BEYOND THE GLOVE: QUANTIFYING THE HIDDEN VALUE OF CATCHERS IN MAJOR LEAGUE BASEBALL

By

Lance Brady (Yale University, B.S. Statistics and Data Science)
Mason Dugas (Yale University, B.S. Statistics and Data Science)

Submitted to Professor Tobias Moskowitz and Professor Nils Rudi
for MGT 817: Sports Analytics
at Yale School of Management

Fall 2024

Introduction

Abstract

This study quantifies the hidden value of Major League Baseball (MLB) catchers beyond traditional defensive metrics such as framing, blocking, and throwing. Utilizing pitch-by-pitch Statcast data and advanced statistical methods, including logistic regression, mixed-effects models, and Bayesian hierarchical models, we evaluate latent traits like game-calling and pitcher-catcher synergy. Our findings reveal year-over-year consistency in catchers' latent contributions to strikeouts and their influence on metrics like xFIP-, highlighting traits not captured by standard statistics. Additionally, team context and organizational strategies play critical roles in shaping catcher performance, with results suggesting that advanced analytics and game-planning can enhance or diminish a catcher's value. By bridging gaps in existing research and quantifying unmeasured attributes, this work provides a comprehensive framework for evaluating catchers' contributions, emphasizing the interplay between individual skills and team dynamics. These insights have significant implications for player evaluation, compensation, and roster construction in modern baseball.

Table of Contents

Review of Literature

I. Introduction to Baseball Data and the Catcher Position

Major League Baseball, long rooted in quantitative analysis, has undergone a major revolution in the last two decades, especially with the advent of pitch location data called PITCHf/x in 2006–which was replaced by the more accurate TrackMan system in 2017–and subsequent Statcast data implementations in 2015, which used doppler radar and high definition video to measure the speed, acceleration, and other aspects for every player on the field.

One position greatly affected by the implementation of new quantitative metrics is the catcher. Catchers have a large role in defensive play, which makes evaluating their contributions quantitatively significant. It is a physically and mentally taxing position, whose jobs include but are not limited to: relaying messages from coaches to the rest of the infield, constantly backing up first or third base, calling pitches in an optimal order, making plays on defense, and being a part of every single pitch in a nine inning baseball game (Garro, 2019). Unlike pitchers, they also have to hit.

II. Quantifying a Catcher's Performance with CERA

Initial Research

Early attempts to evaluate a catcher's performance studied forms of a general metric called Catcher's ERA (CERA). One early approach looked at 104 catchers from 1946 to 1987, and found that as a catcher moved from rookie to veteran with the same team, the ERA of the pitching staffs he handled improved significantly but eventually declined after an ideal period of around 400-800 games. Even with just 104 catchers in the study, it was enough to rule out randomness in obtaining the results, and an average drop of about a third of a run per game from the catcher's rookie season to his prime years was noted (Hanrahan, 1999). Hanrahan identified three main roles of the catcher–calling the game, handling the pitchers, and framing the strike zone–and found that the differences in catchers' abilities to throw out potential base stealers was much less important than his other abilities. He recognized, however, that teams employing a rookie catcher might be in a "rebuilding" year more often than is typical, and thus employed a subpar pitching staff concurrently. His subsequent study found that when looking at pitcher-catcher pairs, pitchers' ERAs would increase when throwing to a new catcher, but soon decreased as more time was accumulated between them (Hanrahan, 2009).

A different approach by Sean Smith in 2015 found conflicting results, and thus, no "rookie catcher effect," as Hanrahan's findings had been coined, even though catchers seemed to get better with experience. This research found that although catching technique and game-calling are two skills that affect the game, they were not metrics that research could put a number on, but instead were small parts of the game, whose impact on winning is apparent by common sense. Smith also observed that while there were analytically good catchers in MLB, those catchers might not always be the best to work with a specific pitcher. This logic is perhaps most obvious in the case of the rare knuckleball pitcher, where a catcher's ability to catch the pitches he has caught throughout his entire career–fastballs, breaking balls, changeups–may not translate to catching for a knuckleball pitcher. Smith's research was important because he pointed out that managers often valued defense *too* much, and that having a catcher who was offensively valuable in the lineup was often more important.

Other early research pointed against catcher's influence entirely, saying that there was no substantial evidence–when looking at pitcher-catcher combinations as Hanrahan did originally or on two catchers working with the same pitcher across years–that catchers had any influence on offensive production of the opponents (Woolner, 2000). Woolner later used a simulation technique to identify a pitcher's above-average and below-average catchers, finding that overall, above-average catchers stay somewhat above-average and below-average catchers stay below-average (Woolner, 2002). In this study, Woolner acknowledged that with significant noise in yearly catcher measurements, the hypothesis that catchers impact pitcher performance could not be proven, but it could also not be refuted. This corroborated early research by Bill James, father of modern baseball analytics, who found that even if catchers do have a significant defensive ability, CERA could not be a reliable indicator of it due to yearly variation. Woolner's research also found an upper bound on the effect's magnitude, and he concluded that most catchers don't deviate from average by a full run of CERA or more, if it is due to intrinsic traits.

Flaws with CERA

CERA has inherent flaws that highlight the need for researchers to refine their approach to evaluating catchers' contributions to run prevention. Foremost of those is that it is a total contribution statistic, which means that it has all of a catcher's duties baked into it, along with other noise like pitcher and defensive strength. That means that using CERA doesn't give us any indication of the importance of each of a catcher's roles–there is nothing to separate framing from blocking and throwing. Secondly, early research methods were often as exploratory as the topics themselves; differing approaches—such as analyzing pitcher-catcher pairs versus entire pitching staffs, or focusing on plate appearance (PA)-level versus inning-level run prevention—resulted in findings that were difficult to compare.

To deal with this, Woolner also made a model using the "three true outcomes" of baseball: strikeouts, walks, and home runs. These are outcomes which take away the defense's role in run prevention, making them better for studying pitcher-catcher pairs. This method involved finding a rate value of a derived variable called Pitching Runs divided by Plate Appearances (PAs) in subsets of PAs to find a Run Prevention Rate representing how many fewer runs the pitcher yielded per batted faced with the catcher in question. He found a distribution resembling the random chance normal distribution, either pointing to the no-catcher-effect hypothesis or just that catchers are normally distributed in skill. He found that with year-to-year catcher-pitcher pairs, how well they worked together in RPR is not correlated with how they'll work the following year relative to other catchers on the club, with r = 0.02. He acknowledged that if there were a true game-calling or latent synergy skill, it was below the threshold of detection, although some catchers might have specific skills in clutch situations, or keeping the pitch count low by determining what his pitcher is throwing well, helping their pitchers go deeper in games (Woolner, 1999).

Later All-Encompassing Metrics using CERA

While early studies used CERA to evaluate a catcher's overall contributions, later metrics refined its application by focusing on specific aspects of a catcher's performance. Retrosheet's Catcher Defensive Rating System, for instance, utilized CERA solely to measure game-calling through the OCERA/CERA ratio, defined as "All Other Catchers on His Team Earned Run Average divided by The Catcher's Earned Run Average" (Rosciam, 2009). This rating system identified five other important attributes: stamina, good glove, good arm, ball handling, and effective game play, and it was a step toward minimizing CERA's flaws by using other standard statistics to try to quantify each of a catcher's roles. Tom Tango's With or Without You (WOWY) model compared a catcher to other catchers on their team, like Hanrahan had previously, but it did not capture a catcher's performance relative to other players around the league. Fangraphs'

Defensive Runs Saved (DRS) metric also attempted to describe a catcher's overall play using the following: normal fielding, bunt runs saved, blocking play, stolen base runs saved, adjusted earned runs saved (essentially game-calling), and strike zone runs saved (framing) (DeForest, 2023). In this way, CERA was used as simply part of the game-calling aspect of the DRS calculation.

III. Baseball Savant Metrics

Since 2015, Statcast has been tracking the "advanced metrics" of baseball, with location data that allows for research on nearly every action in a baseball game. The three main areas in which metrics germinated for the catcher position were catcher framing, blocking, and throwing.

Framing

Catcher framing is defined by MLB as: "the art of a catcher receiving a pitch in a way that makes it more likely for an umpire to call it a strike -- whether that's turning a borderline ball into a strike, or not losing a strike to a ball due to poor framing" (MLB, 2024). Statcast gives quantitative metrics mainly based on the edges of the plate, defined as the area that is the width of two baseballs (one inside the zone, one outside) around the entire edge of the strike zone, and adjust their metrics both for ballpark and pitcher. Below is the leaderboard of the top 10 catchers according to Statcast's Catcher Framing Runs according to a .125 run/strike conversion. In terms of catcher framing, about 86% of the contribution comes from the edge of the zone, 6% from the middle and 8% from far off the plate.

Rk.	Catcher	Team	Pitches	Catcher Framing Runs	Strike Rate	Zone 11	Zone 12	Zone 13	Zone 14	Zone 16	Zone 17	Zone 18	Zone 19
1	Bailey, Patrick		3214	16	52.5%	23.3%	57.4%	28.4%	64.6%	70.5%	29.8%	54.8%	32.2%
2	Raleigh, Cal		3404	13	49.1%	22.4%	55%	30%	60.9%	68.5%	30%	46.3%	28.4%
3	Wells, Austin		3111	12	48.6%	23%	49.3%	23.9%	58.1%	66%	31.5%	55.1%	29.3%
4	Kirk, Alejandro		2586	10	49.8%	21.7%	50.7%	21.9%	67.3%	61.2%	34.6%	56.1%	28.8%
5	Trevino, Jose		1890	10	50.7%	22.3%	47.3%	19%	61%	74.4%	33.6%	52.4%	35.7%
6	Rogers, Jake		2525	9	51%	19.6%	52.6%	26.7%	61.7%	64.3%	35.5%	59%	30.1%
7	Naylor, Bo		3105	8	49.3%	23.9%	48.4%	19.6%	66.9%	68.1%	34.8%	50.2%	22%
8	Grandal, Yasmani		1843	6	48.8%	12.3%	48.8%	29%	47.5%	70.6%	34.4%	52.1%	37.6%
9	Vázquez, Christian		2306	6	49.1%	17.4%	53.5%	29.6%	63.8%	72.8%	27.9%	44.7%	27.7%
10	Alvarez, Francisco		2754	6	46.7%	15.2%	48%	22.1%	52.3%	66.2%	35.7%	53.9%	28.4%

Figure 1. Top 10 Catchers according to 2024 Statcast Framing Leaderboard (Statcast, 2024)

Blocking

The next task of the catcher is to prevent wild pitches (WP) or passed balls (PB). These happen when the pitcher throws it to a spot around the plate that is difficult for the catcher to catch or keep in front of them. Traditional metrics simply used the quantity of wild pitches or passed balls that a catcher gave up or some rate statistic derived from one or both, with the determination of whether a pitch is a wild pitch or passed ball determined by an official scorer. Statcast assigns every pitch a probability of being a passed ball or wild pitch based upon several inputs, such as pitch location, pitch speed, pitch movement, catcher location, and batter/pitcher handedness. Based on that knowledge, each pitch a catcher receives (or fails to) is assigned a value—credited for successfully blocking a difficult pitch or debited for failing to do so—reflecting the pitch's difficulty. For example, if a catcher blocks a pitch that is a PB + WP 10% of the time, he will receive +0.10. If he blocks a pitch that is a PB + WP 90% of the time, he will receive +0.90. Blocks Above Average is thus the difference between actual PB + WP and estimated PB + WP based on opportunities seen. Blocks Above Average / game is a rate stat based on an average catcher receiving 40 blocking chances per game. Catcher Blocking Runs converts blocks to runs saved on a .25 runs/block basis.

									Opportunities			Blocks Above Average			
Rk.	Player	Team	Block Opportunities	Catcher Blocking Runs	Blocks Above Average	Actual PB + WP	Estimated PB + WP	Blocks Above Average / Game	Easy %	Medium %	Tough %	Easy	Medium	Tough	
1	Jansen, Danny		2,977	4	14	14	28	0.19	95.2%	3.5%	1.3%	4	3	7	Go To Visual
2	Amaya, Miguel		3,894	3	12	25	37	0.12	94.5%	4.3%	1.2%	0	9	3	Load Visual
3	Murphy, Sean		2,432	3	11	12	23	0.18	94.9%	4.0%	1.1%	1	5	5	Load Visual
4	Stallings, Jacob		3,026	3	10	17	27	0.13	95.1%	4.1%	0.8%	4	4	2	Load Visual
5	Rutschman, Adley		3,833	3	10	23	33	0.10	95.7%	3.2%	1.1%	3	4	4	Load Visual
6	Fermin, Freddy		2,940	2	9	14	23	0.13	95.7%	3.5%	0.8%	4	4	1	Load Visual
7	Rogers, Jake		2,857	2	8	20	28	0.11	94.4%	4.7%	0.9%	6	-1	3	Load Visual
8	Fortes, Nick		4,223	2	8	29	37	0.08	95.5%	3.5%	1.0%	4	2	2	Load Visual
9	Trevino, Jose		2,862	2	7	19	26	0.10	94.8%	4.1%	1.1%	-2	4	4	Load Visual
10	Sánchez, Ali		1,126	2	6	5	11	0.21	94.6%	4.1%	1.3%	3	1	2	Load Visual

Figure 2. Top 10 Catchers According to Statcast Blocking Leaderboard (Statcast, 2024)

Throwing

Another critical aspect of catcher defense that Statcast quantifies is their ability to prevent runs on the basepaths. Traditionally, this skill was evaluated using caught stealing percentage—the proportion of attempted base stealers a catcher successfully throws out. Early analyses focused primarily on arm strength, but over time, coaches began estimating pop time with stopwatches. Statcast revolutionized this evaluation by introducing an accurate and standardized Pop Time metric, which measures the time from when a pitch hits the catcher's mitt to when the intended fielder, typically the second baseman or shortstop, is projected to receive the throw at the base. This calculation accounts for the precise flight path of the ball, adjusting if the throw lands short of or beyond the midpoint of the base. For reference, the MLB average pop time for throws to second base is 2.01 seconds (MLB, 2024).

						Pop Time 2B (Avg)				Pop Time 3B (Avg)			
Rk.	Catcher	Team	Age	Arm	Exchange	Att.	All	CS	SB	Att.	All	CS	SB
1	Lee, Korey		25	88.3	0.64	46	1.85	1.85	1.85	2	1.49	1.56	1.42
2	Realmuto, J.T.		33	85.1	0.65	39	1.85	1.85	1.84	4	1.42	1.41	1.45
3	Bailey, Patrick		25	84.6	0.60	58	1.85	1.83	1.86	2	1.40	1.42	1.37
4	Marchán, Rafael		25	85.9	0.65	6	1.87	1.90	1.84	2	1.51	-.--	1.51
5	Torrens, Luis		28	85.2	0.61	19	1.87	1.88	1.87	1	1.37	-.--	1.37
6	Davis, Henry		24	86.7	0.65	10	1.87	1.87	1.87	2	1.44	-.--	1.44
7	Driscoll, Logan		26	85.3	0.65	6	1.87	1.87	1.87	2	1.50	-.--	1.50
8	Stubbs, Garrett		31	85.4	0.64	22	1.88	1.88	1.88	5	1.52	-.--	1.52
9	Fermin, Freddy		29	84.3	0.64	23	1.88	1.85	1.92	2	1.54	-.--	1.54
10	Moreno, Gabriel		24	83.5	0.62	26	1.89	1.90	1.88	7	1.48	1.47	1.49

Figure 3. Top Ten Catchers According to Statcast Pop Time Leaderboard

However, a statistical analysis using mixed-effects logistic regression models of MLB games (1978–1990) demonstrated that both pitchers and catchers significantly influence stolen-base attempts and success rates, with pitchers showing greater impact than catchers. Variance components revealed that stolen-base defense is a measurable skill, with 95% of pitchers and catchers having success probabilities ranging between 0.50–0.84 and 0.59–0.79, respectively, under typical conditions (Loughin et al., 2008).

So, in order to account for these effects, Statcast also quantifies catcher throwing in a more general sense, seeking to create a Caught Stealing Above Average metric to quantify the skill of catchers at throwing out runners on steal attempts, given their opportunities. This is created by assigning a probability of each steal attempt being successful or not based on the following: runner distance from second, runner speed, pitch location, pitcher/batter handedness, and awareness of pitch outs or delayed steals. Thus, the Caught Stealing Above Average is the difference between actual caught stealing and estimated caught stealing based on the attempts seen. Catcher Stealing Runs translates this to a run value using a .65 runs/CS basis, the difference between a SB (+ .2 runs) and a CS (-.45 runs) in terms of traditional linear weights for those events.

							At All Bases					feet/second	seconds	mph
Rk.	Player	Team	SB Att. (All)	Catcher Stealing Runs	Catcher's CS Above Average	CS	CS%	Est. CS	Est. CS%	CS Above Avg. / Throw	Runner Speed	Exchange Time	Arm Strength
1	Smith, Will		73	7	11	24	33%	13	17%	0.16	28.5	0.62	79.4
2	Bailey, Patrick		69	6	10	20	29%	10	15%	0.14	28.4	0.58	81.5
3	Kirk, Alejandro		69	5	8	22	32%	14	21%	0.11	28.4	0.64	76.6
4	Fermin, Freddy		27	5	7	14	52%	7	25%	0.27	28.6	0.63	82.1
5	Torrens, Luis		23	4	6	11	48%	5	22%	0.26	28.4	0.60	81.6
6	Realmuto, J.T.		57	4	5	15	26%	10	17%	0.10	28.6	0.61	82.6
7	Langeliers, Shea		84	3	5	21	25%	16	19%	0.06	28.5	0.66	83.1
8	Bethancourt, Christian		32	3	5	11	34%	6	19%	0.15	28.8	0.70	85.3
9	Raleigh, Cal		90	3	4	23	26%	19	21%	0.05	28.4	0.66	81.9
10	Kelly, Carson		57	3	4	17	30%	13	22%	0.08	28.5	0.65	78.6

Figure 4. Top Ten Catchers According to Statcast Throwing Leaderboard (Statcast, 2024)

Statcast combines these metrics into an overall Fielding Run Value, where 1 block saved = 0.25 runs, 1 strike saved = 0.125 runs, and 1 SB saved = 0.65 runs saved. We will use these Statcast metrics in our study.

Fielding

Advanced metrics have also tried to account for catcher fielding, like fielding bunts, pop ups, and tags at the plate. Bunt Runs Saved assesses a catcher's success in fielding and throwing on bunt plays, while catch probability and reaction time are used to measure effectiveness on pop-ups. Tags at the plate are evaluated using metrics like Outs Above Average (OAA), which account for throw accuracy and runner speed. These are less important to the catcher role, given the frequency at which these plays occur, but are a part of the position nonetheless.

IV. Catcher Compensation

Importance of Proper Evaluation

Baseball operations is a game of strategy and optimization, so evaluating a catcher's performance is incredibly important for building a winning team. Time spent playing catcher is a lot of wear and tear on the body, so if there is no catching ability, it could be useful to put your worst defender behind the plate. Or, teams could also switch off catchers often so that they get less wear and tear. As previous research has shown, however, this is not the case, but quantifying these measurements allows teams to accurately compensate catchers. One study analyzed MLB catchers' performance from 2016–2022, finding that strike zone runs saved (rSZ), measuring pitch receiving ability, is the most significant differentiator among defensive skills, suggesting teams should prioritize elite receivers to maximize run prevention. In theory, this is expected due to the structure of the game; catchers receive pitches much more than they do other aspects of their game (DeForest, 2023). This is also important for catchers themselves, who, in order to get the highest compensation, should focus on receiving ability.

Awards

Awards can also be associated with compensation. For years, catcher awards seemingly focused on catching stealing, likely because those are the highlight plays. The best example of this is Salvador Perez, who has won five Gold Glove awards, but is one of the worst framers in the league and generally has poor defensive metrics (In, 2021). So, in order to best determine who should win these awards, it is important to quantitatively describe these catcher traits.

V. Potential Unexplored Metrics

While Statcast and other metrics help to point towards a catcher's overall contributions to preventing runs, certain metrics have yet to be fully quantitatively described.

Game-Calling

Piper Slowinski noted that while no one has figured out how to get a definitive metric on catcher game-calling, it should matter to performance (Fangraphs, 2024). Thus, most studies pre-Statcast attempted to study it theoretically rather than quantitatively. One aspect of this is pitch sequencing, which is the practice of varying pitch types to confuse hitters and prevent runs. However, even just optimal pitch locations, which are most often called for by catchers, can have significant effects. In general, one issue with most public baseball datasets is that they do not include the location data for where the catcher demands a ball, which is essential for obtaining the intent of a catcher in their pitch calling. Using Japanese professional baseball data that included the catcher's request location, researchers found that outside pitching attempts performed significantly better than inside pitching attempts in terms of minimizing allowed run average. The maximum contribution of saving runs by selecting an appropriate pitching strategy was estimated at 0.47 runs/game, demonstrating that pitch calling can be a significant part of the game (Nakahara, 2023). Research of Yadier Molina's performance in the mid-2010s found that the spread of game-calling performance over the course of a season is probably around -10 to +10 runs (Lindbergh, 2014).

Even with this research, we don't know much about who really deserves credit for calling a particular pitch: the pitcher, the catcher, or the coaching staff. Most data also doesn't account for pitchers overruling catchers and choosing another pitch and doesn't account for sign stealing, advanced scouting reports, or what the right pitch distribution should be in any one situation.

Pitcher-Catcher Comfortability

Another aspect of a catcher's defense is comfortability with the pitchers he catches for. A pitcher-catcher battery with more comfortability might be able to better execute a game plan, and trust might allow the pitcher to take more risks for the better. Using a two-way fixed effects model, researchers from Western Kentucky University's Department of Economics assessed the significance of these pitcher-catcher pairings. Their findings indicated that the specific combination of pitcher and catcher (match effects) is nearly as crucial as the individual pitcher's skill in securing outs and more influential than the catcher's individual effect. Regarding strikeouts, match effects surpassed the impact of the catcher, though they were not as significant as the pitcher's contribution. The study also explored factors that enhance the effectiveness of these pairings. It was observed that match effects improved when both players share the same country of origin, particularly if they came from predominantly Spanish-speaking countries, or when they have accumulated more shared playing experiences. These findings suggested that cultural and experiential similarities contribute to better on-field collaboration (Biolsi et al., 2022). Given this research, teams should maximize these collaborations.

It is also worth noting that although catchers might join teams with certain traits, these are also to become spots of emphasis for player development, and as teams explore game-calling or pitcher-catcher relationships, these effects will become more obvious in the data (Kuty, 2022).

VI. Objectives Of Our Study

Our study will explore these metrics which are not described by Statcast yet. We will use four main models to isolate a catcher's hidden traits by accounting for catcher framing, blocking, and throwing.

Gap in Research

While previous research has hinted at the importance of catcher game-calling and comfortability with pitchers, they have not been quantitatively described in a way that incorporates framing, blocking, and throwing with new Statcast data. We seek to bridge that gap and be able to rank catchers based on these latent traits.

Objectives

1. Determine if catchers have a latent effect on run prevention that is not captured by framing, throwing, and blocking.
2. If there is a latent effect on run prevention by catchers, we would like to quantify it to rank catchers based on these unobserved traits.

Methodology

Description of Data

We utilized complete pitch-by-pitch data from Baseball Savant, Statcast's platform, spanning the 2021 to 2024 seasons. This time frame was selected to account for post-COVID rule changes and to ensure a sufficient dataset for identifying significant trends. Note that our models use the following variables:

<u>xFIP-:</u> Expected Fielding Independent Pitching Minus, scaled to league average, which isolates a pitcher's performance by focusing on strikeouts, walks, and home runs while neutralizing factors like defense and ballpark effects.

<u>RA-:</u> Runs Allowed Minus, a team-level statistic measuring the effectiveness of pitchers and catchers in preventing runs, adjusted for league context.

<u>DRS:</u> Defensive Runs Saved, a comprehensive metric quantifying a team's or player's defensive contributions to run prevention.

<u>FRV:</u> Fielding Run Value, a Statcast metric that for catchers combines framing, throwing, and blocking into a single metric.

Procedure

We created four models to study whether a catcher has a latent effect on run prevention:

1. Logistic Regression predicting strikeouts based on quantifiable information about catchers, batters, and pitchers. The residuals from this model represent a combination of random noise and unaccounted latent effects. To determine if these residuals reflect meaningful patterns, we analyzed their year-over-year correlations. A positive correlation suggests the presence of factors influencing strikeouts that are not captured by the model or explained by random variability.
2. Mixed Effects Model predicting xFIP- for pitcher-catcher duos, using the pitcher's season-long xFIP- as a baseline and treating the catcher as a random effect. A framing term was included to quantitatively account for framing contributions. This model attributes any remaining variability in xFIP- to the catcher's latent effects, enabling us to rank catchers based on these unmeasured contributions.
3. Bayesian Hierarchical Model predicting xFIP- for pitcher-catcher duos, incorporating the pitcher's season-long xFIP-, catcher, team, and framing as factors. This model allows us to disentangle the influence of a catcher's latent abilities from the impact of team-level factors, such as the effectiveness of analytics departments in developing game plans. By accounting for both individual and organizational effects, this approach provides deeper insights into the combined contributions of catchers and team strategies.
4. Mixed Effects Model predicting RA- for pitcher-catcher duos, using the pitcher's season-long RA- as a baseline and incorporating Defensive Runs Saved (DRS) by the team to account for overall defensive contributions. By focusing on runs allowed—a direct measure of success or failure in baseball—this model evaluates the interplay between pitcher, catcher, and team defense in limiting runs, providing a holistic view of run prevention.

Results

I. Preliminary Models: Logistic Regression on Strikeouts

Our initial models aimed to predict the likelihood of a strikeout using quantitative predictors related to the pitcher, catcher, and batter. By incorporating these variables, the residuals represented any unexplained variability, capturing factors not accounted for by the model. Our primary objective was to develop a model that aligned with the structure of a baseball plate appearance and appropriately reflected the relationships in the data.
Our preliminary models were the following:

Model 0.1

$$P(\text{Strikeout} = 1 \mid \text{Pitcher K\%}, \text{Batter K\%}, \text{Throwing}, \text{Blocking}, \text{Framing}, \text{FRV}) = \frac{1}{1+\exp(-\eta)}$$

where:

$$\eta = \beta_0 + \beta_1 \cdot \text{Pitcher K\%} + \beta_2 \cdot \text{Batter K\%} + \beta_3 \cdot \text{Throwing} + \beta_4 \cdot \text{Blocking} + \beta_5 \cdot \text{Framing} + \beta_6 \cdot \text{FRV}$$

```
glm(formula = strikeout ~ pitcher_k_rate + batter_k_rate + Throwing +
    Blocking + Framing + FRV, family = "binomial", data = df_all)

Coefficients:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)    -3.882943   0.020189 -192.325   <2e-16 ***
pitcher_k_rate  5.608776   0.062279   90.058   <2e-16 ***
batter_k_rate   5.874066   0.055145  106.521   <2e-16 ***
Throwing       -0.007251   0.005877   -1.234    0.217
Blocking       -0.004923   0.005841   -0.843    0.399
Framing        -0.003728   0.005692   -0.655    0.512
FRV             0.004817   0.005622    0.857    0.392
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 580003  on 543031  degrees of freedom
Residual deviance: 560209  on 543025  degrees of freedom
  (940 observations deleted due to missingness)
AIC: 560223
```

Figure 5: Summary Statistics for Model 0.1

This model predicted strikeouts using pitcher and batter strikeout rates, throwing, blocking, framing, and Fielding Run Value (FRV). We chose these variables to use in our initial model because we wanted to track everything that should have an effect on a pitcher-catcher duo getting a strikeout. Pitchers bring their K%, batters bring their K%, and catchers bring their skills that can contribute to strikeouts. For the catcher metrics, we used Statcast’s Catcher Framing Runs, Catcher Block Runs, Catcher Stealing Runs, and FRV. We used these metrics throughout the rest of this paper as well. While blocking and throwing won’t directly affect strikeouts by gaining strikes quantitatively, they may allow pitchers to be more confident pitching with men on base, indirectly contributing to strikeouts. However, in this model, only pitcher and batter K% were the only significant predictors, while catcher skills like framing and blocking have negligible impact. However, FRV is a statistic that combines aspects of throwing, blocking, and framing, so it likely caused multicollinearity.

Model 0.2

$$P(\text{Strikeout} = 1 \mid \text{Pitcher K\%}, \text{Batter K\%}, \text{Throwing}, \text{Framing}) = \frac{1}{1 + \exp(-\eta)}$$

where: $\eta = \beta_0 + \beta_1 \cdot \text{Pitcher K\%} + \beta_2 \cdot \text{Batter K\%} + \beta_3 \cdot \text{Throwing} + \beta_4 \cdot \text{Framing}$

```
glm(formula = strikeout ~ pitcher_k_rate + batter_k_rate + Throwing +
    Framing, family = "binomial", data = df_all)

Coefficients:
                 Estimate Std. Error  z value Pr(>|z|)
(Intercept)    -3.8829773  0.0201858 -192.361   <2e-16 ***
pitcher_k_rate  5.6087416  0.0622781   90.060   <2e-16 ***
batter_k_rate   5.8739987  0.0551443  106.521   <2e-16 ***
Throwing       -0.0023851  0.0014803   -1.611   0.1071
Framing         0.0011102  0.0005787    1.918   0.0551 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 580003  on 543031  degrees of freedom
Residual deviance: 560209  on 543027  degrees of freedom
  (940 observations deleted due to missingness)
AIC: 560219
```

Figure 6: Summary Statistics for Model 0.2

Our next model took out FRV–to reduce multicollinearity–and blocking; in theory, blocking shouldn't contribute to the ability to get a strikeout. Framing and throwing became more significant, although they weren't below a typical threshold of α = 0.05. Pitcher and batter strikeout percentages were still significant.

Model 0.3

$$P(\text{Strikeout} = 1 \mid \text{Pitcher K\%}, \text{Batter K\%}, \text{Framing}) = \frac{1}{1 + \exp(-\eta)}$$

where: $\eta = \beta_0 + \beta_1 \cdot \text{Pitcher K\%} + \beta_2 \cdot \text{Batter K\%} + \beta_3 \cdot \text{Framing}$

```
glm(formula = strikeout ~ pitcher_k_rate + batter_k_rate + Framing,
    family = "binomial", data = df_all)

Coefficients:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)   -3.8834450  0.0201837 -192.405   <2e-16 ***
pitcher_k_rate  5.6091839  0.0622765   90.069   <2e-16 ***
batter_k_rate   5.8740415  0.0551443  106.521   <2e-16 ***
Framing         0.0010908  0.0005784    1.886   0.0593 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 580003  on 543031  degrees of freedom
Residual deviance: 560212  on 543028  degrees of freedom
  (940 observations deleted due to missingness)
AIC: 560220
```

Figure 7: Summary Statistics for Model 0.3

Model 0.3 took out throwing as it shouldn’t have too great an effect on the ability to get a strikeout, and was not significant in Model 0.2. In this model, pitcher and catcher strikeout percentages were significant again, and framing was very nearly significant.

Model 0.4

$$P(\text{Strikeout} = 1 \mid \text{Pitcher K\%}, \text{Batter K\%}, \text{Framing}, \text{Season}) = \frac{1}{1 + \exp(-\eta)}$$

where: $\eta = \beta_0 + \beta_1 \cdot \text{Pitcher K\%} + \beta_2 \cdot \text{Batter K\%} + \beta_3 \cdot \text{Framing} + \beta_4 \cdot \text{Season}$

```
glm(formula = strikeout ~ pitcher_k_rate + batter_k_rate + Framing +
    as.factor(season), family = "binomial", data = df_all)

Coefficients:
                      Estimate Std. Error  z value Pr(>|z|)
(Intercept)          -3.9235821  0.0213961 -183.378  < 2e-16 ***
pitcher_k_rate        5.6350525  0.0624550   90.226  < 2e-16 ***
batter_k_rate         5.8776290  0.0551658  106.545  < 2e-16 ***
Framing               0.0009553  0.0005795    1.649   0.0992 .
as.factor(season)2022  0.0496637  0.0095657    5.192 2.08e-07 ***
as.factor(season)2023  0.0368291  0.0094022    3.917 8.96e-05 ***
as.factor(season)2024  0.0447489  0.0094154    4.753 2.01e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 580003  on 543031  degrees of freedom
Residual deviance: 560179  on 543025  degrees of freedom
  (940 observations deleted due to missingness)
AIC: 560193
```

Figure 8: Summary Statistics for Model 0.4

In Model 0.4, we added a season factor that acted as a coefficient on the seasons, with 2021 serving as the reference season. Based on this model, where we observed that season factors were statistically significant, we figured it would likely be better to separate our future logistic regression models by season.

II. Model 1: Logistic Regression on Strikeouts

Model 1.1

After careful consideration and a further investigation of our data, we found that pitchers often pitched to just one or two catchers over the course of a season, so a catcher’s latent effect would be baked into that pitcher’s K%. So, in this model, we sought to predict strikeouts based on distinct statistics from the three players involved in whether a batter strikes out or not: the batter, the catcher, and the pitcher. The pitcher’s contribution is modeled by Pitching+, a Fangraphs statistic that combines “stuff” and location to get the performance of a pitcher based on what only they can control: the movement of the ball. We kept Batter K% as modeling our batting’s contribution, and kept framing as the contribution from the catcher. We assumed that umpire contributions would act as random noise, and analyzing year-over-year residuals would allow us to isolate and evaluate these effects.

Our all-year logistic regression model was specified as:

$$P(\text{Strikeout} = 1|\text{Pitching+}, \text{Batter K\%}, \text{Catcher Framing}) = \frac{1}{1 + \exp(-\eta)}$$

Where:

$$\eta = \beta_0 + \beta_1 \cdot \text{Pitching+} + \beta_2 \cdot \text{Batter K\%} + \beta_3 \cdot \text{Catcher Framing}$$

```
glm(formula = strikeout ~ sp_pitching + batter_k_rate + Framing,
    family = "binomial", data = df_all)

Coefficients:
                Estimate Std. Error  z value Pr(>|z|)
(Intercept)   -5.8137576  0.0565603 -102.789   <2e-16 ***
sp_pitching    0.0327433  0.0005508   59.444   <2e-16 ***
batter_k_rate  5.6090710  0.0421786  132.984   <2e-16 ***
Framing        0.0048333  0.0005095    9.486   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 767289  on 716220  degrees of freedom
Residual deviance: 745100  on 716217  degrees of freedom
  (2384 observations deleted due to missingness)
AIC: 745108
```

Figure 9: Summary Statistics for All-Year Logistic Regression Model

However, as mentioned in our preliminary models, creating a distinct model for each season would provide more accurate and tailored insights. Seasonal models account for potential year-to-year variations, such as changes in league-wide trends, player performance, and environmental factors like rule changes or shifts in offensive and defensive strategies. By isolating each season, we aimed to reduce the impact of confounding variables and better identify meaningful patterns in the data. Therefore, we developed four separate models, each tailored to its respective season, with the summary statistics for each year detailed below to highlight these variations.

```
Summary for model: m5.2021

Call:
glm(formula = strikeout ~ sp_pitching + batter_k_rate + Framing,
    family = "binomial", data = season_data)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.905643   0.109791 -53.790   <2e-16 ***
sp_pitching    0.034108   0.001069  31.915   <2e-16 ***
batter_k_rate  5.384411   0.073693  73.065   <2e-16 ***
Framing        0.001738   0.001179   1.474    0.141
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 191981  on 177200  degrees of freedom
Residual deviance: 185173  on 177197  degrees of freedom
  (783 observations deleted due to missingness)
AIC: 185181
```

```
Summary for model: m5.2022

Call:
glm(formula = strikeout ~ sp_pitching + batter_k_rate + Framing
    family = "binomial", data = season_data)

Coefficients:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)   -6.242815   0.115973 -53.830  < 2e-16 ***
sp_pitching    0.036809   0.001129  32.593  < 2e-16 ***
batter_k_rate  5.700053   0.086765  65.695  < 2e-16 ***
Framing        0.007036   0.001100   6.399 1.56e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 186479  on 175323  degrees of freedom
Residual deviance: 180964  on 175320  degrees of freedom
  (840 observations deleted due to missingness)
AIC: 180972
```

```
Summary for model: m5.2023

Call:
glm(formula = strikeout ~ sp_pitching + batter_k_rate + Framing,
    family = "binomial", data = season_data)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.6343802  0.1093868 -51.509  < 2e-16 ***
sp_pitching    0.0307062  0.0010630  28.888  < 2e-16 ***
batter_k_rate  5.7400697  0.0896435  64.032  < 2e-16 ***
Framing        0.0043340  0.0008775   4.939 7.85e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 194471  on 181687  degrees of freedom
Residual deviance: 189379  on 181684  degrees of freedom
  (594 observations deleted due to missingness)
AIC: 189387
```

```
Summary for model: m5.2024

Call:
glm(formula = strikeout ~ sp_pitching + batter_k_rate + Framing,
    family = "binomial", data = season_data)

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)   -5.5084341  0.1182223 -46.594  < 2e-16 ***
sp_pitching    0.0294097  0.0011527  25.515  < 2e-16 ***
batter_k_rate  5.7381133  0.0908913  63.132  < 2e-16 ***
Framing        0.0059530  0.0009976   5.967 2.41e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 194320  on 182007  degrees of freedom
Residual deviance: 189526  on 182004  degrees of freedom
  (167 observations deleted due to missingness)
AIC: 189534
```

Figure 10. Summary Statistics by Year for Logistic Regression Models

Based on these outputs, we found that Pitching+ and Batter K% were consistently significant predictors of strikeouts across all seasons. However, Framing was only statistically significant from 2022 to 2024. This suggests that the lack of statistical significance for Framing in the All-Years model was likely driven by its insignificance in 2021. It is unknown why framing was less important to strikeouts that year, but it is an area for future research. Now that we had models for each year predicting strikeouts from a set of significant predictors (except for 2021’s Framing coefficient), we analyzed year-over-year residuals to uncover whether the values not accounted for by our model stayed consistent for certain catchers.

Residuals 2023 vs. 2024

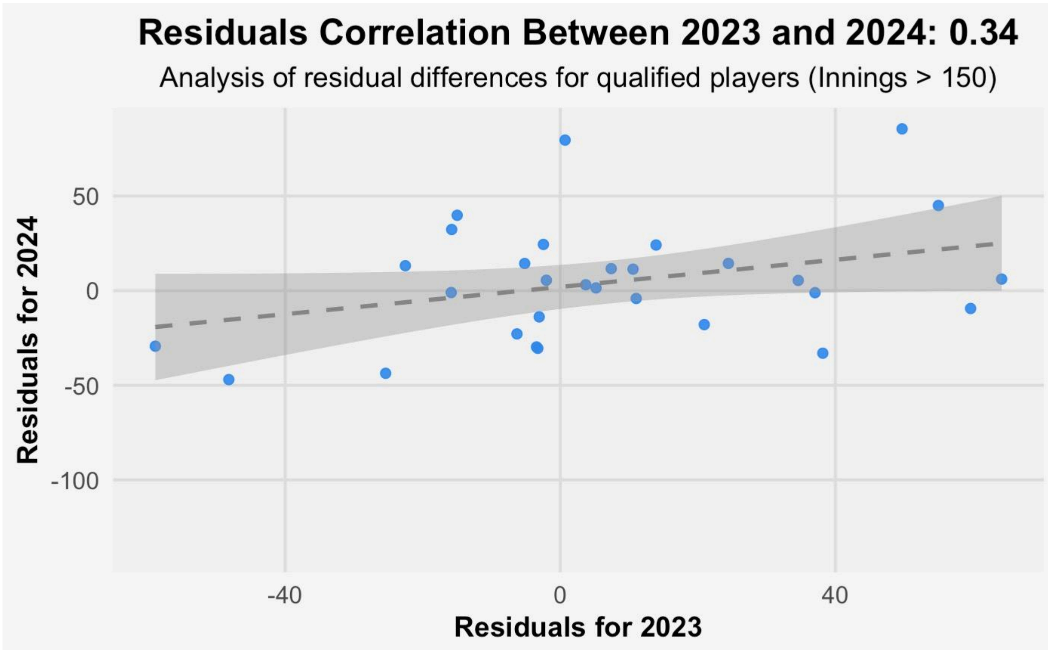

Figure 11: Residuals Correlation Between 2023 and 2024

Figure 11 illustrates, for 29 catchers who caught over 150 innings in both 2023 and 2024, the correlation between residuals from our strikeout prediction model, representing unexplained variability after accounting for pitcher, batter, and catcher contributions. The scatter plot shows a positive linear relationship with a correlation coefficient of 0.34, indicating moderate consistency in the residuals across seasons. This suggests the presence of latent factors that persist year-over-year. The shaded confidence interval around the trendline confirms the reliability of this relationship, though some uncertainty exists at the extremes.

Residuals 2022 vs. 2023

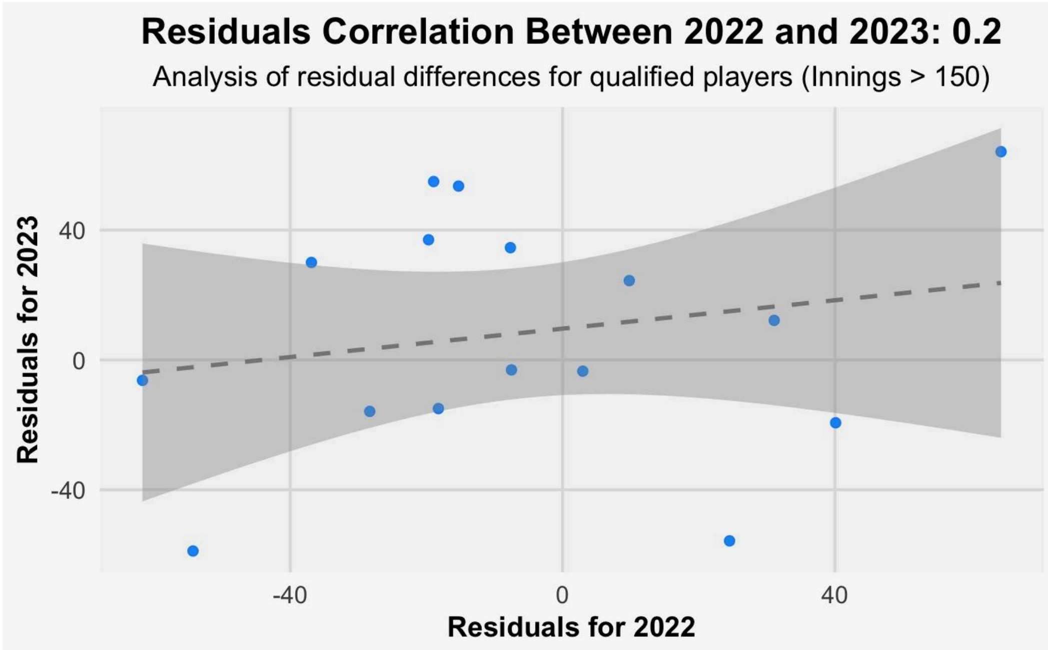

Figure 12: Residuals Correlation Between 2022 and 2023

Figure 12 shows, for 16 catchers who caught over 150 innings in both 2022 and 2023, the correlation between residuals from our strikeout prediction model. The dashed trendline demonstrates a weak positive linear relationship, with a correlation coefficient of 0.2. This low correlation suggests minimal consistency in the unexplained variability between these two years. The reason for this lack of consistency is discussed in Takeaway 6 of the Conclusion section. The broad shaded confidence interval around the trendline highlights the uncertainty in the linear relationship, particularly at the extremes.

Residuals 2021 vs. 2022

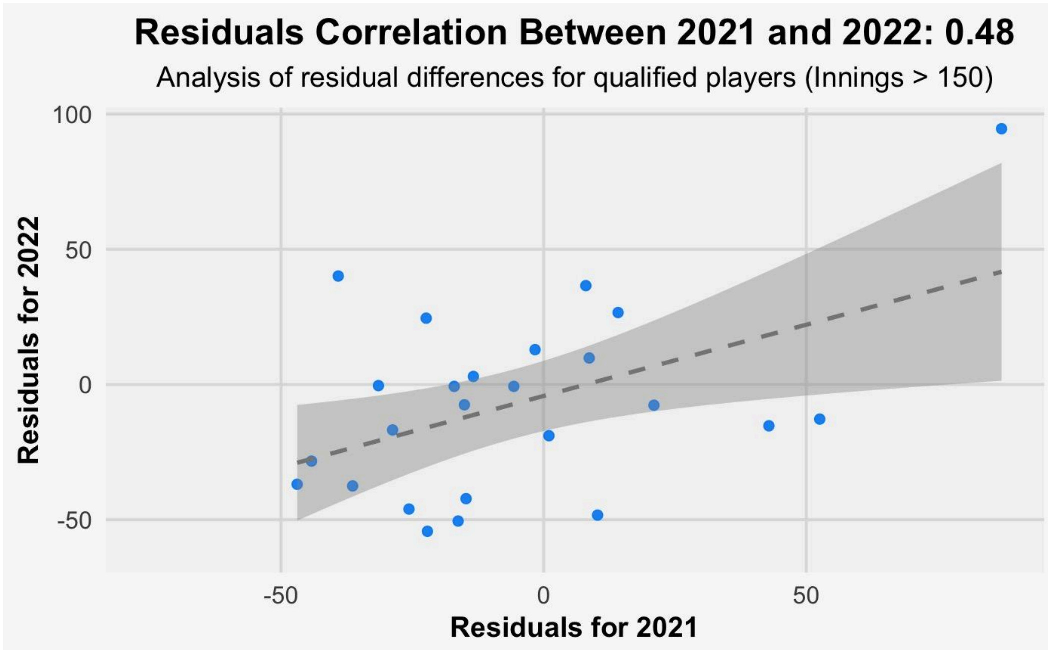

Figure. 13: Residuals Correlation Between 2021 and 2022

Figure 13 displays, for 25 catchers who caught over 150 innings in both 2021 and 2022, the correlation between residuals from our strikeout prediction model. The scatter plot shows a moderate positive linear relationship, with a correlation coefficient of 0.48, indicating greater consistency in the unexplained variability compared to other year-to-year comparisons, but not much better than the 2023-2024 residuals comparison. The dashed trendline highlights this moderate correlation, while the shaded confidence interval shows some uncertainty.

Based on these three residual graphs, there is strong evidence supporting the hypothesis that latent factors, such as catcher game-calling, contribute to year-over-year consistency in strikeout outcomes. However, the varying levels of correlation between years suggest that these latent effects are influenced by external changes, such as evolving gameplay dynamics or rule modifications, which may impact their stability over time.

III. Model 2: Mixed Effects Model on xFIP-

Mixed-effects modeling was critical in this analysis for several reasons. First, pitchers often throw to multiple catchers within a season, creating a nested data structure where observations are naturally grouped under individual catchers. Mixed-effects models effectively account for this nesting by modeling within-catcher correlations, ensuring more accurate estimates. Additionally, from the perspective of unmeasured latent effects, the inclusion of random effects enables us to capture variability attributable to catchers' latent contributions, such as intangible factors like game-calling ability or chemistry with pitchers. The random effect coefficient for each catcher serves as a proxy for their latent contribution to xFIP-, providing a measurable indicator of their impact. By partitioning variance between fixed and random effects, this approach not only enhances generalizability but also allows us to make more robust predictions that apply across diverse pitcher-catcher pairings, strengthening the overall reliability of the model.

Model 2.1

In this model, we sought to predict xFIP- (Expected Fielding Independent Pitching scaled to league average) based on key contributions from both the catcher and pitcher, while accounting for latent effects inherent to specific catchers. Using a statistic like xFIP-, which only considers a pitcher's fly ball rate, walks, and strikeouts, allowed us to continue to disregard the defense in our calculations. Unlike a straightforward logistic regression, a mixed-effects approach was necessary here to handle the hierarchical structure of our data, where multiple observations were nested under individual catchers. By treating Name (the catcher) as a random effect, we aimed to capture unmeasured catcher-specific influences that might explain variability in xFIP- beyond the included fixed effects.

Our first mixed-effects model was specified as:

$$xFIP\_minus \sim FB.season + BB.season + K.season + Throwing + Blocking + Framing + (1|Name)$$

FB.season, BB.season, and K.season represent the pitcher's flyball rate, walk rate, and strikeout rate for the season, respectively, capturing the pitcher's skill and tendencies. Throwing, Blocking, and Framing represent catcher-specific defensive contributions. The random effect (1|Name) captures the unexplained variance attributable to each catcher.

Interestingly, the random effect associated with the catcher (Name) was estimated with a coefficient of 0 for each catcher in this model. This outcome suggests that the variability in xFIP- attributable to individual catchers is already explained by the fixed effects included in the model. Similar to the influence of pitcher strikeout rate observed in Model 0.4, the walk rate, flyball rate, and strikeout rate are shaped by the latent effects of catchers, likely because pitchers typically throw to only a few catchers throughout a season. As a result, the catcher's latent impact is effectively baked into these fixed effects, leaving little residual variability to be captured by the random effect.

Model 2.2

To further refine our analysis, we introduced a second model designed to decompose the contributions of season-wide pitcher performance and catcher-specific latent effects. The model was specified as:

$$xFIP\_\text{minus} \sim xFIP\_\text{minus.season} + \text{usage} + \text{Framing} + (1|\text{Name:season})$$

xFIP_minus.season captures the pitcher's season-long xFIP- without considering specific catcher contributions, serving as a baseline for the pitcher's independent performance. "Usage" measures the frequency of the pitcher-catcher pairing and may account for small differences in xFIP- due to how often a catcher is utilized. Specifically, if a catcher is used less frequently, their coefficient may appear more extreme, reflecting greater variability in their influence over a smaller sample size. The random effect (1|Name:season) isolates the unique influence of each catcher within a specific season, accounting for catcher-year variability.

Top 5 Catchers by Season (2024 to 2021)

Rank	2024	2023	2022	2021
1	Jacob Stallings	Salvador Perez	Elias Diaz	Sean Murphy
2	Travis d'Arnaud	Kyle Higashioka	Kyle Higashioka	Elias Diaz
3	Kyle Higashioka	Sean Murphy	Austin Hedges	Kyle Higashioka
4	Danny Jansen	Yasmani Grandal	Yasmani Grandal	Austin Hedges
5	Austin Hedges	Austin Hedges	Martin Maldonado	J.T. Realmuto

Figure 14: Top 5 Catchers by Season (Mixed Effects Model on xFIP-)

Bottom 5 Catchers by Season (2024 to 2021)

Rank	2024	2023	2022	2021
1	Christian Vazquez	Ryan Jeffers	Ryan Jeffers	Alejandro Kirk
2	J.T. Realmuto	Elias Diaz	James McCann	Jacob Stallings
3	Martin Maldonado	Jose Trevino	Alejandro Kirk	Jose Trevino
4	Salvador Perez	Travis d'Arnaud	Travis d'Arnaud	Travis d'Arnaud
5	Willson Contreras	Willson Contreras	Will Smith	Will Smith

Figure 15: Bottom 5 Catchers by Season (Mixed Effects Model on xFIP-)

The rankings in Figures 14 and 15 provide a compelling perspective on latent catcher contributions, with significant implications for how catchers are valued. The top 5 catchers consistently include players known for their leadership and defensive prowess, aligning with expectations. Conversely, the bottom 5 catchers feature players who are generally considered weaker defensively or less impactful behind the plate. This separation highlights the effectiveness of the model in identifying latent contributions that may not be fully captured by traditional defensive metrics. We observed that certain players, such as Travis d'Arnaud and Martin Maldonado, appear in both the top and bottom rankings across different years, warranting further analysis. Maldonado's performance shift may be attributed to his team change between the 2023 and 2024 offseasons, which could have impacted his synergy with new pitchers or his ability to game plan. Similarly, d'Arnaud's improved ranking aligns with his fifth year with the Braves, suggesting that his growing familiarity and rapport with the pitching staff may have started to positively influence his latent contributions. These cases underscore the dynamic nature of pitcher-catcher relationships and highlight the need for a deeper exploration of how team context and evolving chemistry impact catcher performance over time. Thus, Model 2.3 will incorporate teams as a random effect.

It is noteworthy that these rankings are not merely reflective of typical defensive statistics, but rather the unique latent effects each catcher brings to their pitcher-catcher partnerships. For example, while J.T. Realmuto is widely regarded as an elite defender by Statcast metrics, his relatively poor ranking in our model suggests that his latent effects, as captured here, may differ from his measurable defensive skills. Similarly, Salvador Perez, who often struggles in traditional defensive evaluations, performs well in our model, potentially due to intangible qualities like game-calling or rapport with pitchers. These findings suggest that latent contributions, distinct from measurable defensive metrics, play a critical role in evaluating catchers' overall value and their impact on run prevention. Specific cases will be explored further in the Conclusion section.

Model 2.3

In Model 2.3, we extended the predictors from previous models by incorporating a mixed effect for the pitcher's team, specified as (1|pitcher_team). This addition aimed to capture latent variability at the team level, recognizing that a team's analytics infrastructure, game-planning, and overall pitching strategy can significantly influence catcher performance.

$$xFIP\_\text{minus} \sim xFIP\_\text{minus.season} + \text{usage} + \text{Framing} + (1|\text{Name:season}) + (1|\text{pitcher\_team})$$

Interestingly, however, the random effect for the team in Model 2.3 was estimated with no significant coefficient, similar to what we observed in Model 2.1. This result suggested that variability between teams is already accounted for by the fixed effects in the model, particularly metrics like xFIP_minus.season, usage, and Framing. In other words, while team dynamics likely play a role, their influence may already be baked into the observed season-wide pitcher performance and other catcher-specific contributions. Thus, the addition of the team-level random effect did not significantly improve the model's ability to explain the remaining variance.

IV. Model 3: Bayesian Hierarchical Model on xFIP-

The Bayesian approach offers significant advantages for modeling complex hierarchical data, especially when accounting for interactions and random effects, such as catcher-team dynamics. Unlike traditional frequentist methods, which produce point estimates and often ignore the full distribution of possible values, Bayesian models incorporate uncertainty into the estimation process by providing posterior distributions for parameters. This allows us to quantify not just the most likely value of an effect but also the range and confidence associated with it, which is critical when random effects, such as team influence, may seem negligible in frequentist models due to insufficient data or overfitting concerns.

In this context, uncertainty reflects the variability and lack of precision in estimating catcher or team-level effects, given the data at hand. For example, if data on a specific catcher-team pairing is sparse, the Bayesian approach accounts for this uncertainty by widening the posterior distribution, signaling lower confidence in precise effect estimates. This framework allows us to better capture subtle interactions between catchers and teams, which Model 2.2 highlighted as an area requiring further refinement. Additionally, Bayesian methods regularize extreme estimates by “shrinking” them toward the overall mean, reducing the risk of overfitting and ensuring a more balanced and interpretable understanding of the hierarchical relationships in the data. This nuanced approach makes Bayesian modeling particularly well-suited for exploring complex dynamics in catcher-pitcher-team interactions.

Model 3.1

$$xFIP\text{\_minus} \sim xFIP\text{\_minus.season} + \text{usage} + \text{Framing} + (1|\text{Name:season}) + (1|\text{pitcher\_team}) + (1|\text{Name:pitcher\_team})$$

This representation included fixed effects (xFIP_minus.season, usage, and Framing) and random effects to capture variability at multiple levels: (1|Name:season) for individual catchers within seasons, (1|pitcher_team) for team-specific effects, and (1|Name:pitcher_team) for interactions between catchers and teams.

Top 5 Teams Over 2021-2024 Seasons

Team	Mixed Effect Value
WSH	−0.10296551
CLE	−0.10117262
DET	−0.07825696
PIT	−0.05965774
AZ	−0.04538583

Bottom 5 Teams Over 2021-2024 Seasons

Team	Mixed Effect Value
MIN	0.04541245
TB	0.05017947
SD	0.05460133
CHC	0.11056326
LAD	0.14696250

Figure 16: Top/Bottom 5 Teams (Bayesian Hierarchical Model on xFIP-)

Inverting the values by multiplying the data by -1 to align with the interpretation of FIP makes conceptual sense because, in baseball metrics, a lower FIP (Fielding Independent Pitching) generally indicates better performance. By adjusting the data this way, teams with lower xFIP_minus mixed-effect values represent those that contribute positively to reducing pitchers' expected runs, making them "better" in this context.

However, the results challenge conventional expectations. Teams like the Tampa Bay Rays and Los Angeles Dodgers, widely regarded for their cutting-edge analytics departments, rank surprisingly low in terms of catcher contributions to xFIP-. This unexpected outcome raises important questions about how organizational strategies influence xFIP-. For instance, it may reflect differences in how these teams manage their pitching staffs, prioritize defensive metrics, or implement game strategies that interact uniquely with catcher performance. Notably, the primary goal of baseball is not to minimize xFIP-, but to prevent runs. Teams focused on limiting soft contact or inducing weak batted balls may succeed in run prevention in ways that are not captured by xFIP-, which only encompasses strikeouts, walks, and fly balls. These findings highlight the complexity of the relationship between team analytics and on-field performance, suggesting that even teams with strong analytical capabilities may make strategic trade-offs in areas like framing and game-calling to achieve broader defensive goals.

It is also worth noting that our ranking of catchers was essentially equivalent to Model 2.2, as we explain in Takeaway 5 of the Conclusion section, signalling that the hierarchical structure and Bayesian approach introduced in Model 3.1 primarily serve to refine the uncertainty estimates and provide a more robust framework for understanding the latent effects of catchers and teams. While the rankings themselves remained consistent with those generated by Model 2.2, the Bayesian framework adds value by quantifying the reliability of these rankings.

V. Model 4: Mixed Effects Model on Runs Allowed Minus

Runs Allowed Minus (RA-) serves as a critical measure of a team's ability to prevent runs, encompassing contributions from pitchers, catchers, and overall defensive performance. We decided to model RA- to better understand how the interplay between a catcher and pitcher impacts run prevention, as this provides a more comprehensive view of their effectiveness compared to metrics focused solely on certain pitching outcomes.

Model 4.1

$RA\_\text{minus} \sim RA\_\text{minus.season} + \text{usage} + \text{Framing} + \text{Blocking} + \text{Throwing} + \text{DRS} + (1|\text{Name:season}) + (1|\text{pitcher\_team})$

We included key catcher-specific metrics such as Framing, Blocking, and Throwing, which capture the defensive skill set of the catcher. Additionally, we incorporated Defensive Runs Saved (DRS) for the pitching team to account for the broader defensive contributions beyond the pitcher-catcher pairing. The random effect (1|Name:season) captures variability specific to each catcher-season pairing, further isolating the unique impact of catchers in the context of team defense. The team-level random effect (1|pitcher_team) had a non-zero coefficient, indicating that organizational strategies and overall team defense significantly contribute to run prevention. Again, we created a table of the top and bottom 5 catchers by their catcher-season random effect.

Top 5 Catchers by Season (2024 to 2021)

Rank	2024	2023	2022	2021
1	Elias Diaz	Jonah Heim	Ryan Jeffers	Jonah Heim
2	Jacob Stallings	Alejandro Kirk	Alejandro Kirk	Kyle Higashioka
3	Austin Hedges	Will Smith	Austin Hedges	Elias Diaz
4	Jonah Heim	Jose Trevino	Jonah Heim	J.T. Realmuto
5	Ryan Jeffers	J.T. Realmuto	J.T. Realmuto	Jose Trevino

Figure 17: Top 5 Catchers By Season (Mixed Effects Model on Runs Allowed)

Bottom 5 Catchers by Season (2024 to 2021)

Rank	2024	2023	2022	2021
1	Travis d'Arnaud	Danny Jansen	James McCann	Travis d'Arnaud
2	Sean Murphy	Austin Hedges	Travis d'Arnaud	Christian Vazquez
3	Danny Jansen	Elias Diaz	Kyle Higashioka	Salvador Perez
4	Christian Vazquez	Travis d'Arnaud	Christian Vazquez	Danny Jansen
5	J.T. Realmuto	Kyle Higashioka	Elias Diaz	Will Smith

Figure 18: Bottom 5 Catchers By Season (Mixed Effects Model on Runs Allowed)

The results from the catcher rankings are surprising compared to earlier estimates, revealing unexpected trends in performance. Players like Kyle Higashioka, who was previously ranked highly, experienced a significant drop after the 2021 season and now appears near the bottom of the list. Similarly, James McCann, another previously solid performer, is ranked among the bottom catchers in later seasons. In contrast, players like Alejandro Kirk consistently rank near the top.

The rankings also underscore a notable lack of consistency across seasons. Many catchers, including Higashioka, show significant fluctuations in performance, suggesting variability in their effectiveness or the influence of evolving team dynamics and pitcher-catcher synergies. This inconsistency raises the possibility that our Runs Allowed Minus (RA-) model may have introduced additional noise or unaccounted-for variables, complicating the interpretation of these results. This highlights the challenges of isolating latent effects and suggests that further refinement of the model is necessary to better capture the underlying factors influencing catcher performance.

Top 5 and Bottom 5 Teams in Model 4 (Values in Scientific Notation)

Top 5 Teams		**Bottom 5 Teams**	
Team	**Value**	**Team**	**Value**
PIT	-7.61×10^{-8}	NYM	2.89×10^{-8}
TEX	-6.83×10^{-8}	STL	3.09×10^{-8}
TB	-4.22×10^{-8}	NYY	3.30×10^{-8}
PHI	-3.03×10^{-8}	LAA	3.66×10^{-8}
DET	-2.81×10^{-8}	SF	3.92×10^{-8}

Figure 19: Top/Bottom 5 Teams (Mixed Effects Model on Runs Allowed)

This table is a markedly different output than we got in Model 3 for team effects, signalling that there are likely other variables that we are not accounting for in these coefficients.

Latent Traits: Independent of Statcast's Defensive Metrics

To ensure that the latent effects derived from Model 3.1 were not simply a derivative of existing catcher statistics, we analyzed their correlation with Statcast's blocking, framing, and throwing metrics. Initially, we examined the relationships among the three Statcast metrics themselves to test the notion that "the best defensive catchers excel in all areas." Interestingly, the inter-metric correlations were generally weak, with absolute values below 0.09, except for framing and blocking, which had a modest but significant correlation of 0.263. This suggests that catchers who excel in one defensive area, such as blocking, are not necessarily the best in others, such as throwing, challenging the assumption of universal defensive excellence among elite catchers.

We then investigated the relationship between our latent trait metric and each of the three Statcast metrics. As shown in Figures 20, 21, and 22, the correlations were negligible—values of -0.01, -0.09, and 0.01 for framing, throwing, and blocking, respectively. These results confirmed that the latent trait metric captured an entirely distinct dimension of catcher performance, separate from measurable defensive skills. This independence reinforced the idea that the latent metric reflects nuanced contributions, such as game-calling or pitcher-catcher synergy, which are not directly quantifiable through existing metrics.

These findings are consistent with the results of our models, where latent effects persisted even after accounting for measurable defensive skills. Together, these results highlight the value of incorporating latent metrics alongside traditional ones to fully evaluate a catcher's impact on the game.

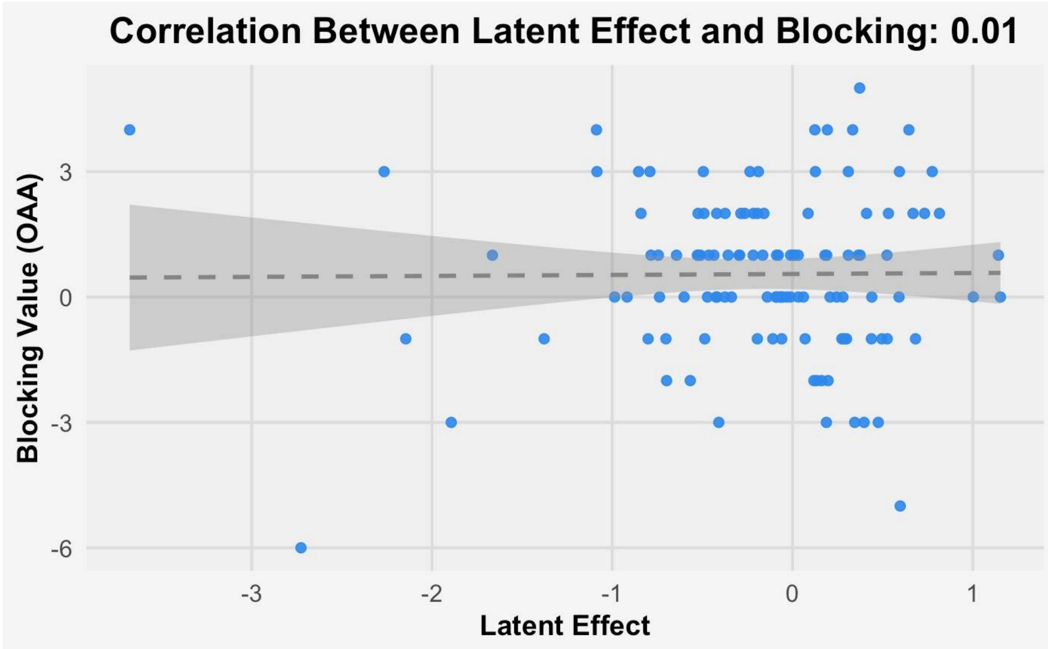

Figure 20. Correlation between Latent Effect and Blocking Value (OAA)

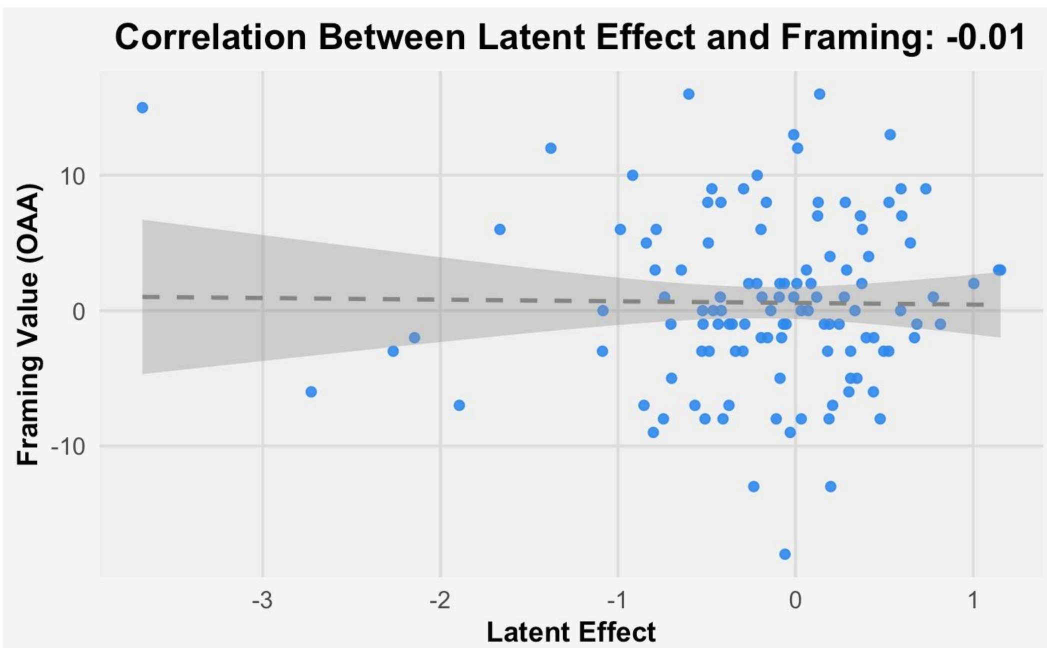

Figure 21. Correlation between Latent Effect and Framing Value (OAA)

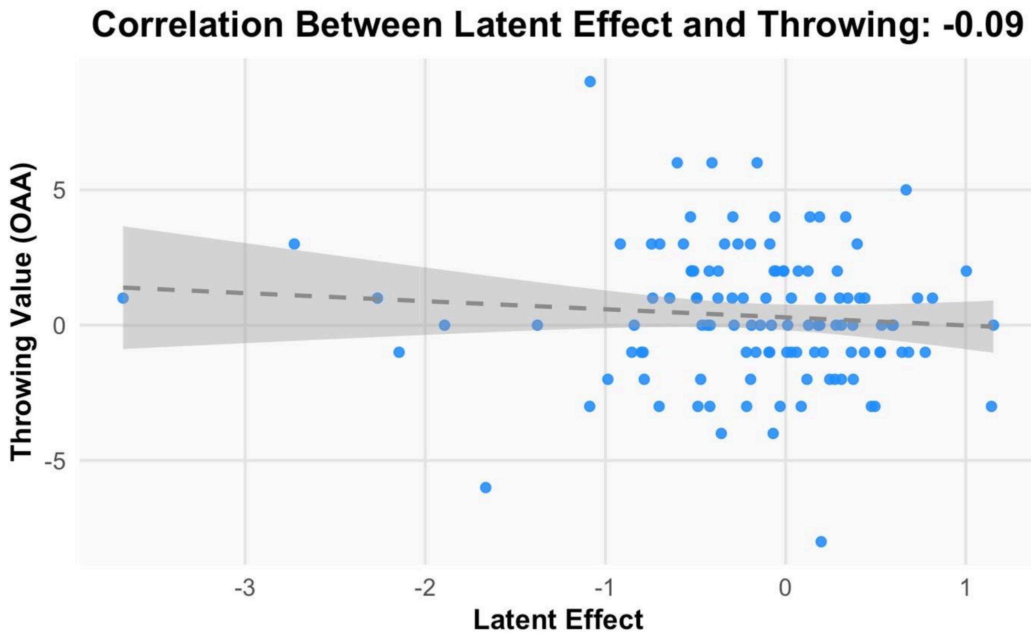

Figure 22. Correlation between Latent Effect and Throwing Value (OAA)

Conclusion

I. ***Takeaway 1:*** Catchers have some sort of latent effect on strikeouts.

Residuals in our Logistic Regression

Our year-over-year residuals showed consistent catcher ability in getting strikeouts, even when accounting for framing, and the abilities of pitchers and batters. This is important because it points to the idea that these catchers have a consistent latent effect.

II. ***Takeaway 2:*** Catchers contribute to lowering pitcher's xFIP- outside of just framing.

Consistently Good Catchers

Kyle Higashioka was for years Gerrit Cole's personal catcher, and he stands out in our xFIP- mixed-effects model for his consistent placement among the top catchers, sitting in the top 5 every season. Cole, one of the best pitchers in the league, had Higashioka catch for him most of the time they were both on the Yankees, potentially reflecting the importance of his latent traits, whether that be their comfortability or his game-calling.

Team Switching

Our xFIP- model suggests that teams may influence game-calling, either directly through someone in the dugout making pitch calls or indirectly via effective pre-game planning. This is exemplified by Martin Maldonado, who consistently performed above average in Model 2.2 and even ranked in the top 5 in 2022. However, his move to the Chicago White Sox in 2024 coincided with a significant decline in performance of his latent traits, as he ranked third worst in our metric that year. While the White Sox struggled in 2024, this result underscores the potential impact of joining an organization with a different approach to game planning and analytics. It's also possible that Maldonado's latent comfort and synergy with pitchers were disrupted by the transition. To better account for such team-level influences, we introduced Team as a factor in Model 3, allowing us to capture the variability introduced by organizations with differing analytics capabilities and game-planning strategies. This underscores the difficulty of capturing team effects, especially as they may change year-to-year; a team with an analytical advantage in 2023 may not have that same advantage in 2024.

Improving Analytics

Another significant factor in Model 2 was the possibility that analytics teams were improving and that a player could improve within his organization. Travis d'Arnaud, who had been with the

Atlanta Braves throughout the four seasons studied in this paper, likely benefitted from the Braves' improving analytics department. This stability allowed him to build familiarity with the pitching staff and enhance his performance. So, his improved performance in Model 2's latent trait metric could point to the idea of his game-calling improving with better analytics or his increased comfortability with a particular pitching staff.

III. ***Takeaway 3:*** Teams have a latent effect on xFIP, which could be due to differing perspectives on analytics or overall capabilities.

Highly-regarded teams like the Rays and Dodgers may rank low in xFIP- mixed effects because they prioritize other metrics, such as weak contact or defensive positioning, which don't directly impact xFIP-. Additionally, their reliance on high-strikeout pitchers and frequent rotation of pitching staffs and catchers might dilute the measurable impact of catcher-team synergy in this context. These results highlight the complexity of organizational strategies and suggest that advanced analytics may prioritize broader team outcomes over metrics like xFIP-.

IV. ***Takeaway 4:*** xFIP- and RA- are measuring two markedly different things.

In the discrepancies between Model 3 and Model 4, we point to the idea that teams and catchers may be prioritizing different objectives. While some teams might focus on xFIP- to emphasize preventing walks and home runs, others may adopt strategies that optimize for outcomes not fully captured by xFIP-, such as inducing weak contact or ground balls. The team rankings in the Runs Allowed Minus mixed-effects model, while designed to reflect analytics and game-planning ability, were likely influenced by external factors, including ballpark characteristics and overall run environments.

For example, teams with more pitcher-friendly ballparks could naturally suppress runs, giving the appearance of stronger game-planning or analytics contributions, even if these effects were primarily environmental. Conversely, teams in hitter-friendly parks may rank lower despite having effective game-planning strategies, as their environments inherently inflated run metrics. This intermixing of ballpark effects with team-level rankings complicates efforts to isolate and evaluate organizational strengths in analytics and game-planning. The team rankings in the model, while aiming to reflect analytics and game-planning ability, may also be heavily influenced by ballpark factors and run environments. It is worth noting that our xFIP- models are designed to minimize the influence of ballpark factors and run environments, making them a more reliable measure for evaluating the latent contributions of both catchers and teams in terms of analytics and game-planning effectiveness; using fly-ball rate instead of home-run rate in xFIP- directly minimizes the effect of ballpark size and the inconsistency of home-run rates.

V. ***Takeaway 5:*** Model Comparison

Models 3.1 and 2.2 produced very similar results, as seen in Figure 23, which is unsurprising given that both utilize xFIP- as a measure to uncover the latent traits of a catcher.

Both models aim to quantify how catchers impact pitching performance beyond the obvious contributions of the pitcher or team defense. By focusing on xFIP-, these models provide a consistent framework for evaluating a catcher's effect on pitching outcomes, resulting in aligned results that reinforce the idea that latent traits, like framing ability and synergy with pitchers, play a critical role in a catcher's value. The similarity in their findings highlights the robustness of xFIP- as a tool for isolating the repeatable, skill-based impact of catchers on pitcher performance.

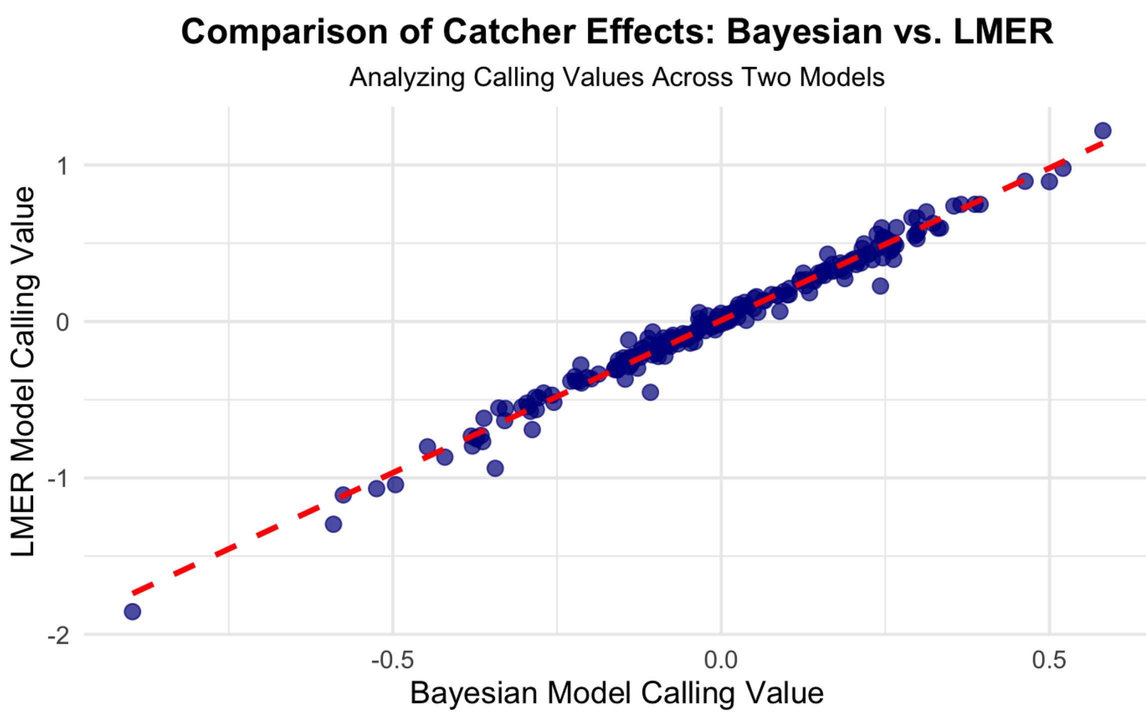

Figure 23: Comparison of Model 3.1 and Model 2.2

Models 4.1 and 2.2, both mixed-effects models, show limited correlation in Figure 24 because they predict fundamentally different outcomes: Model 4.1 focuses on RA-, while Model 2.2

predicts xFIP-. Runs Allowed Minus incorporates a wide range of factors, including defense, sequencing, and luck, making it a broader and less controlled measure of performance. In contrast, xFIP- isolates a pitcher's direct contributions by focusing on strikeouts, walks, and home runs, while neutralizing external factors like defense and ballpark effects.

We consider Model 2.2 to be superior for evaluating catcher influence because xFIP- is a more controlled statistic. By removing the variance introduced by environmental and defensive factors, xFIP- provides a clearer signal of how a catcher impacts outcomes like pitch framing and game-calling. This controlled framework allows for a more accurate assessment of a catcher's latent traits, reducing the noise inherent in run-based metrics like RA-.

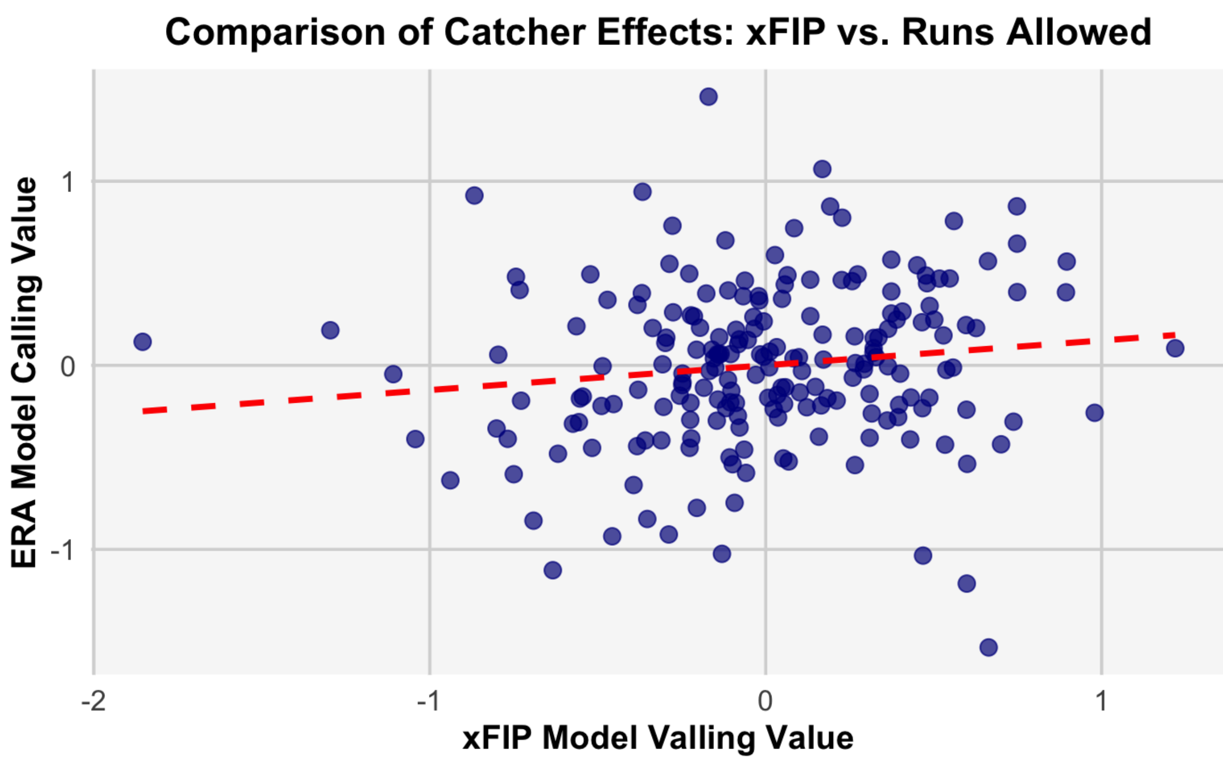

Figure 24: Comparison of Model 4.1 and Model 2.2

VI. Takeaway 6: There is a need to adapt models to account for evolving gameplay dynamics.

The low correlation observed in 2022 and 2023 residuals for our Logistic Regression model (Figure 12) may have been influenced by external changes introduced after the 2022 season.

Major League Baseball implemented several rule changes, including the introduction of a pitch clock, which forced pitchers and catchers to reduce their time between pitches, and PitchCom, a technology enabling pitch calling without traditional hand signals (MLB, 2023). These changes could have disrupted established rhythms between pitchers and catchers, potentially reducing the stability of latent effects and contributing to the observed variability.

VII. Takeaway 7: Latent catcher effects reflect unmeasured dimensions of performance.

Our analysis revealed that the latent catcher effect derived from Model 3.1 is distinct from traditional Statcast defensive metrics such as framing, blocking, and throwing. The weak correlations between these metrics—values of -0.01, -0.09, and 0.01, respectively—indicate that the latent trait metric captures contributions beyond the scope of measurable defensive skills. This suggests that the latent effects represent unique dimensions of catcher performance, such as game-calling, pitcher-catcher synergy, or situational decision-making, which are not encapsulated in existing metrics.

Interestingly, even the inter-metric relationships among Statcast's framing, blocking, and throwing metrics were weak, except for a modest correlation of 0.263 between framing and blocking. This finding challenges the assumption that elite defensive catchers excel uniformly across all skill sets. Instead, it underscores the specialized nature of catcher performance, where different skills may serve distinct roles rather than functioning as a unified defensive package.

Discussion

The findings from our study shed light on the nuanced and often hidden contributions of catchers in Major League Baseball, moving beyond traditional metrics to explore latent traits such as game-calling and pitcher-catcher comfortability. While contemporary tools like Statcast have revolutionized the evaluation of defensive metrics such as framing, blocking, and throwing, our research demonstrates the importance of incorporating unmeasured attributes to fully understand the value of a catcher.

Latent Effects on Strikeouts

Our logistic regression analysis revealed year-over-year correlations in residuals that likely cannot be explained by measurable factors like framing or batter/pitcher characteristics alone. This consistency suggests the presence of intangible skills that enable certain catchers to have a persistent impact on strikeouts. These latent traits highlight the importance of incorporating qualitative aspects of a catcher's role into quantitative models to better capture their contributions.

Catchers' Impact on xFIP-

Using mixed-effects modeling, we demonstrated that catchers influence a pitcher's xFIP- beyond just framing and measurable skills. Consistent appearances by catchers like Kyle Higashioka at the top of the rankings emphasize the importance of chemistry and game-calling. Higashioka's steady performance alongside elite pitcher Gerrit Cole illustrates the significance of familiarity, while Travis d'Arnaud's improvement aligns with the Atlanta Braves' commitment to advanced analytics and synergy with his pitching staff. Conversely, Martín Maldonado's decline after joining the struggling Chicago White Sox illustrates how team context and a disrupted synergy with a pitching staff can hinder a catcher's latent effectiveness.

Organizational and Team Effects

Our Bayesian hierarchical model expanded on these insights by incorporating team-level effects, revealing intriguing trends. Teams traditionally recognized for their analytical prowess, such as the Dodgers and Rays, ranked lower than expected in xFIP-related metrics. This could reflect differing organizational priorities, such as an emphasis on weak contact or defensive positioning, rather than metrics like framing or game-calling. The model underscored the complexity of team-catcher dynamics, where broader strategies may overshadow individual contributions.

Runs Allowed and the Broader Picture

Modeling Runs Allowed Minus (RA-) presented unique challenges because RA- captures a broader set of influences than xFIP-. While RA- directly measures a team's success in preventing runs, it incorporates not only a pitcher's and catcher's contributions but also external factors like team defense, sequencing, and even luck. This makes RA- a less controlled measure and complicates efforts to isolate the specific effects of individual catchers. For instance, teams with strong infield defense or advantageous ballpark environments may suppress runs more effectively, which could inflate a catcher's perceived contributions. Conversely, teams with weaker defenses or in hitter-friendly parks may see inflated RA- values, obscuring a catcher's potential impact. These complexities highlight the trade-offs inherent in using RA- to evaluate catcher performance.

Challenges and Implications

While our models provide compelling evidence of latent catcher contributions, they also highlight the challenges in isolating these effects. Organizational strategies, player development, and evolving analytics departments introduce variability that can blur the lines between individual and team-level impacts. For example, the improvement seen in d'Arnaud's performance could result from better analytics or enhanced comfortability with the pitching staff, making it difficult to separate these traits.

Broader Implications for Evaluation and Strategy

The results of our study have meaningful implications for how teams evaluate and compensate catchers. Quantifying latent traits like game-calling and comfortability can lead to better-informed decisions in player development, contract negotiations, and roster construction. Additionally, understanding the influence of team-level factors emphasizes the need for organizations to invest in analytics and game-planning to maximize catcher contributions.

Our research highlights the complexity and importance of latent traits in evaluating catchers. While traditional metrics provide a solid foundation, incorporating these unmeasured attributes offers a more complete picture of a catcher's value. Future work should continue to refine these models, extend the dataset, and explore how organizational dynamics and team strategies interact with individual performance.

Future Work

In the future, we'd like to continue this project over more years of data and by teasing out more specifics of the outputs of our models. While we have accumulated a good amount of evidence that catchers exert some sort of hidden influence on run prevention, our models are not sufficient for teasing out exactly where these effects originate or for attributing run values to these qualities.

Acknowledgements

We would like to thank our professors Dr. Tobias Moskowitz and Dr. Nils Rudi for giving us the opportunity to conduct this research project. We would also like to acknowledge https://baseballsavant.mlb.com/ for providing us with the data necessary to conduct our analysis.

References

1. Biolsi, C., Goff, B., & Wilson, D. (2021). Task-level match effects and worker productivity: Evidence from pitchers and catchers. *Applied Economics*, *54*(25), 2888-2899. https://doi.org/10.1080/00036846.2021.1999898
2. Castrovince, A. (2023, March 28). 8 changes fans need to know for the '23 season. MLB.com. https://www.mlb.com/news/mlb-rules-changes-refresher-2023-season/
3. *Catcher framing*. (2024). MLB.com. https://www.mlb.com/glossary/statcast/catcher-framing
4. DeForest, W. (2023). *Modeling the Performance-based Compensation of MLB Catchers* [Master's thesis]. https://scholarship.claremont.edu/cmc_theses/3155/
5. *Fielding run value*. (2024). MLB.com. https://www.mlb.com/glossary/statcast/fielding-run-value
6. Garro, A., & Chesterton, E. (2019, February 11). *Which is the more demanding job: Pitcher or catcher?* MLB.com. https://www.mlb.com/cut4/which-is-harder-being-a-catcher-or-a-pitcher-c303719102
7. Hanrahan, T. (1999). Catchers--better as veterans. *By the Numbers*, *9*(3), 13-20. https://sabr.org/research/statistical-analysis-research-committee-newsletters/
8. Hanrahan, T. (2009). Catcher ERA — Once more with feeling: It’s tough to be a rookie. *Baseball Research Journal*. https://sabr.org/journal/article/71351/
9. In, S. (2021, April 22). *The catcher who can't catch as well as we might think*. Pitcher List. https://pitcherlist.com/the-catcher-who-cant-catch-as-well-as-we-might-think/
10. Lindbergh, B. (2014, July 14). *Yadi’s secret sauce*. Grantland. https://grantland.com/features/st-louis-cardinals-yadier-molina-catcher-value-game-calling/
11. Loughin, T. M., & Bargen, J. L. (2008). Assessing pitcher and catcher influences on base stealing in Major League Baseball. *Journal of Sports Sciences*, *26*(1), 15-20. https://doi.org/10.1080/02640410701287255
12. Nakahara, H., Takeda, K., & Fujii, K. (2023). Pitching strategy evaluation via stratified analysis using propensity score. *Journal of Quantitative Analysis in Sports*, *19*(2), 91-102. https://doi.org/10.1515/jqas-2021-0060
13. *Pop time (POP)*. (2024). MLB.com. https://www.mlb.com/glossary/statcast/pop-time
14. Rosciam, C. (2009). *Catcher’s Defensive Rating System*. SABR. https://retrosheet.org/Research/RosciamC/Catcher%20Defensive%20Rating%20System.pdf
15. Slowinski, P. (2024, December 9). *Catcher defense*. Sabermetrics Library. https://library.fangraphs.com/defense/catcher-defense/
16. Smith, S. (2015, March 25). *Do catchers have an ERA?* The Hardball Times. https://tht.fangraphs.com/classic-tht-annual/do-catchers-have-an-era/
17. *The special insight the Yankees had into new catching whiz*. (2022, April 7). nj. https://www.nj.com/yankees/2022/04/the-special-insight-the-yankees-had-into-new-catching-whiz.html
18. *Statcast catcher blocking Leaderboard*. (2024). baseballsavant.com. https://baseballsavant.mlb.com/leaderboard/catcher-blocking?players=643376-2024-111&selected_idx=0
19. *Statcast catcher framing Leaderboard*. (2024). baseballsavant.com. https://baseballsavant.mlb.com/catcher_framing?year=2024&team=&min=q&type=catcher&sort=4,1

20. *Statcast catcher throwing Leaderboard*. (2024). baseballsavant.com. https://baseballsavant.mlb.com/leaderboard/catcher-throwing
21. Tango, T. (2019, March 26). *WOWY Framing, part 3 of N: run value of a called strike*. Tangotiger blog. https://tangotiger.com/index.php/site/wowy-framing-part-3-of-n-run-value-of-a-called-strike
22. Weinberg, N. (2014, November 14). *What do we know about catcher defense?* Sabermetrics Library. https://library.fangraphs.com/what-do-we-know-about-catcher-defense/
23. *Without and with you baseball dictionary | Baseball almanac*. (n.d.). Baseball Almanac: MLB Stats, History, Records & Research | 1876-2024. https://www.baseball-almanac.com/dictionary-term.php?term=Without+and+With+You
24. Woolner, K. (2000, January 11). *Catching up with the general: A postscript: A second look at catcher defense*. Baseball Prospectus. https://www.baseballprospectus.com/news/article/436/catching-up-with-the-general-a-postscript-a-second-look-at-catcher-defense/
25. Woolner, K. (2000, January 10). *Field general or backstop?: Evaluating the catcher's influence on pitcher performance*. Baseball Prospectus. https://www.baseballprospectus.com/news/article/432/field-general-or-backstop-evaluating-the-catchers-influence-on-pitcher-performance/#problems
26. Woolner, K. (2002, May 29). *Aim for the head: Simulating catcher's ERA*. Baseball Prospectus. https://www.baseballprospectus.com/news/article/1489/aim-for-the-head-simulating-catchers-era/