decision interval (h) to be more specific also decreases the sensitivity.

- The ***number of game appearances before an out-of-control signal*** is observed is an important metric for evaluating the reliability of the approach.

- The results related to the number of game appearances before an out-of-control signal is observed are provided in ***Next Table***.

Mean Game Appearances Before Reinjury Classification by Decision Interval Value (h)

Season	2021	2022	2023
Mean Game Appearances Before Second Stint	12.12	12.26	9.09
h			
1	2.72	1.94	1.64
2	3.96	2.26	1.91
3	4.38	3.69	3.03
4	5.00	3.64	3.25
5	5.35	4.36	3.35
6	6.32	5.44	4.24
7	6.53	6.09	4.61
8	6.74	6.29	4.71
9	7.85	6.60	4.96
10	8.02	7.20	6.14
11	9.13	7.35	6.70
12	9.33	7.90	7.12
13	9.49	8.11	7.16
14	9.35	8.31	7.20
15	9.15	8.40	6.64
16	9.67	8.88	7.29
17	10.48	9.45	7.76
18	10.36	10.93	9.29
19	10.29	11.00	10.68
20	11.10	10.81	8.76

Areas for Refinement & Future Work

- The analysis did not consider factors like;
  - ✓ *injured pitcher quality*,
  - ✓ *replacement pitcher quality*,
  - ✓ and *injury severity*,

  which may be more important than just the quantity of injuries.

- The study did not incorporate other player-specific factors like;
  - ✓ *workload*
  - ✓ and injury type,

  which could help better predict and prevent re-injury.

Recommendation

Sensitivity & Specificity Analysis: Sensitivity was high for most h values, but specificity was lower.

Decision Interval (h): Values between **7 – 12** are recommended for practical use.

Why: The cost of misclassifying a struggling pitcher as injured is low (e.g., imaging by a doctor between games).

Effective Monitoring: using the h range of 7–12 provides a balanced approach for injury prediction.

Ethical Considerations

Avoiding false injury accusations

The End

Thanks For Listening

Your Final Thoughts

Q&A