

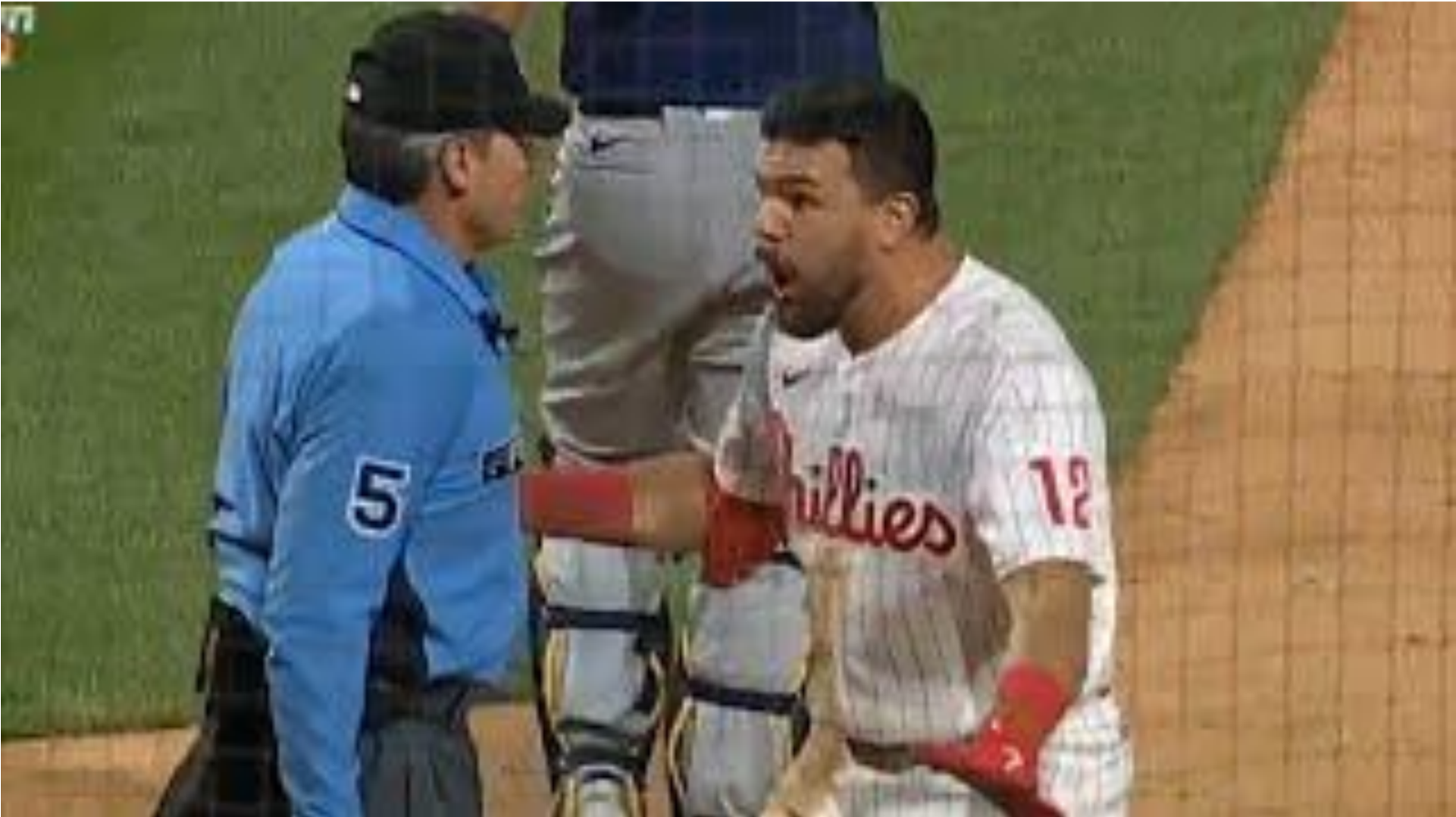


Changes In Minor League Umpire Tendencies With The Challenge and Automatic Ball-Strikes Systems

Jonah Soos, Dr. Jeremy Losak, & Dr.
Jason Maddox



April 24th 2022: Schwarber Bat Slam



April 13th 2023: Alec Bohm's Blow Up



1997 NLCS: “The Eric Gragg Game”



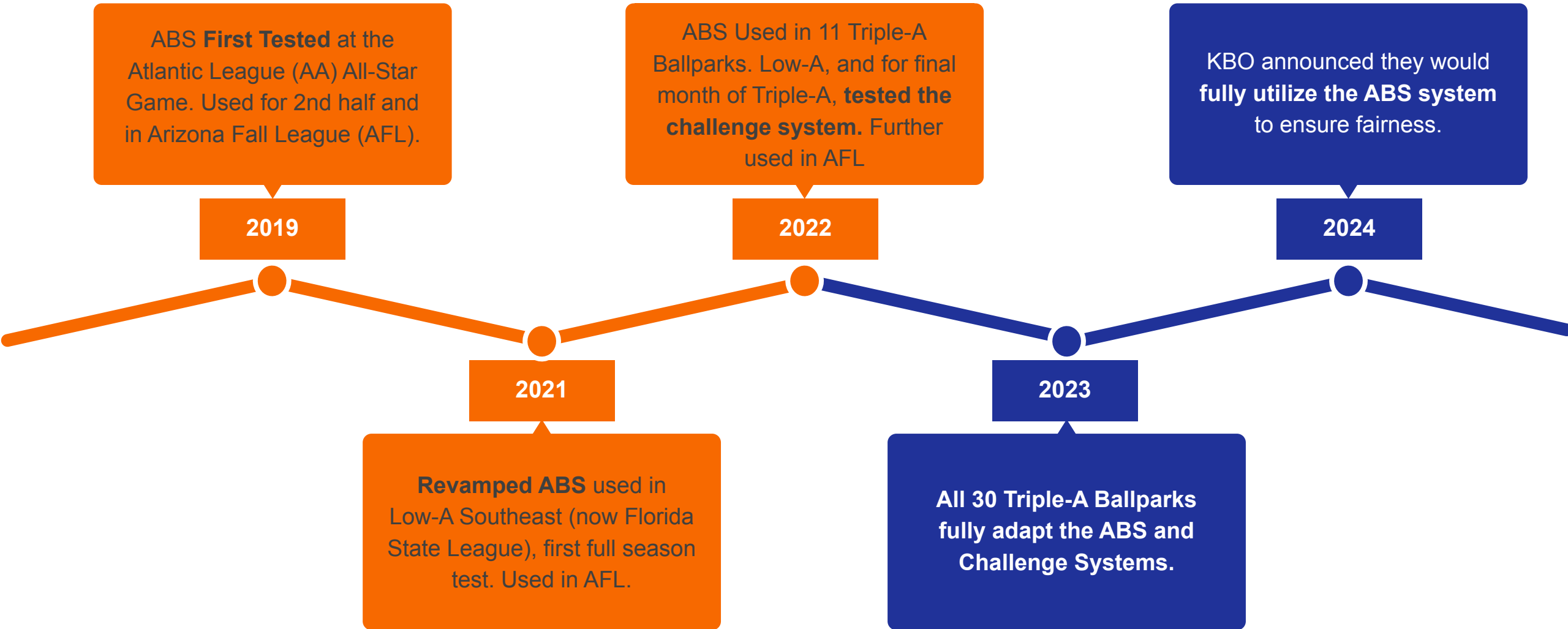
Responding to Scrutiny

Sparked discourse amongst the community about the **effect of umpires** on the game of baseball.

- Lead MLB to test two new alternatives in recent MiLB seasons
 - Automatic Ball-Strike (ABS) System
 - Challenge System

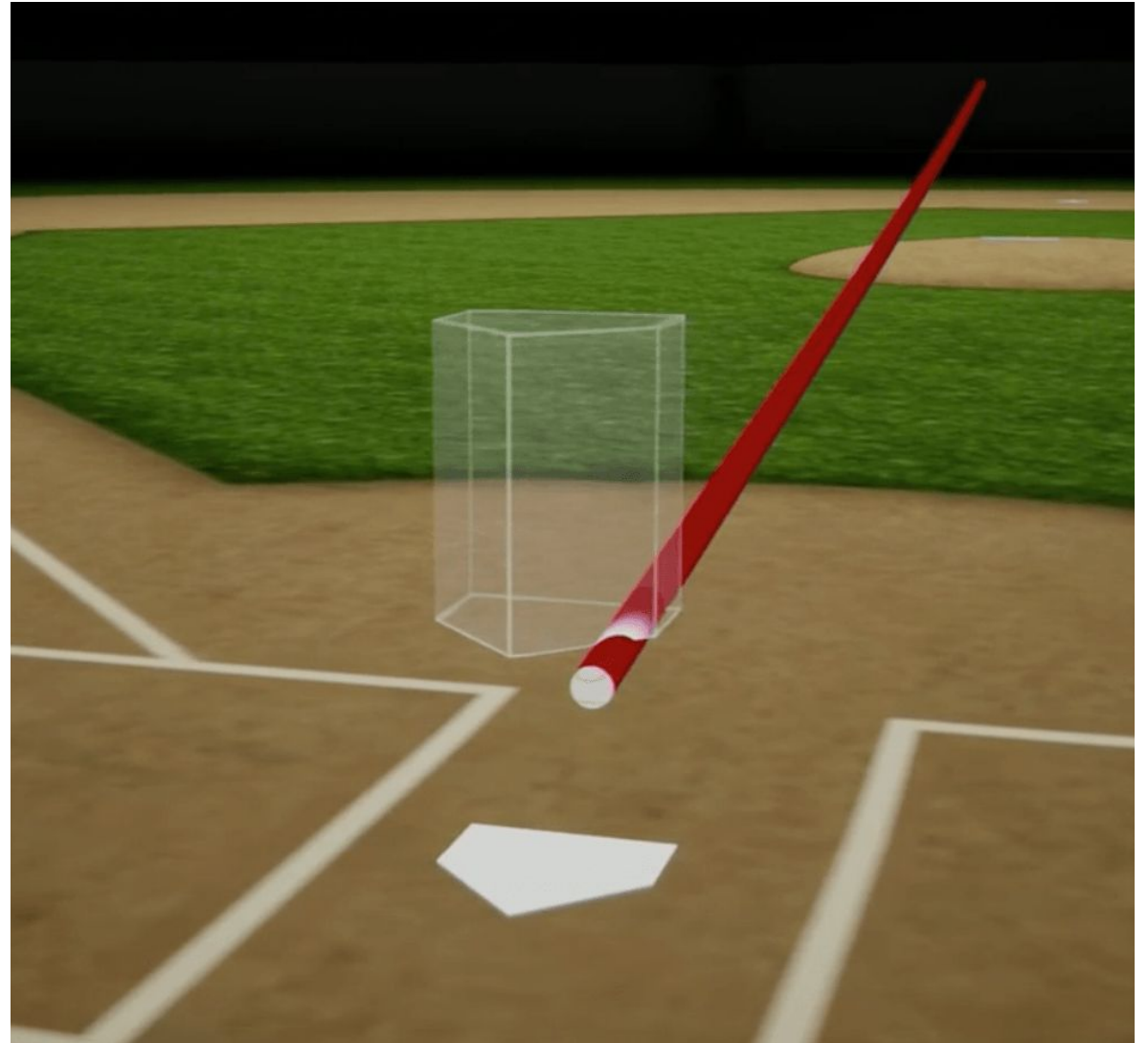


History of Automated Strike Zones



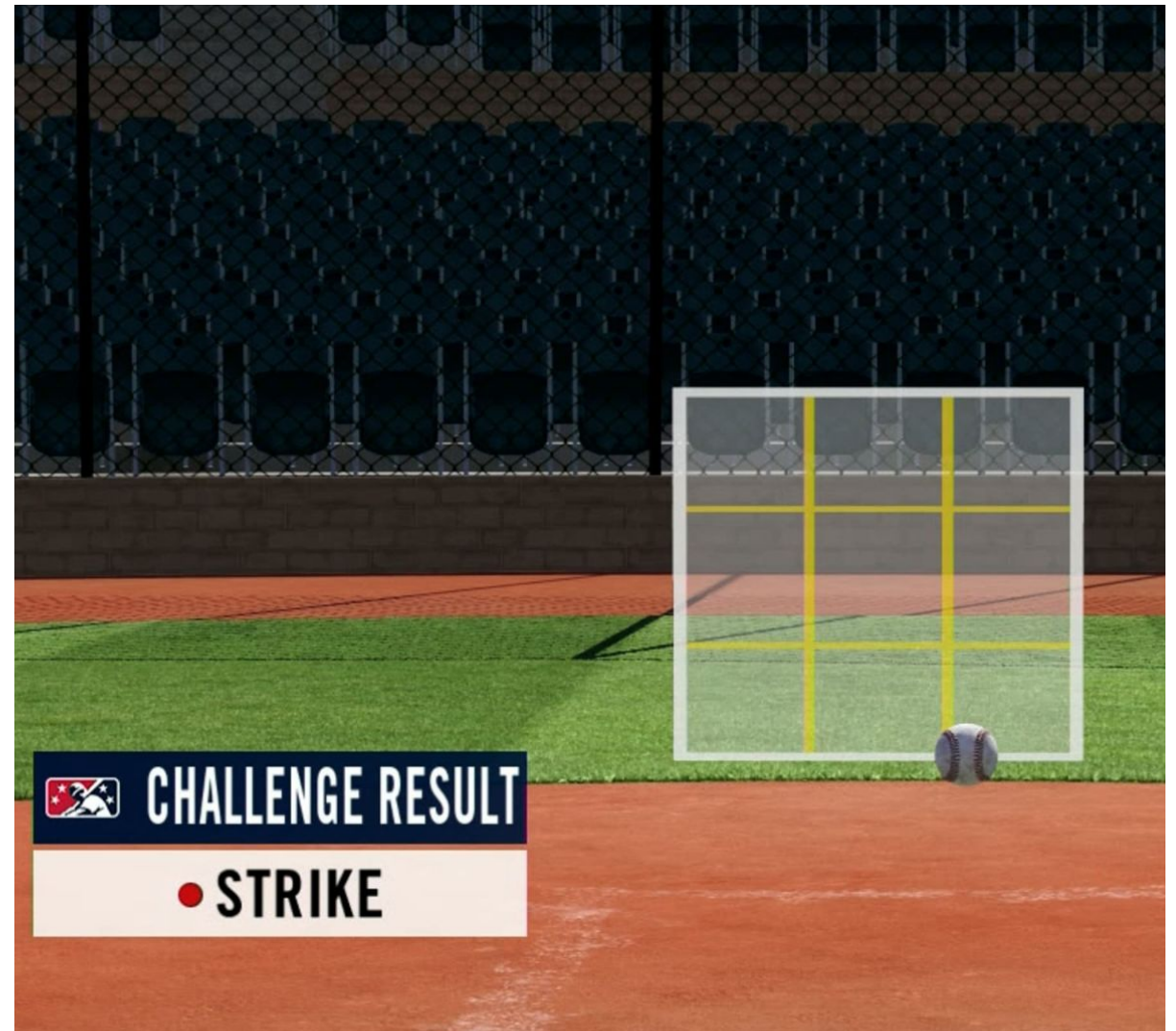
The Automatic Ball-Strike (ABS) System

- Uses Hawk-Eye 3D Tracking system to determine whether a ball touches any area of the strike zone.
- Negates any umpire variability – every call is automated
- Received criticism and multiple adjustments for being too pitcher/batter friendly



The ABS-Challenge System

- Uses same Hawkeye Technology to determine balls and strikes
- Batter, Catcher, or Pitcher can initiate a challenge after the pitch
- Teams get unlimited challenges, until they reach 3 unsuccessful attempts
- Maintains umpire variability



How a Challenge Works



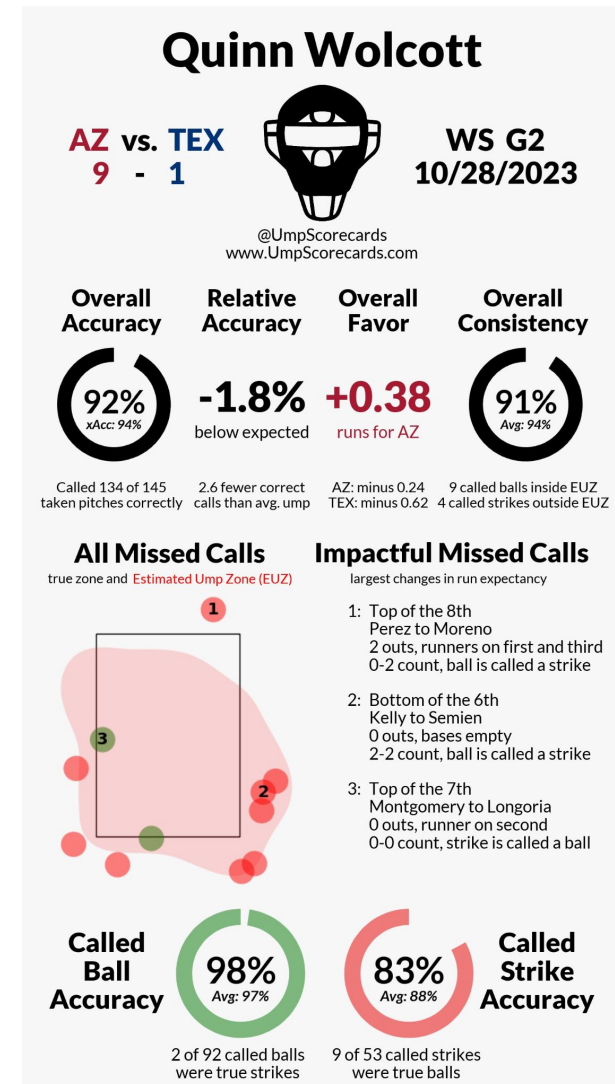
Area of Focus

- Utilized the 2023 Triple-A season as our focus group for umpire systems
- First full season to use both the ABS and Challenge Systems interchangeably
- Higher caliber of play & umpires can allow for results to be more translatable



What We Know About Umpires

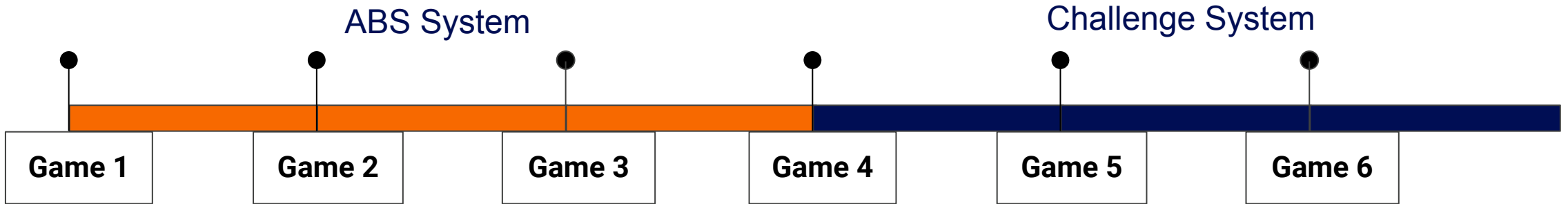
- More scrutiny of umpires with increased access to data
 - Umpire scorecard
 - Thorough review of umpire behaviors
- Umpires impacted by external social pressures
 - “Heckling” umpires (Guérette et al., 2024)
- Umpire pitch calling has positively progressed as strike-zone technology improves (Mills, 2016)
 - Expediting skill development
 - Especially prominent in younger umpires



Goal: Analyze umpire accuracy and tendency shifts under the two new systems.

Question: Do umpires change their strike calling tendencies?

Test Groups



- 2 Groups: ABS and Challenge
- In September, raised the bottom of the strike zone to account for player feedback
- Used International League Games pre April 25th as our control group.



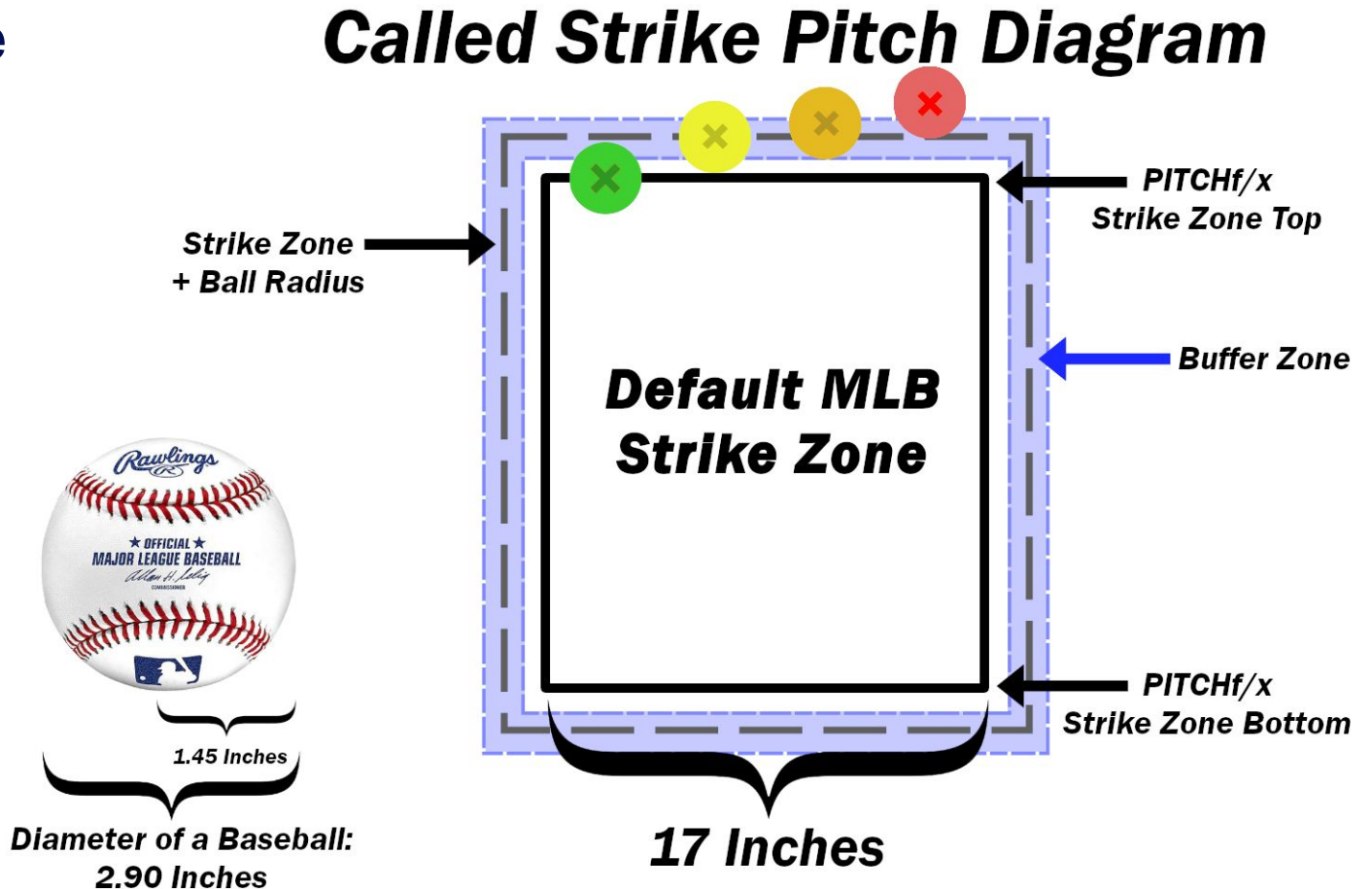
Available Data

- Collected Pitch-By-Pitch data for the 2023 Triple-A Season.
 - First season with access to PITCHf/x data
- Binary Indicator for an Incorrect Call by the Umpire
 - Determined through a predefined 2D strike zone
- Looked specifically at 3-0, 0-2, and 3-2 counts in analysis.



Defining the Strike Zone

- Classified our 2D strike zone with different parameters
- Strike Zone Edge: One ball radius from the defined MLB Strike Zone
- Buffer Zone: Half Ball Radius zone on both sides of the Strike Zone Edge



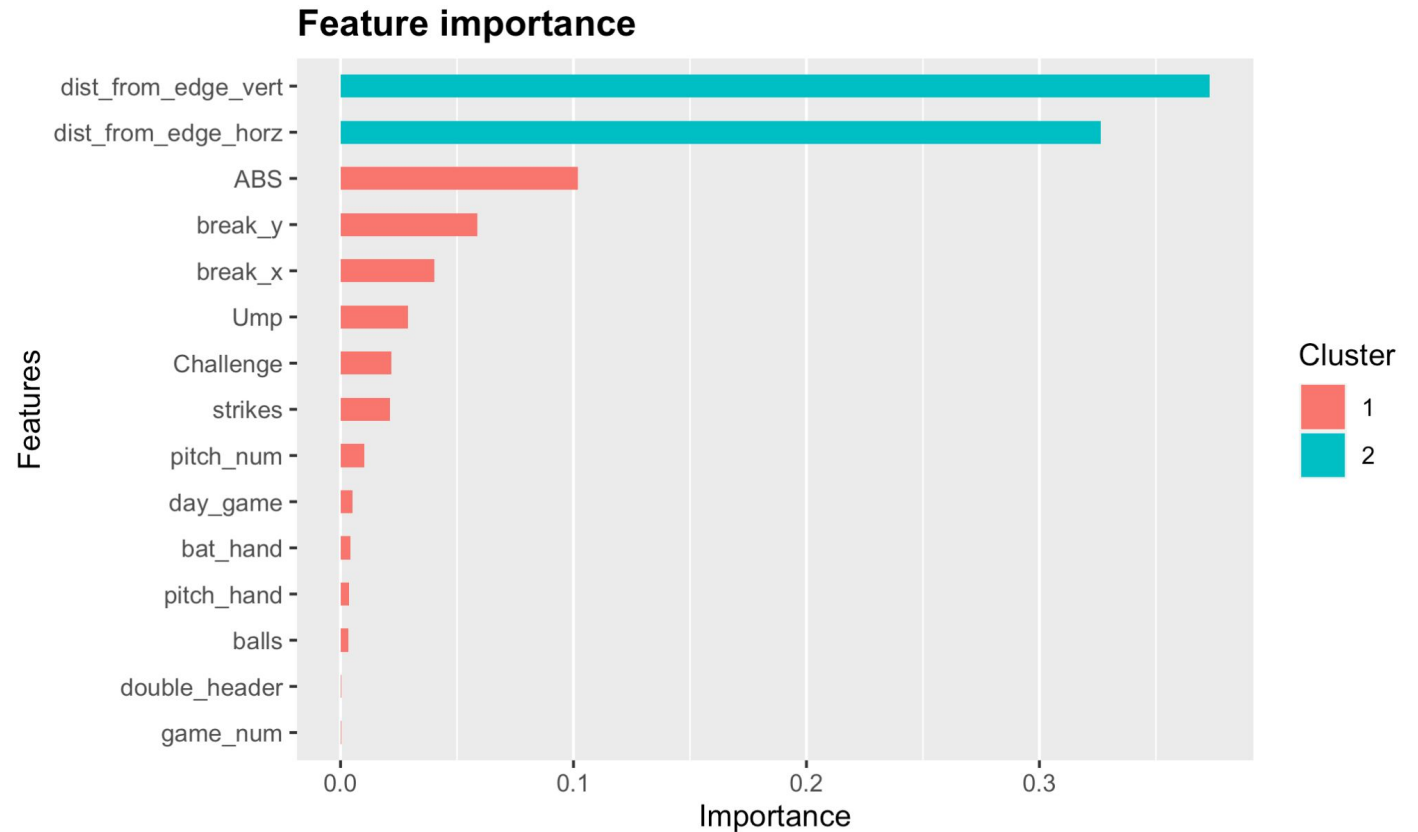
Modeling Methodology

- Limit the data to non-swing pitches.
- Model predicted probability of the umpire making the correct call
 - Effects of ball/strike system using average predictions.
- Attempted logistic regression, but too many complex interactions.



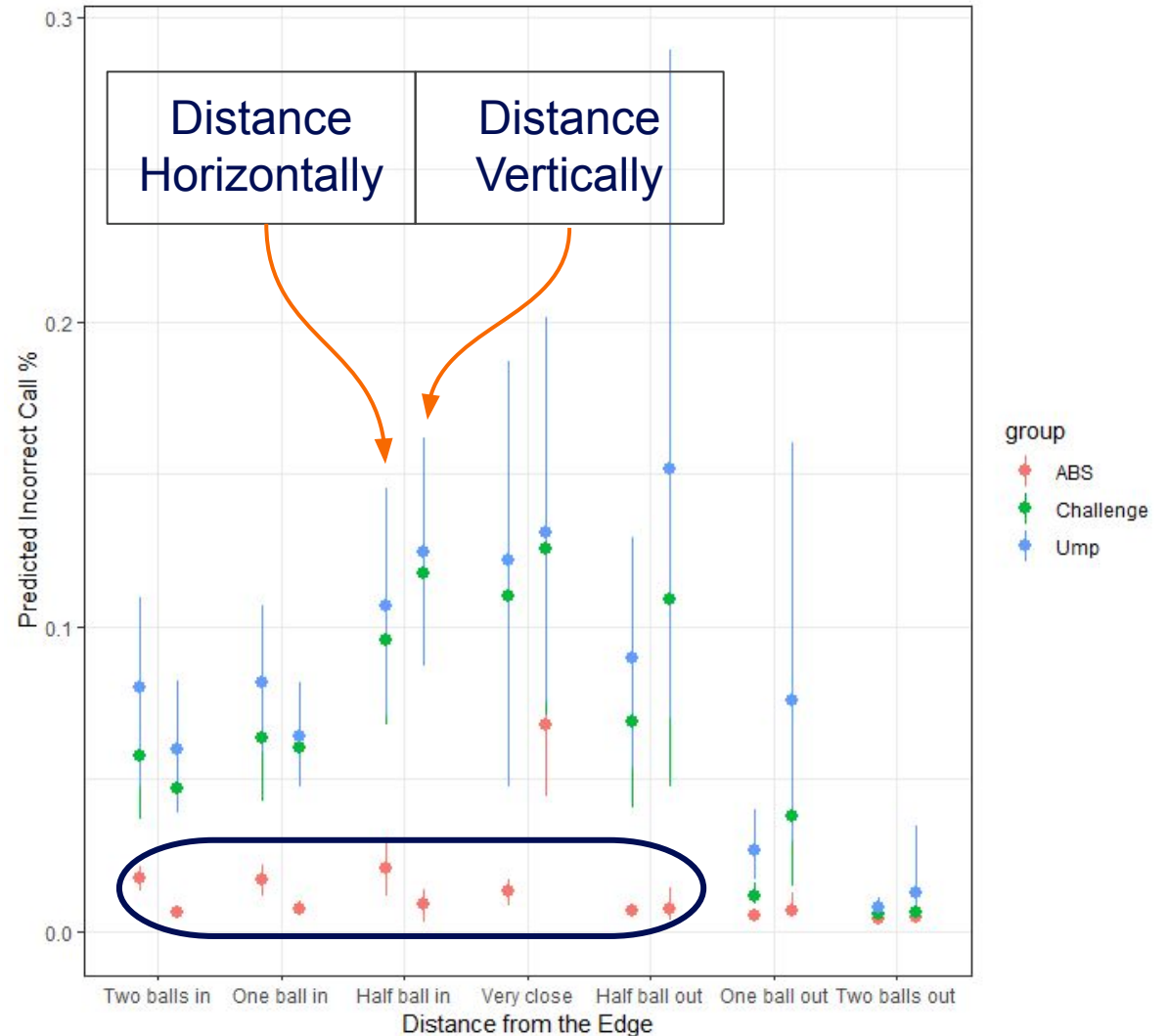
Modeling Methodology

- Fit an xgboost model to predict probability of an incorrect call by umpire.
- Tuned parameters:
 - max tree depth: 8
 - eta: 1/700
 - trees: 3360
- 100 bootstrap sample models ran to approximate standard errors of incorrect call probability.



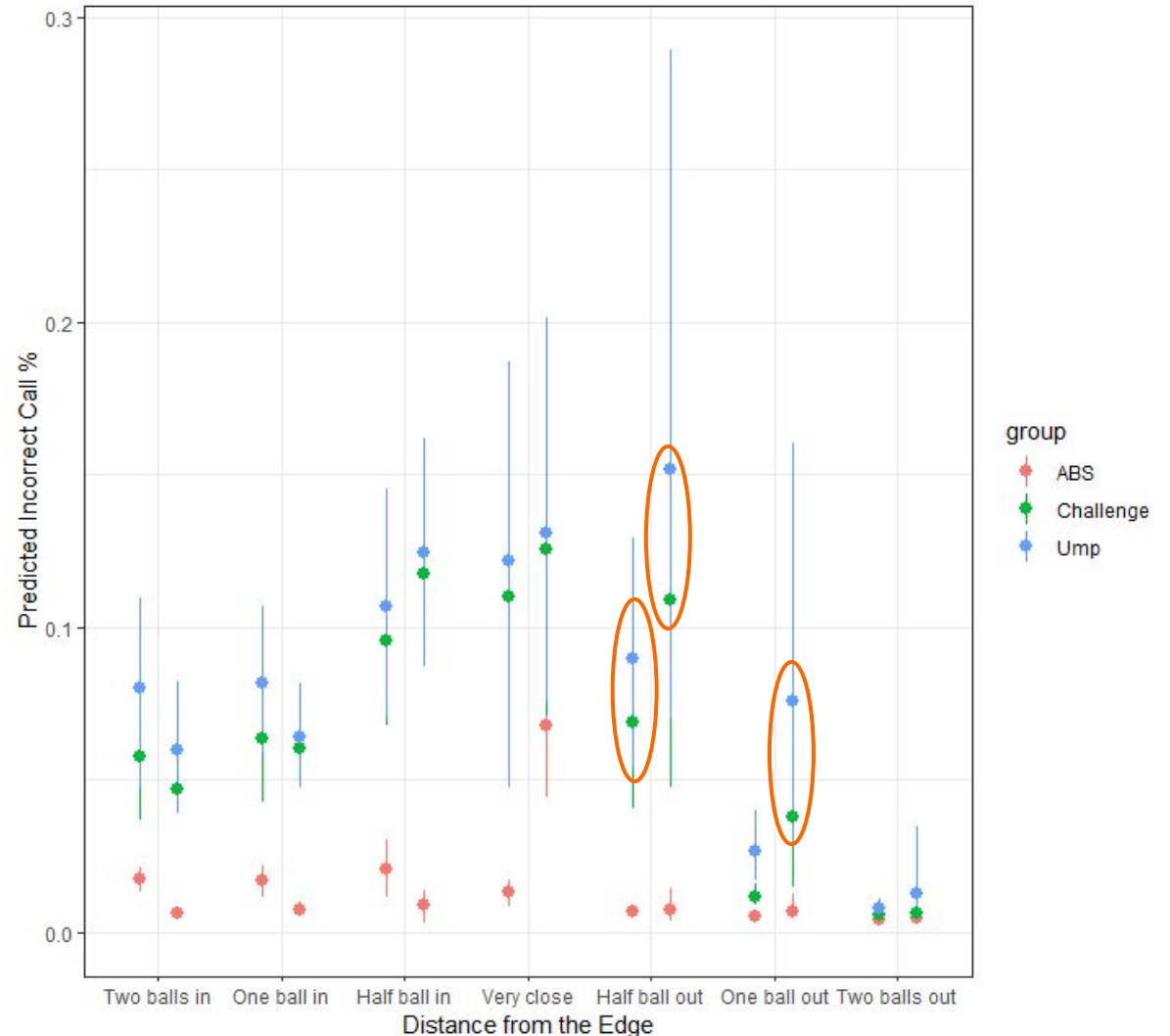
Results: ABS Accuracy

- Predicted Inaccuracy in the ABS system is nearly 0
- Should not be exactly zero – limitation of a 2D strike zone
- Most notable jump is Very Close vertically

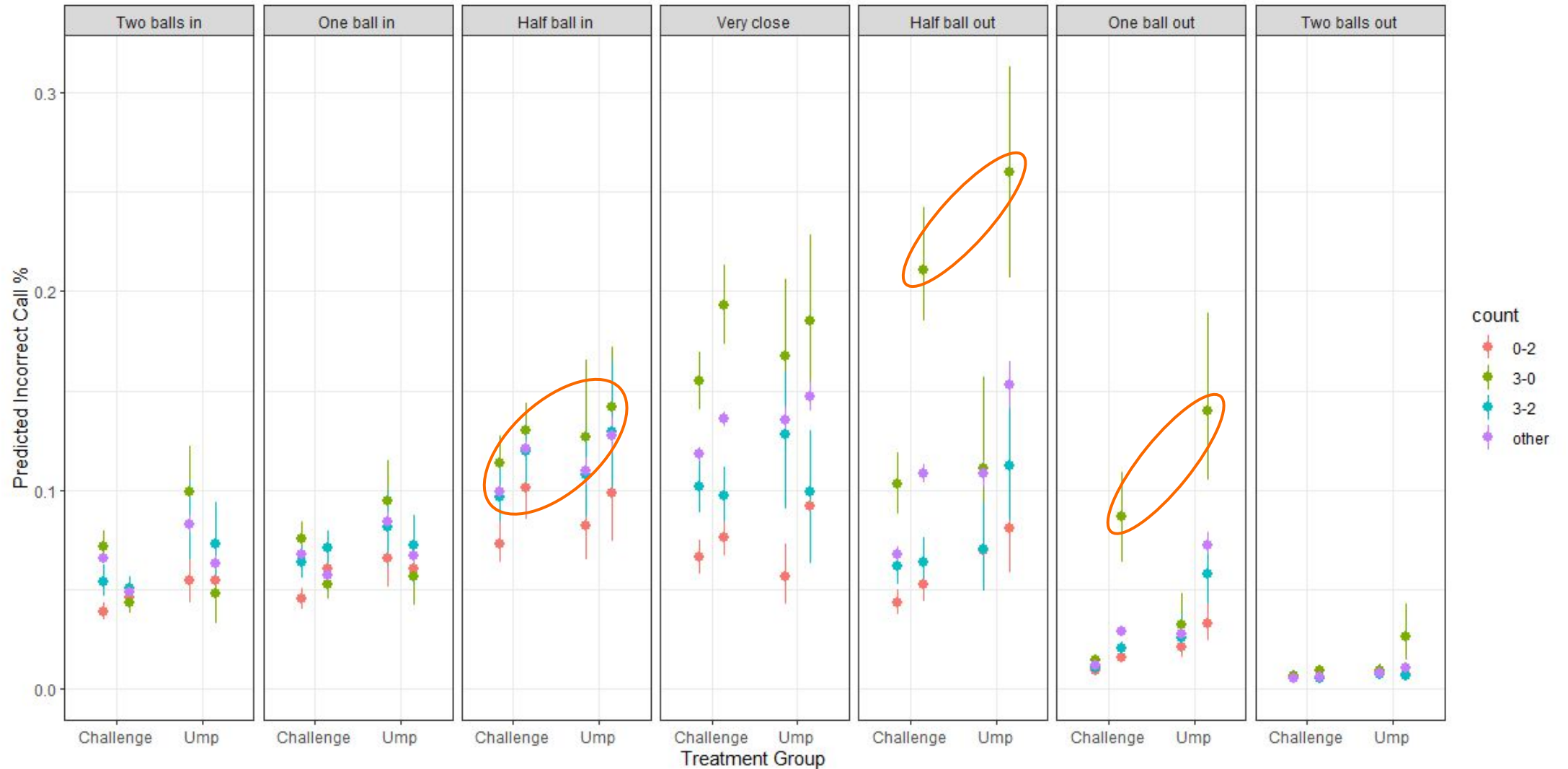


Results: General Umpire Accuracy

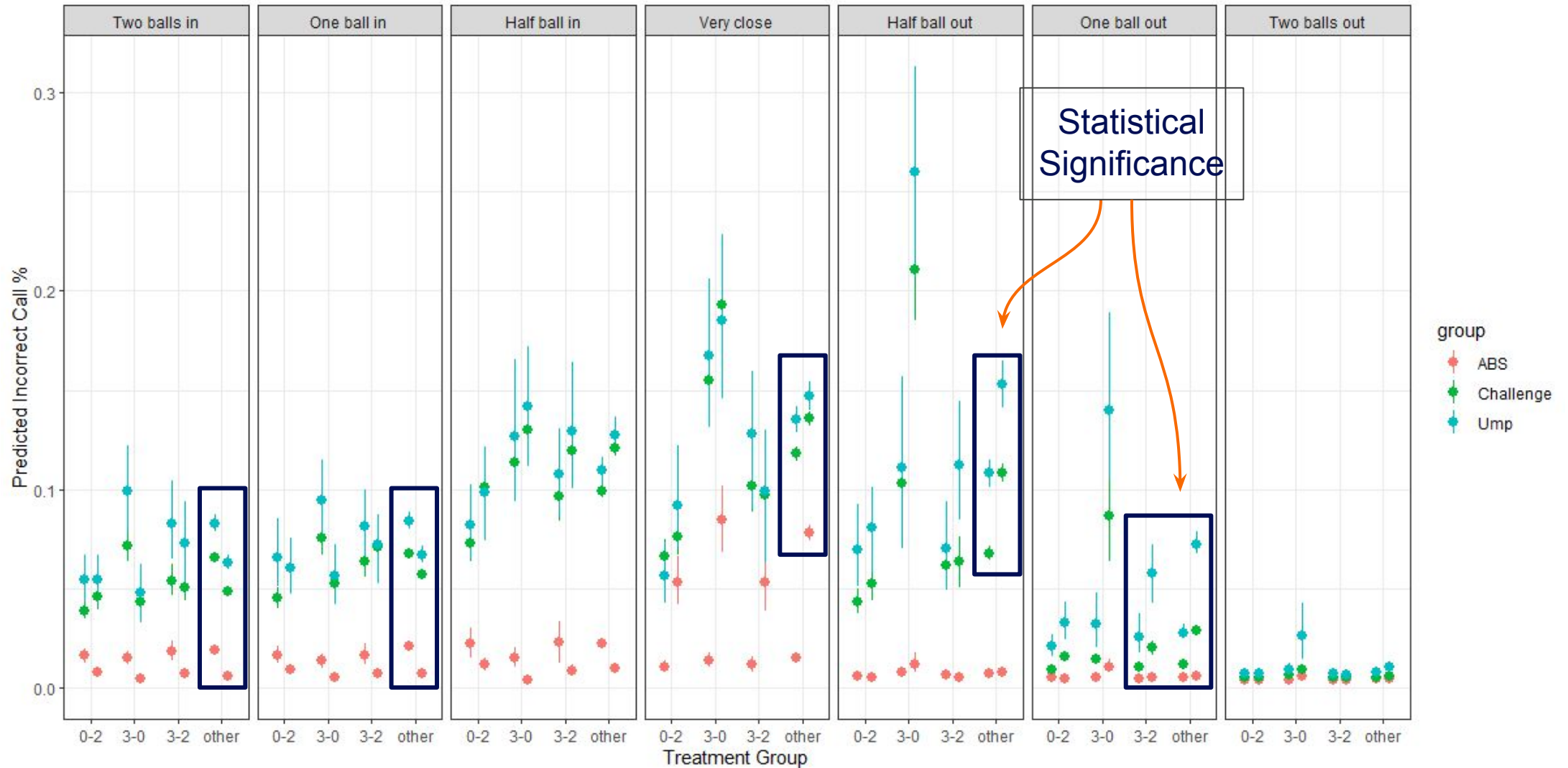
- Some evidence that Umpires are more accurate in the Challenge system over the control group.
- General trend, some instances of statistical significance
- Sample Size of control group make it difficult to draw definitive conclusions



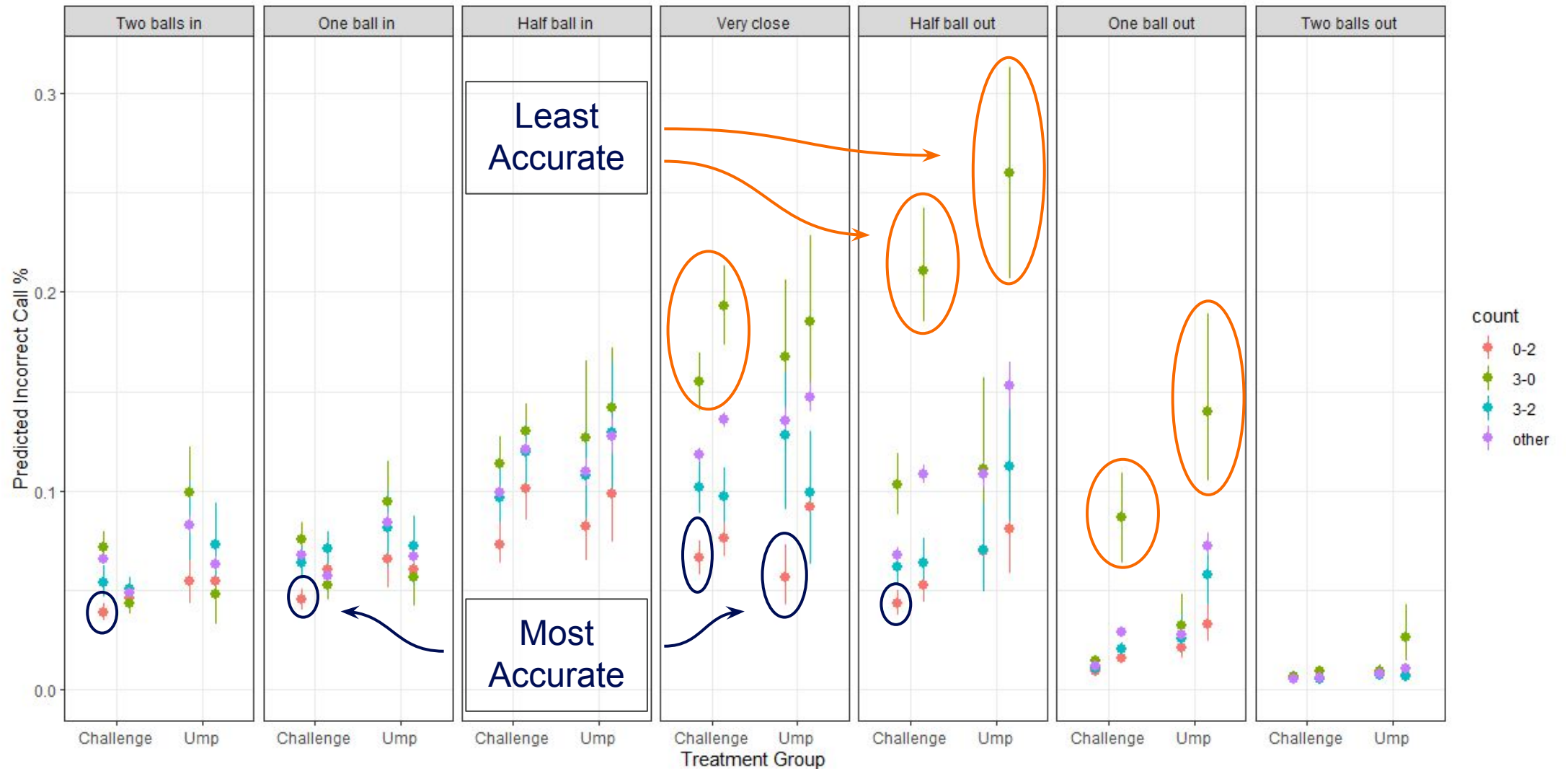
Results: General Umpire Accuracy



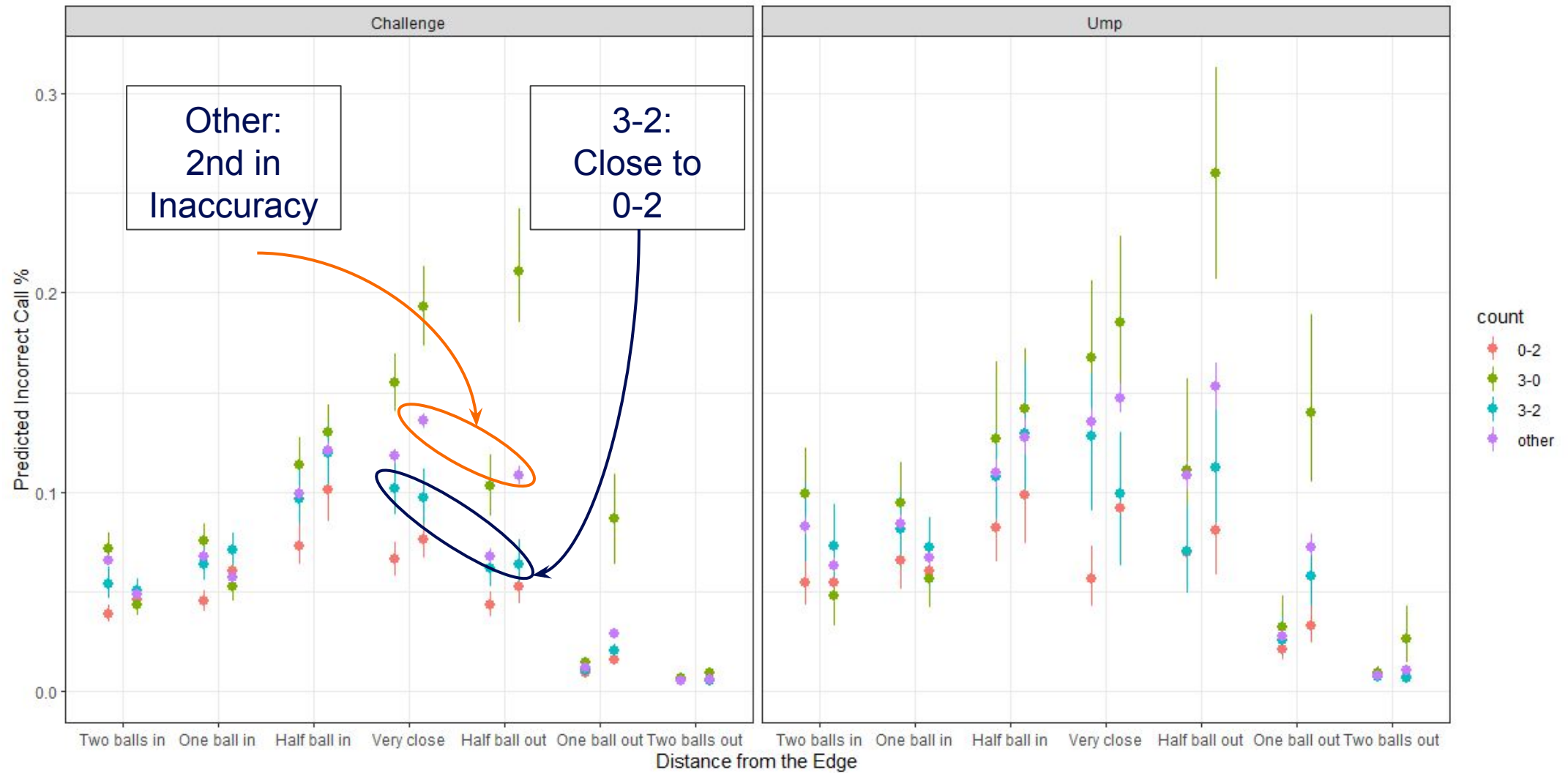
Results: General Umpire Accuracy



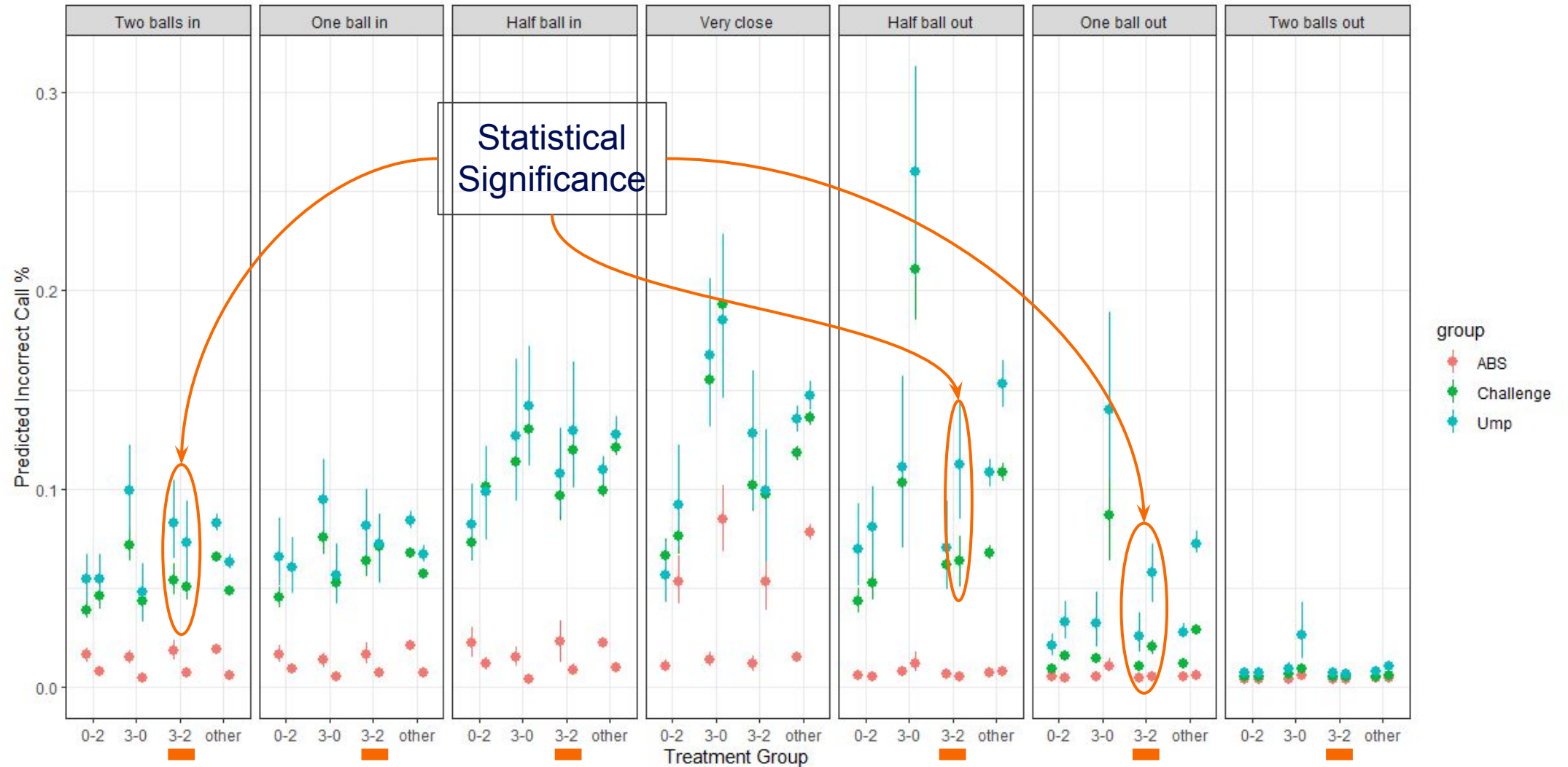
Results: Accuracy Based on Count



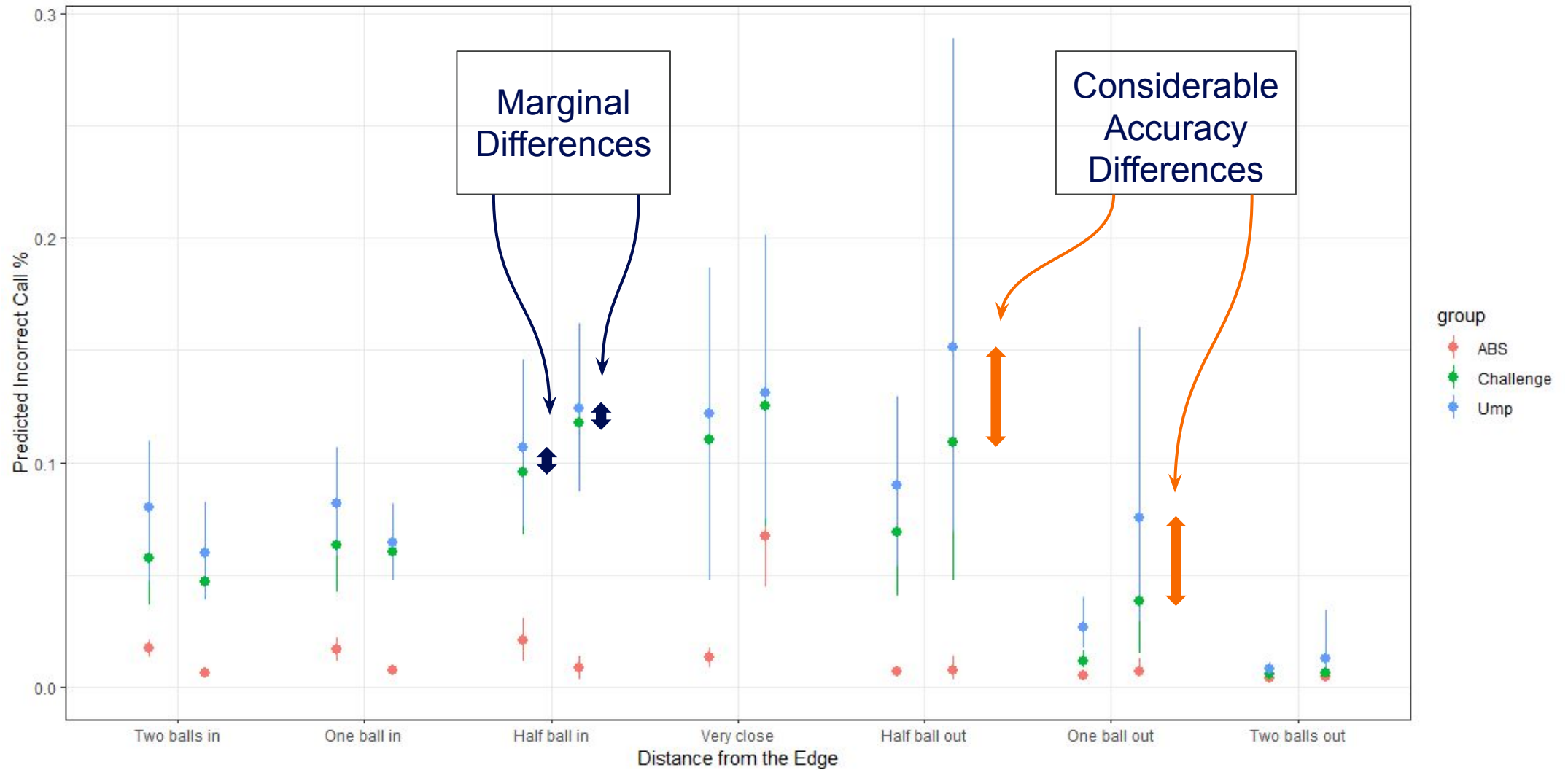
Results: Accuracy Based on Count



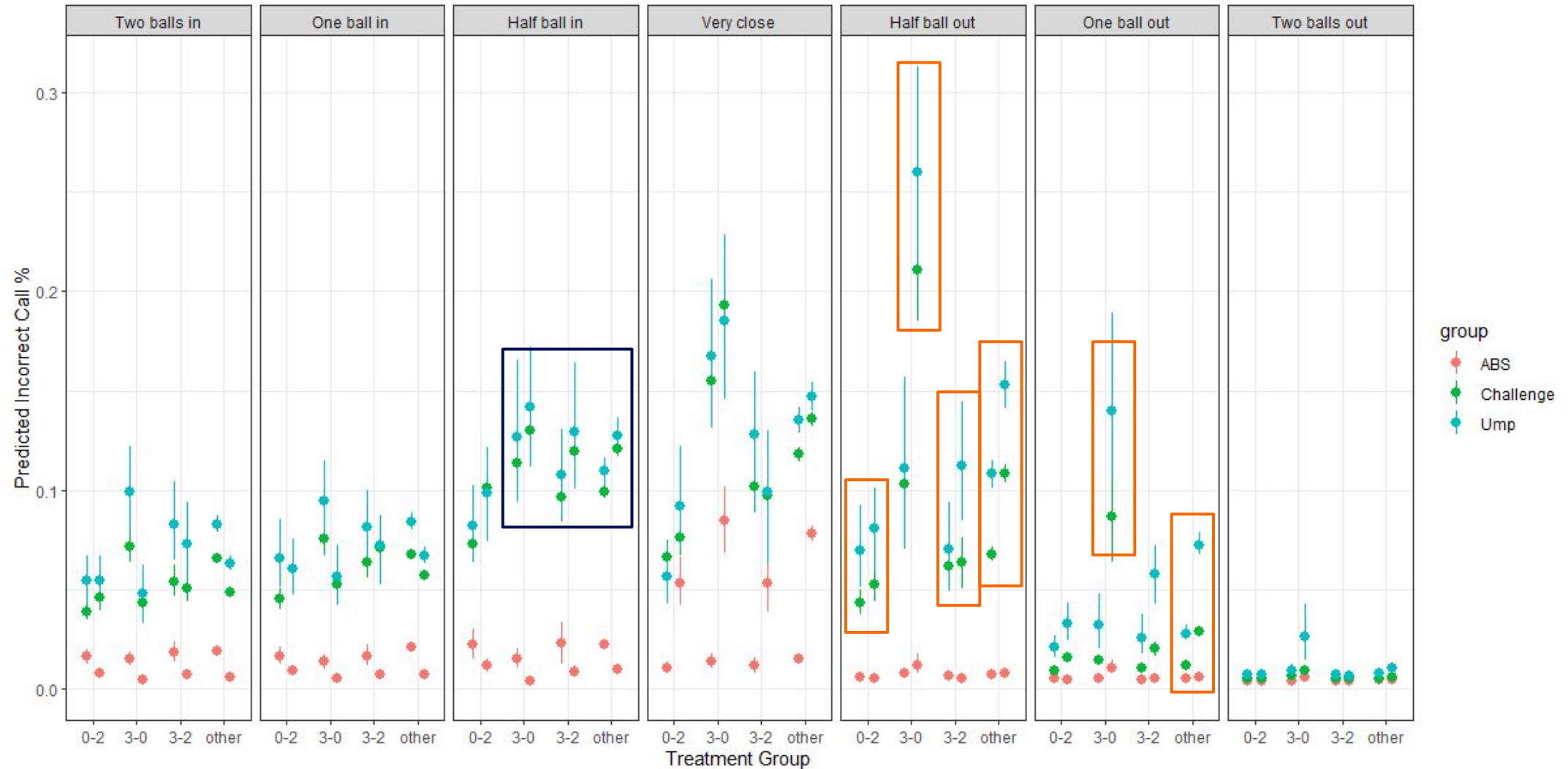
Results: Accuracy in High Leverage



Results: Shrinking the Strike-Zone



Results: Shrinking the Strike-Zone



Results: Accuracy in High Leverage

Largest Challenge %, low overturn rate

Count	Challenges	Pitches	Challenge %	Challenged Ball %	Challenged Ball Accuracy	Challenged Strike %	Challenged Strike Accuracy
3-2	494	16122	3.06%	39.47%	31.28%	60.53%	41.81%
3-1	232	8445	2.75%	52.59%	45.90%	47.41%	53.64%
3-0	113	4665	2.42%	61.95%	38.57%	38.05%	74.42%
2-1	429	17745	2.42%	46.62%	43.50%	53.38%	46.29%
2-0	278	12853	2.16%	51.08%	38.03%	48.56%	50.37%
2-2	480	25107	1.91%	44.38%	39.44%	55.63%	38.95%
1-1	577	31508	1.83%	46.10%	44.36%	53.90%	43.09%
1-0	619	34263	1.81%	52.18%	49.85%	47.82%	54.39%
0-0	1416	81164	1.74%	53.81%	51.44%	46.12%	51.76%
0-1	623	37737	1.65%	45.43%	56.54%	54.57%	45.88%
1-2	425	28521	1.49%	45.41%	43.01%	54.59%	34.91%
0-2	280	19624	1.43%	45.36%	58.27%	54.64%	37.91%

Conclusions

- Umpires trended to be more accurate under observation within the challenge system
 - Found some statistical significance
- Found evidence Umpires adapted strike zones to account for inaccuracies in 3-2 counts.
- Umpires tended to be more accurate in high leverage situations.
 - Batters and Pitchers challenged unsuccessfully





Thank You

Jonah Soos
@jonahsoos24
jasoos@syr.edu

Jeremy Losak
@JeremyLosak
jmlosak@syr.edu

Jason Maddox
@jasontmaddox
jtmaddox@syr.edu

