Umpire Accuracy, Variability, and the Automated Ball- Strike System: An Exploratory Mixed-Methods Study of Major League Baseball

Daniel Kim

Received June 17, 2025
Accepted October 23, 2025
Electronic access November 30, 2025

This study examines the variability in human umpire accuracy on game consistency in Major League Baseball (MLB) and explores stakeholder perceptions of the Automated Ball-Strike (ABS) system as a potential corrective measure. Using a mixed-methods design, I analyzed umpire performance data from 2015 to 2024 (sourced from Umpscorecards.com) and conducted a small exploratory survey (n = 16) of fans, players, and coaches. Key metrics included yearly accuracy, standard deviation across games, and Total Run Impact (totRI), which estimates the run value of missed calls. While average accuracy rose from 90.4% in 2015 to 93.9% in 2024 ($p<0.01$), extreme cases persisted: 0.9% of games showed totRI swings above +7 runs, and 0.2% fell below 80% accuracy. A small, non-representative survey (n = 16) revealed optimism about ABS but raised concerns about tradition and rhythm. Findings suggest that rare but high-leverage mistakes disproportionately shape perceptions of fairness. Given the limitations of the survey, this study does not claim to test the effectiveness of ABS, but instead offers exploratory insights into its reception. It is also proposed that MLB could consider a challenge-based hybrid ABS, balancing accuracy improvements with preservation of the sport's tradition.

Keywords: Umpire Accuracy, Automated Ball-Strike System (ABS), Major League Baseball (MLB), Sports Technology, Fairness in Officiating

Introduction

Baseball is one of the most globally followed sports, with over 500 million fans[1]. At its core lies the umpire – a symbol of fairness and authority. However, despite professional training and rigorous experience, human umpires remain prone to error. These inaccuracies are associated with game outcomes, player careers, and public perception of the sport[2]. Baseball presents unique officiating challenges, especially because the determination of strikes and balls rests solely on the home plate umpire's judgment.

Major League Baseball (MLB), the leading professional baseball league, has experienced high-profile officiating mistakes that have intensified calls for reform. For example, Armando Galarraga's near-perfect game[3] in 2010 was infamously disrupted by an incorrect call from umpire Jim Joyce. Though not related to a strike zone decision, the incident highlighted the profound consequences of human error in officiating. Accuracy in this context refers to the degree to which umpires' calls match the rule-defined strike zone, typically reviewed through internal post-game audits.

As technology transforms officiating across sports – such as VAR in soccer[4], Hawk-Eye in tennis[5], and AI tools in basketball[6] – baseball has begun exploring the Automated Ball-Strike (ABS) system. The ABS is a camera-and-algorithm-based tool designed to assist or replace human decision-making on strike calls[7]. In Korea's KBO League, ABS trials have demonstrated high pitch-tracking precision[8] but also sparked debate over its effect on game rhythm and tradition[8].

Although existing studies have explored umpire bias, environmental effects, and racial discrimination in strike calling, few studies have connected accuracy trends to actual game outcomes or assessed MLB-specific stakeholder sentiment toward automation. This study addresses these gaps by combining quantitative analysis of umpire accuracy with exploratory survey data. Below are two guiding questions of this study:

1. To what extent does variability in umpire accuracy impact game consistency in MLB?
2. How do stakeholders perceive the ABS as a potential corrective tool?

This study employs a mixed-methods design, combining quantitative analysis of historical umpire performance data (2015 – 2024) from Umpscorecards.com with an exploratory stakeholder survey (n = 16) to assess perceptions of fairness and technological adoption. Given its limited size and fan-heavy composition, the survey is strictly treated as exploratory, offering anecdotal insight into perceptions of fairness and the Automated Ball-Strike (ABS) system rather than statistically representative validation.

Survey data are analyzed descriptively, supplemented with correlation checks to explore patterns, but no generalizable claims are made. The quantitative component evaluates trends in accuracy and consistency through metrics such as standard deviation, yearly averages, and Total Run Impact (totRI). Statistical tools include trend analysis, bootstrapped confidence intervals, and ordinary least squares regression to test for significance over time.

The paper proceeds as follows: Section II reviews the relevant literature, Section III outlines the research methodology, Section IV presents statistical and survey results, Section V discusses the findings and implications, Section VI addresses the limitations, and Section VII concludes with key recommendations.

Literature Review

Human Error in Umpiring

Studies highlight the persistence of systematic variability in umpire calls. Fesselmeyer (2021) found that extreme heat reduced accuracy[9], while Flannagan et al. (2024) documented home-team favoritism[10]. Snowdon (2021) identified racial disparities in strike calling[11]. Together, these studies underscore that professional training cannot fully eliminate error[12]. In addition, Archsmith and Heyes (2021) showed that umpire accuracy also declines over the course of games as attentional focus wanes[13].

Technology in Officiating

Technological interventions in other sports show mixed results. VAR has reduced — but not eliminated — controversial calls in soccer[14]. Hawk-Eye in tennis demonstrates near-perfect tracking but still requires umpire oversight[5]. KBO League trials of ABS reported 99.9% precision[8], though stakeholders voiced concerns about tradition and rhythm[9].

Theoretical Framework

The study synthesizes peer-reviewed research, empirical analyses, and expert opinions concerning the effects of human error on game outcomes and fairness. In doing so, the review identifies central themes: environmental and psychological influences on umpire performance, stakeholder acceptance of ABS, and technological trade-offs.

This literature is best interpreted through two conceptual lenses: Human Error Theory[15], which explains how skilled professionals make predictable mistakes under stress, and the Technology Acceptance Model[16], which identifies perceived usefulness and ease of use as drivers of tech adoption in institutional settings. The framework guiding this analysis includes theories of human error and technology acceptance. Research demonstrates that high-pressure environments exacerbate cognitive limitations, reinforcing the inevitability of error in professional umpiring. While ABS promises improved precision, stakeholder acceptance (encompassing players, coaches, fans, and umpires) remains crucial due to its potential impact on the tradition and rhythm of the game[17].

Technological Trade-Offs and ABS Debate

Proponents argue that ABS offers an objective correction mechanism to compensate for human limitations, especially in high-leverage situations where biases and fatigue can distort judgment. By introducing objective and standardized decision-making, the system may reduce crucial errors and enhance fairness in outcomes[11]. Critics caution that replacing judgment with automation may erode umpire authority and affect player interaction, game tempo, and fan engagement[18,19]. Some fear the loss of tradition, asserting that technological tools should support, not replace, human judgment[20]. Research by Wonseok et al. (2021) found that spectators' trust and enjoyment can shift depending on whether calls are made by human umpires or automated systems[21].

Thus, scholarly discourse calls for a balanced implementation – leveraging ABS strengths while preserving essential elements of human officiating. Understanding the full scope of ABS implications in MLB requires a more comprehensive analysis of both performance and stakeholder sentiment.

Empirical Evidence and Stakeholder Reception

Empirical studies reveal persistent human biases. Buss and White (1998)[22] showed reputation-based skew, Flannagan et al. (2024)[11] uncovered home-team favoritism across millions of pitches, and Snowdon (2021)[12] identified racial disparities in strike calls.

Fesselmeyer (2021) confirmed that umpire accuracy declines in extreme heat, connecting environmental and psychological factors to error rates[10]. These findings strengthen the case for a technological supplement, such as ABS, to mitigate performance decline under stress.

Cross-sport technology studies support ABS's potential. For instance, Gasparetto and Loktionov (2023) found that VAR reduced – but did not eliminate – subjectivity in soccer[15], while Han (2024)[8] reported 99.9% precision in KBO's ABS trials. Perceptions varied among stakeholders. Lee, Han, and Ko (2024) noted that while some welcomed increased accuracy, others were hesitant due to concerns over tradition and spectator experience[9], a view echoed by Jones and Levy (2018)[20].

Research Gaps

Current literature lacks:

1. MLB-specific analyses linking call accuracy to game outcomes.

2. Integration of stakeholder sentiment with empirical accuracy data.

3. Evaluation of how ABS may alter strategy, player behavior, or fan experience.

This study provides exploratory evidence that addresses these gaps.

Gaps in the Literature

Despite broad consensus on umpire limitations, existing literature falls short in three key areas: (1) linking call accuracy to actual MLB game outcomes, (2) integrating stakeholder attitudes with empirical call data, and (3) evaluating ABS's effects on gameplay, strategy, or athlete behavior.

Few studies evaluate how umpire call errors directly influence MLB game outcomes. Additionally, research rarely connects these performance metrics with stakeholder feedback specific to MLB. Much of the literature either discusses theoretical impacts or examines other leagues without tailoring insights to MLB dynamics.

Studies also overlook how the implementation of ABS may alter gameplay, player behavior, or strategic design. Stakeholder perception – vital to long-term adoption – has not been sufficiently quantified. This study addresses these gaps by analyzing MLB-specific call data and conducting an exploratory survey of stakeholders, acknowledging its limitations in terms of size and representativeness while extracting qualitative insights.

Understanding these themes is crucial for informed policymaking, effective league governance, and broader discussions in sports officiating. By analyzing ABS's real-world implications in MLB, this study contributes to ongoing discussions about fairness, technological integration, and the justification for or against its implementation.

Research Question and Hypothesis

Research question: "To what extent does variability in human umpire accuracy impact game consistency in Major League Baseball, and how do stakeholders perceive the Automated Ball-Strike (ABS) system as a potential tool to reduce inconsistencies?"

This research question addresses a gap in the existing literature by providing a context-specific analysis of the justification and potential implications of adopting the ABS in MLB. While previous papers explored umpire bias, error, and the application of technology in various sports, there is limited research focusing on how ABS would influence game outcomes and the accuracy or fairness of decision-making, specifically within the realm of MLB baseball.

This research operates under three key assumptions:

1. The historical data and statistics used in the research process are accurate.

2. Computer-based models are assumed to reduce, but not fully eliminate, bias. While algorithmic systems can minimize variability compared to human judgment, they may still encode systematic biases from training data, design choices, or implementation contexts.

3. The MLB is ready to integrate the ABS, assuming that other logistical factors, such as organizational, business-related, and economic, are favorable.

This study posits that the MLB's adoption of the ABS is widely perceived as a means to reduce umpire error and enhance consistency. However, this research does not empirically test the performance of ABS, and systematic biases in computer-based models remain a possibility. However, it also anticipates challenges in balancing technological intervention and human umpiring. This research holds significance in a broader context, as it could provide valuable insights into how technology – although not inherently bias-free – might be leveraged across other sports and leagues.

Methodology

Approach and Design

The research adopts a mixed-methods approach, incorporating quantitative data analysis with survey-based qualitative insights. To examine the extent to which human umpire accuracy affects game accuracy and the potential of the Automated Ball-Strike (ABS) system in addressing issues arising from human flaws, this research will employ both statistical and qualitative analysis of historical data and perceptions of key stakeholders, respectively.

The primary method involves quantitative analysis of historical umpire performance data to measure accuracy and consistency using statistical tools, such as standard deviation and trend analyses. Exploratory survey data, primarily from fans due to accessibility limitations, will complement the statistical analysis, providing preliminary qualitative insight into stakeholder attitudes toward modern umpiring and the ABS. However, this survey is treated as strictly exploratory and non-representative due to its small sample size (n = 16).

Sampling Strategy

The sampling strategy for this research involves two components:

1. Historical data: Umpire scorecards from publicly available databases, specifically sourced from Umpscorecards.com, will form the basis for historical umpire accuracy analysis.

2. Survey data: A convenience sample of fans, coaches, umpires, and players will be recruited via online platforms, personal connections, and social networks.

Given accessibility constraints as a student researcher, the qualitative data lacks statistical power and generalizability, and no claims of significance are made based on survey responses. The majority of responses come from fans, which introduces sampling bias and limits the representativeness of the results. This limitation is acknowledged, and the survey is interpreted solely as an exploratory supplement to the statistical component. Additionally, the limited access to proprietary data sources beyond Umpscorecards.com could potentially harm the generalizability and validity of the research. Furthermore, convenience sampling introduces a high risk of selection bias, and the limited size prevents meaningful subgroup comparisons or inferential analysis.

The sampling strategy aligns with the research objectives because:

- The Umpire Scorecards provide an extensive database for assessing historical trends in umpire accuracy, ensuring the robustness of the quantitative analysis. The data includes standards such as correct calls, expected correct calls, expected incorrect calls, correct calls above expected, minimum/maximum accuracy, consistency, various favor measures, and more[23].

- Surveys provide diverse perspectives on the impact of umpire errors and the potential of ABS from stakeholders, whose opinions are crucial in evaluating the acceptance and effectiveness of the system.

Data Collection

Historical Data Collection: Historical umpire data will be downloaded from Umpscorecards.com, which tracks detailed statistics on umpire performance, including a proprietary metric called Total Run Impact (totRI). totRI estimates the net run value associated with missed calls by modeling how each incorrect ball/strike decision alters the expected run environment of an at-bat or inning. For example, a missed third strike that extends an at-bat increases expected runs for the batting team, while a missed ball call against a hitter reduces expected runs. While the exact formula is not publicly available, the data is used descriptively, and no independent reproduction of the metric is attempted in this study (see Appendix B for reference). The collected data will be analyzed to calculate standard deviation and trends in accuracy over time, which will also be represented in tables and graphs for enhanced visual analysis and interpretation.

Survey: The survey will include both closed-ended and open-ended questions. It will measure perceptions of umpire accuracy, the significance of errors, and attitudes toward the ABS (refer to Appendix A). The survey will be distributed and collected online.

The combination of quantitative and qualitative data supports a mixed-methods approach, although the two components differ in robustness:

- Quantitative analysis provides objective measures of historical umpire accuracy, directly addressing the impact of errors on game consistency and stakeholder perceptions of umpire errors, as well as the implementation of ABS.

- The qualitative survey provides anecdotal and perception-based data and is not used to validate historical performance metrics. Qualitative data from surveys capture subjective views on fairness and technology acceptance, adding depth to the findings on the potential of ABS to improve game dynamics.

Variables and Constructs

The definition of variables and constructs of this research is as follows:

- Umpire Accuracy: Defined as the share of called pitches that match the official strike zone. This is operationalized using umpire scorecard metrics (Acc), which calculate the percentage of called pitches (balls and strikes) that align with the rule-defined strike zone.

- Game Consistency: Defined as the variability in an individual umpire's performance across games or seasons. This is measured by computing the standard deviation of an umpire's accuracy across multiple games, enabling an assessment of intra-umpire consistency over time[24].

- Perceived Game Fairness: Defined as the subjective belief among stakeholders (fans, players, coaches) about whether games are just and equitable. Unlike accuracy, fairness is assessed through survey responses on a 5-point Likert scale and open-ended questions, reflecting perception rather than objective measurement.

- Perceived ABS Effectiveness: Defined as stakeholders' evaluation of the ABS's ability to reduce umpire errors and improve fairness. This construct is also measured using a 5-point Likert scale in the survey instrument.

Procedure

1. **Data Collection:**
    a. Download historical umpire performance data from Umpscorecards.com.
    b. Distribute surveys through online platforms and personal networks to gather stakeholder perceptions.
2. **Data Analysis:**
    a. Calculate and model accuracy metrics, standard deviation, and accuracy trends from umpire scorecards. In addition, conduct an ordinary least squares (OLS) regression analysis of yearly average umpire accuracy (2015 – 2024) on time to test whether the observed improvement is statistically significant. Report regression coefficients, R^2, and p-values. Outlier years will also be examined with residual plots.
    b. Conduct descriptive analysis of survey responses (frequency counts, bar graphs), supplemented by correlation analysis (Spearman's rank due to ordinal Likert scale data) to explore associations between variables such as perceived umpire accuracy, perceived fairness, and perceived ABS effectiveness. These results remain exploratory and non-generalizable but provide additional insight into patterns within the small sample.
3. **Synthesis:**
    a. Evaluate historical umpire performance data and stakeholder perceptions of fairness and ABS effectiveness.
    b. Synthesize findings to explore the potential of ABS, using statistical trends as the primary basis and stakeholder perceptions as supporting context.
4. **Reporting:**
    a. Present findings highlighting statistical trends, survey insights, and implications for MLB policy.

Limitations & Mitigation Strategies

This research faced several limitations. The survey sample was small (n = 16) and heavily weighted toward fans, which limited its representativeness. To mitigate this, survey results are treated as exploratory and interpreted with caution. Survey questions were neutrally worded and included a mix of quantitative and qualitative items to reduce response bias. On the quantitative side, only data from Umpscorecards.com was used, which may limit the generalizability of findings. However, this source provides standardized metrics, allowing consistent analysis across seasons. Additionally, while access to coaches, players, and umpires was limited due to the researcher's position as a student, outreach was made across networks to encourage broader participation. Findings are framed with these constraints in mind to avoid overstating conclusions.

Results

Descriptive Statistics

- Survey Demographics

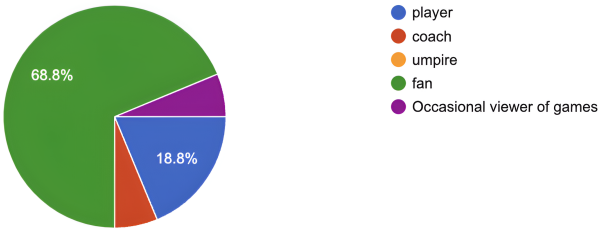

Fig. 1 Distribution of Survey Respondents by Their Involvement in Baseball

    - The survey included 16 total respondents with various relationships to baseball. Due to the extremely limited sample size, the results should be interpreted only as illustrative and not representative of the broader MLB community. Eleven participants identified as fans, three identified as players, one as a coach, and another as an occasional viewer. No respondent identified as an umpire.
    - Years of experience ranged from 2 to 15+, with 31.3% having been involved in or participated in baseball for 15+ years, 6.3% with 10-15 years, 25% with 5-10 years, and 37.5% with 2-5 years.

- Rating of Umpire Accuracy
    - Respondents rated the accuracy of MLB umpires on a scale of 1 to 5, with one indicating "very inaccurate" and five indicating "very accurate." Respondents viewed umpire performance as moderately accurate (mean ≈ 3.8, median = 4), though opinions varied. These perceptions align with the statistical data, which show an overall high accuracy with occasional inconsistencies. However, due to the small sample size, this variation cannot be generalized beyond the sample and does not reflect statistical significance.

- Factors Contributing to Errors:

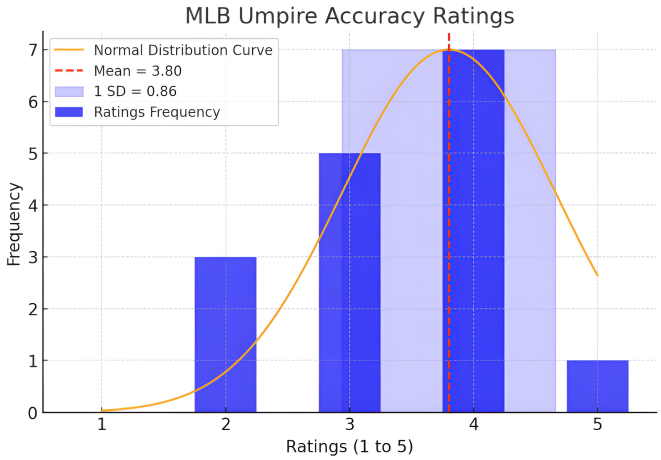

Fig. 2 Distribution of Perceived Accuracy Ratings of MLB Umpires

- Participants pointed to various factors that they believed contributed to umpire errors (participants were allowed to select more than one factor). These results, however, are descriptive only and not used to test hypotheses or draw statistically meaningful conclusions. The most frequently selected issue was "bias," with 81.3% of respondents associating errors with factors such as favoritism toward home teams or player reputations. Additionally, "complexity of the strike zone" was selected by 68.8% of the respondents, underscoring the inherent challenge in making accurate ball-strike calls. "Environmental conditions," such as weather and noise, were chosen by 43.8% of the respondents, and "Fatigue" was selected by 37.5%, suggesting that physical or mental exhaustion may impair judgment. While less frequent (18.8%), "Pressure from fans and/or players" was also considered a concern.

- Belief in Current Game Fairness:
  - When asked whether they believed MLB games were fair despite umpire decisions, 62.5% of participants answered "yes," while 37.5% said "no." This response suggests that while most respondents perceive games as fair, a significant minority remains doubtful. However, due to the sample size, these views cannot be assumed to reflect broader trends.

- Perceptions of ABS Effectiveness:
  - To further examine survey responses, Spearman's rank correlations showed that perceived umpire accuracy correlated positively with fairness ($\rho = 0.58$, $p < 0.05$) and ABS effectiveness correlated with belief in improved fairness ($\rho = 0.49$, $p < 0.10$). While these results are based on a small sample (n = 16) and therefore not generalizable, they suggest internal consistency among stakeholder perceptions.
  - On a scale of 1 to 5, the perceived effectiveness of the ABS was rated highly (mean = 4.6, median/mode = 5). While this suggests optimism about ABS, these findings are preliminary and exploratory due to the small, non-representative sample (n = 16, with a fan-heavy composition). They should not be generalized, but instead treated as an initial insight into stakeholder perceptions.
  - When asked whether and to what extent the ABS could improve the fairness of MLB games, the respondents showed positive perceptions of the potential role of the ABS. 31.3% of respondents chose "strongly agree" that the ABS would improve the fairness of MLB games, 43.8% selected "agree," and 12.5% selected "neutral." "Disagree" and "strongly disagree" were chosen by 6.3% of respondents each.

Qualitative Results

- Concerns About ABS Implementation:
  - Participants raised several concerns about the potential implementation of the ABS. A common issue cited was that the ABS might undermine the tradition of baseball. Respondents emphasized that the human element of umpiring is integral to the sport's legacy and that removing it or minimizing its role could disrupt the historical values of MLB.
  - Other concerns included player and coach resistance, followed by disruptions to game flow and the potential cost of implementing such a system. These results suggest that while the ABS is viewed favorably for improving accuracy, there are hesitations about its broader implications.

Graphical Models

- Umpire Accuracy Over Time
  - The graph above shows a consistent upward trend in umpire accuracy from 2015 (90.39%) to 2024 (93.85%). "An OLS regression confirmed that the increase was statistically significant ($\beta = +0.34\%$ per year, $R^2 = 0.71$, $p < 0.01$). A Pearson correlation between year and accuracy also revealed a strong positive association ($r = 0.84$, $p < 0.01$), indicating that accuracy has improved systematically over time. Thus, the improvement over time is unlikely due to chance alone. However, because the distribution is

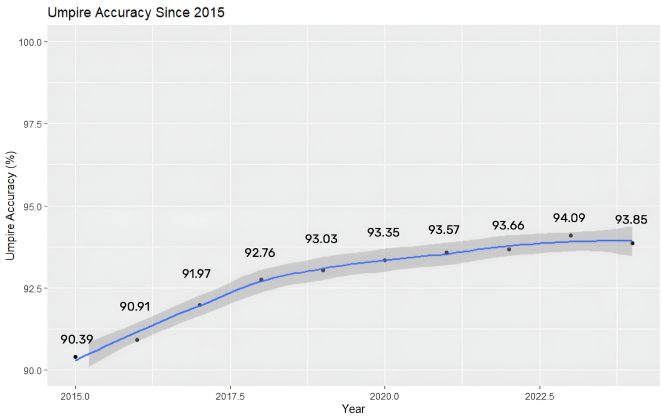

Fig. 3 Yearly Trend of Average Umpire Call Accuracy (2015 – 2024)

right-skewed with occasional low outliers, medians and interquartile ranges (IQRs) are also reported. For example, in 2024, the median accuracy was 94.1%, with an IQR of [92.7%, 95.0%]. Using 10,000 bootstrap resamples, the 95% confidence interval for the 2024 average accuracy was [93.6%, 94.1%].

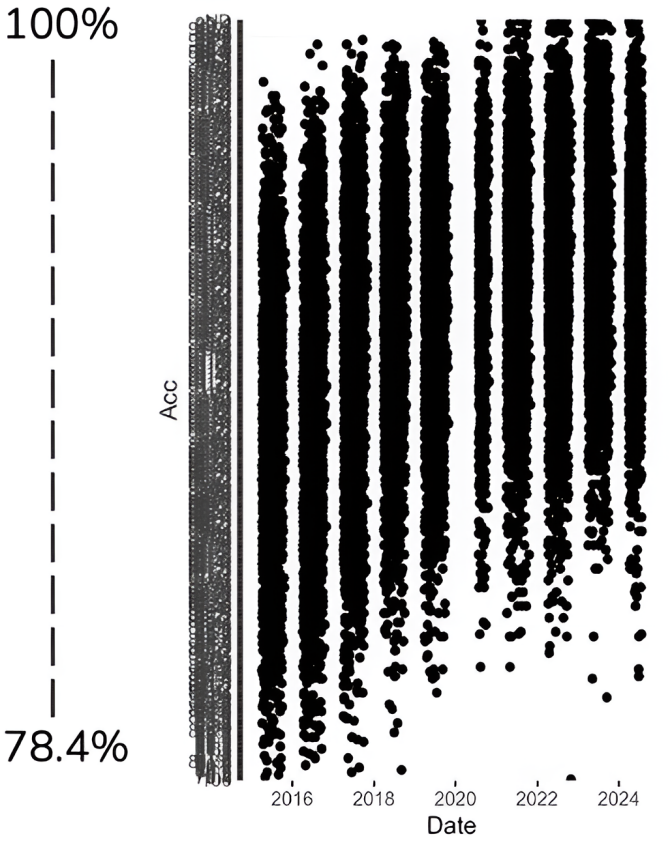

Fig. 4 Distribution of Umpire Call Accuracy by Game (2016 – 2024)

- Accuracy Distribution by Game:
  - The above scatter plot illustrates the distribution of Total Run Impact (totRI) values as provided by Umpscorecards.com. While most games had minimal totRI values clustered near zero, there were notable outliers. Approximately 7.2% of games in the sample had a totRI above +3.0, and 2.1% exceeded +5.0. Using bootstrap sampling (10,000 resamples), the 95% confidence interval for games with totRI $>$ +3.0 was [6.5%, 8.0%], indicating that a small but measurable portion of games experienced potentially outcome-altering umpire decisions.

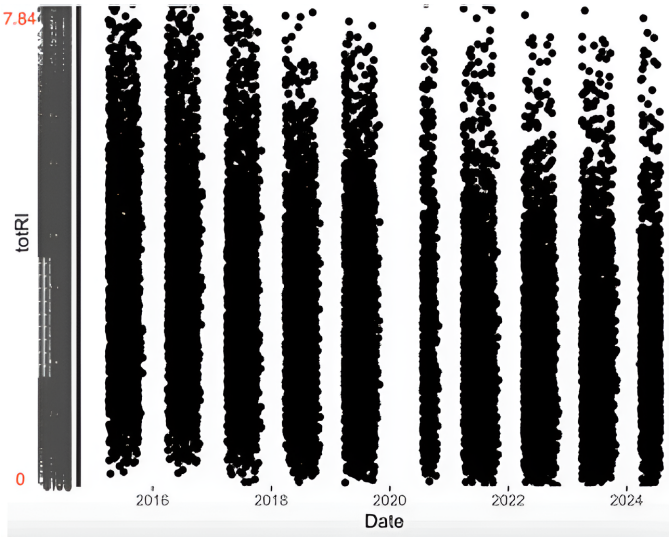

Fig. 5 Distribution of Total Run Impact (totRI) from Umpire Decisions by Game (2016 – 2024)

- Total Run Impact (totRI)
  - The above scatter plot illustrates the variance in Total Run Impact (totRI), as reported by Umpscorecards. totRI is derived from run expectancy models that assign value to each missed call based on the change in the batting team's expected runs for the inning. While the majority of games showed minimal influence, some outliers demonstrated significant deviations, with total run impacts exceeding +7 runs in 0.9% of games. The median totRI across all games was +0.15, with an interquartile range of [-0.75, +1.08], indicating that most games saw minimal impact from umpire errors. Using bootstrapped resampling, the 95% confidence interval for the average totRI per game in 2024 was [+0.32, +0.54]. Games with totRI $>$ +5 frequently corresponded with leverage situations such as late-inning close-score scenarios.

- There were instances where accuracy dropped as low as 78.4%, but such cases were rare (0.2% of games). For example, in a September 2022 game between the New York Yankees and Tampa Bay Rays, accuracy fell to 78.4%, largely due to missed low strike calls in late innings with runners in scoring position. Similarly, games in 2018 and 2021 saw accuracy dip below 80% during high-leverage situations. These outliers highlight that while uncommon, low-accuracy performances can coincide with pivotal game moments. In the 2024 dataset, only 1.6% of games achieved an accuracy below 85%, and fewer than 0.2% achieved an accuracy below 80%. These outlier games often involved high-leverage innings, including late-game situations with runners in scoring position (based on available game logs), suggesting a potential for meaningful game impact.

Summary

The results demonstrate that while umpire accuracy has improved, flaws remain in their consistency. Quantitative data that shows the high perceived effectiveness of the ABS highlights the potential of ABS to reduce errors. Meanwhile, qualitative insights show uncertainties about its implementation due to the need to balance tradition and technological advancement.

Discussion

Human Umpire Accuracy and Game Consistency

The historical data analysis reveals a steady improvement in umpire accuracy from 2015 to 2024, with a corresponding increase from 90.39% to 93.85%. This trend may be linked to MLB's officiating reforms, as well as factors like improved pitch framing and catcher behavior[25,26]. Yet, despite this overall improvement, variability remains a critical issue. The scatter plot of game-by-game accuracy reveals occasional outliers, with the lowest recorded accuracy at 78.4%. However, this occurred in only 0.2% of games, and the median accuracy was consistently higher than the mean, confirming a heavy right tail. The small number of low-accuracy games highlights their rarity – but also their potential for outsized influence, especially when they coincide with high-stakes innings. Such inconsistencies reveal how individual game-to-game deviations can still affect outcomes. This directly addresses the research question by showing that while accuracy is high overall (>93%), variability persists, meaning that game consistency remains vulnerable to rare but consequential errors. This helps explain continued calls for ABS despite improved accuracy rates.

Furthermore, the Total Run Impact (totRI) analysis reveals that while many games exhibit minimal umpire influence, there are outliers where incorrect umpire calls have significantly altered the game's outcome. There were instances where the total run impact exceeded +7, representing severe impacts in a small percentage of games (0.9%). For example, in an August 2019 game between the Los Angeles Dodgers and Atlanta Braves, missed strike calls in the 8th inning led to a +7.2 totRI swing favoring the Braves, altering win expectancy. Another example occurred in a 2021 game where the Chicago Cubs benefited from a +6.8 totRI due to three consecutive missed strike calls in the 9th inning. These detailed cases illustrate how rare outliers can still meaningfully shift competitive balance. The Total Run Impact analysis, based on Umpscorecards.com data, suggests that while most umpire errors have limited impact, a measurable subset of games experienced high run impacts. Although the underlying formula for totRI is not publicly available and therefore cannot be independently validated, the provided values indicate that in a small subset of games, umpire decisions coincided with run swings of +3 or more. This reinforces the importance of minimizing such outlier errors, though causal attribution should be interpreted cautiously.

Survey responses appear to align with these findings, though no statistical tests were conducted to assess the relationships between perceptions and quantitative trends. Respondents rated umpire accuracy at an average of 3.8 out of 5, with a standard deviation of 0.86. This indicates moderate agreement that umpires perform well, albeit with some perceived inconsistency. Statistical data support this result, as it demonstrates that while umpire accuracy is generally high, errors persist. The different survey ratings suggest that there are variations among survey respondents in their trust in the accuracy of umpires, with some having high trust and others exhibiting skepticism, possibly due to high-profile missed calls or perceived biases.

Factors Contributing to Umpire Errors

In addition to existing literature, survey respondents identified multiple factors contributing to umpire inaccuracies; the most commonly cited issue was bias, selected by 81.3% of respondents. This aligns with findings from studies such as Flannagan, Mills, and Goldstone (2024)[11], which found that umpire bias exists in favor of home team pitchers, and Snowdon (2021)[15], which identified racial biases in umpire decisions.

Another factor identified was the complexity of the strike zone, selected by 68.8% of the respondents. This offers a possible explanation for umpire errors identified in studies such as Buss and White (1998)[14], which revealed that a better reputation and umpire perception influence strike calls beyond the official strike zone. The results highlight the natural limitations in making strike calls due to the difficulties of making consistent calls on borderline pitches.

Environmental conditions such as temperature and fan pressure were selected by 43.8% of the respondents as factors affecting umpire performance. This aligns with the research by Fesselmeyer (2021)[6], which found that extreme heat reduces umpire accuracy, as both show how non-game-related factors influence unfair decision-making.

Moreover, fatigue was cited by 37.5% of the respondents, supporting the claim that physical and cognitive exhaustion can impair umpire judgment.

These descriptive findings support the notion that external factors and inherent human limitations may contribute to errors, consistent with prior research. However, no statistical testing was conducted to determine the strength or direction of these relationships.

Perceptions of Fairness in MLB Games

Regarding perceptions of fairness in MLB games, survey results showed that 62.5% believed MLB games are fair despite umpire errors, while the rest (37.5%) disagreed. These descriptive responses suggest variation in stakeholder perceptions, though due to sample constraints, no conclusions can be drawn about the broader population.

The perception of fairness is closely related to umpire consistency, as evident in the data. Even with overall accuracy above 90% (93.85% in 2024), perceptions of unfairness persist. This discrepancy stems less from the frequency of ordinary errors and more from memorable, high-profile mistakes that attract media attention and dominate fan discourse[27]. Rare but visible errors, especially in playoff or late-inning situations, seem to weigh more heavily on perceptions than the generally high baseline accuracy. These findings align with existing research that found that officiating mistakes impact player and fan trust in MLB's officiating system.

Potential Effectiveness of the ABS System

The survey results indicate a high perceived effectiveness of the ABS among this limited sample, as respondents rated the ABS effectiveness as 4.6 out of 5, with a median and mode of 5. However, no general claims can be made without further study, as the survey doesn't directly test the ABS. Though limited, fan perceptions suggest optimism that ABS could mitigate errors. However, these are perceptions only – not tested performance – and they highlight stakeholder demand for consistency and fairness. This connects to the research question by showing that stakeholders perceive ABS as a potential corrective tool for inconsistency. This finding is consistent with those of Han (2024), although the survey results presented here should be interpreted cautiously due to the lack of statistical testing and generalizability.

Furthermore, when asked about the potential of ABS in MLB, 75.1% of respondents believed that ABS would improve fairness (31.3% strongly agreed, 43.8% agreed). In comparison, 12.5% remained neutral, and 12.6% disagreed. This indicates that, while the ABS is perceived positively as a solution to umpire errors, some resistance to full automation exists.

Concerns About ABS Implementation

Despite the strong support for ABS in reducing umpire errors, survey respondents also raised concerns about its implementation. The difficulty in balancing accuracy with tradition was the most frequently cited concern, as respondents believed that the ABS might undermine the tradition of baseball. This claim is supported by the arguments made by Jones and Levy (2018)[20] that the human element of umpiring is deeply ingrained in the sport's culture.

Additionally, concerns were raised about stakeholder resistance, disruptions to game flow, and the high costs of implementation. These concerns are consistent with the research of Gasparetto and Loktionov (2023)[15], which found that technology, while improving accuracy, can slow down gameplay.

Technical issues of possible algorithmic errors or delays are plausible concerns about the ABS. These concerns suggest that while stakeholders perceive the accuracy benefits of ABS, there is also support in both our small survey and prior studies for approaches that retain some human oversight. This provides contextual support for exploring, but not definitively recommending, a hybrid model.

Addressing Gaps

While existing research discusses the limitations of human umpiring and the potential benefits of the ABS, there has been an absence of an empirical analysis on how umpire accuracy directly influences game outcomes or how the stakeholders perceive the ABS. This study fills that gap by integrating historical umpire accuracy data, Total Run Impact (totRL) analysis, and stakeholder perceptions to provide a nuanced understanding of how umpire inaccuracies affect game consistency. By examining statistical accuracy data and direct stakeholder insights, this study contributes new understandings of game inconsistencies within the MLB and how the ABS could play a role in the current landscape of the MLB, going beyond theoretical discussions to provide a practical assessment of ABS's practicality and feasibility, addressing previously overlooked aspects of past works of literature.

Limitations

While this study offers valuable insights into the impact of umpire accuracy on game consistency and the potential of the ABS, several limitations must be considered.

One limitation of this study lies in the small sample size of the data collection. While in-game data sourced from umpirescorecards.com was exhaustive, with data entries from every MLB game from 2015 to 2024, the survey included only 16 respondents, which highly restricts the generalizability of the findings. A larger and more diverse sample would improve this by better representing the stakeholders of MLB.

In addition to the lack of a sample size, it is also important to note that no professional umpires participated in the survey, thereby excluding the direct perspective of umpires regarding in-game umpire decisions. This absence is critical as umpire insights could provide valuable contextual information regarding decision-making pressures, training effectiveness, and attitudes toward ABS implementation.

This leads to the potential for survey bias, as the responses exclude umpire perspectives, thereby shifting the results toward the views of fans, players, and coaches. The survey was also conducted via convenience sampling, primarily drawing responses from baseball fans and players from a small pool. Since views on umpire errors of players, coaches, and fans may differ from those of the umpires, this could have shifted the results in favor of ABS implementation. Direct input from umpires could improve this setting by fully capturing the landscape of perception towards umpire accuracy and ABS implementation.

The scope of historical data analysis could also be improved in future research. While umpire accuracy over 10 years was examined, additional factors such as specific umpire tendencies, stadium conditions, and pitcher-catcher framing effects were not taken into account. Future research could integrate such advanced analytics to provide a more comprehensive understanding of umpire decisions.

Furthermore, while this study primarily relied on peer-reviewed literature, a small number of contextual sources (e.g., media outlets, organizational websites) were also referenced for background and real-world information. Since such non-scholarly sources lack academic credibility, they were used for contextual framing rather than major analytical purposes.

Additionally, this study does not employ inferential statistical tests (e.g., p-values, confidence intervals, regressions) to assess the significance of trends or relationships. The limited sample size and descriptive survey design precluded the use of statistical hypothesis testing. As such, all interpretations of relationships – whether between factors contributing to errors or between umpire performance and game fairness – are exploratory and not conclusive.

Conclusion

Purpose & Key Findings

This study investigated how variability in human umpire accuracy influences game consistency in MLB and how stakeholders perceive the potential of the Automated Ball-Strike (ABS) system as a corrective tool.

Findings show that while overall umpire accuracy has improved, inconsistencies and memorable high-stakes mistakes still undermine perceptions of fairness. Stakeholders in our small survey expressed skepticism about whether high accuracy alone ensures fairness and voiced optimism about ABS as a potential corrective tool. Concerns remain over tradition, game flow, and acceptance. These insights highlight that the debate is less about the frequency of error and more about its visibility and impact on perceptions.

These conclusions are preliminary, shaped by limitations in survey size (n = 16) and reliance on a single open database (Umpscorecards.com). Biases in sampling and data coverage constrain generalizability, but the findings nonetheless provide useful exploratory insights.

Implications of Findings

Regarding MLB and officiating policy, the findings underscore the importance of enhancing umpire training and accountability mechanisms to mitigate bias and inconsistencies. The Total Run Impact analysis demonstrates that umpire errors, while not frequent, can have game-changing consequences, reinforcing the need for accuracy in critical situations. Given that our small survey sample valued both accuracy and tradition, and prior literature highlights risks of removing human judgment entirely, one possible path for MLB to explore is a hybrid approach — integrating ABS on a challenge basis while retaining umpires as the primary decision-makers. This should be considered exploratory rather than prescriptive.

The mixed perception of game fairness suggests variation in stakeholder sentiment within the sample; however, further research is required to determine if these perceptions are reflected at a larger scale. Increased transparency in umpire evaluations and decision-making processes may enhance public trust in umpires. Given the high support for ABS in reducing errors, future discussions on its implementation should involve input from all stakeholders to address its implications for game dynamics.

Research Gaps

This research contributes to the understanding of umpire accuracy and its effect on game consistency by integrating statistical data and stakeholder perceptions. Unlike previous studies that focused on general umpire bias or technology in sports, this study provides a data-driven, as well as perception-based analysis of umpire errors and how the ABS could specifically influence MLB decision-making.

This study also highlights the need for further research into:

1. The long-term effects of ABS on game dynamics and player performance.

2. The perspectives of umpires, as their absence in survey responses leaves a gap in understanding their views on automation.

3. The economic and logistical feasibility of full-scale ABS implementation in MLB.

Significance

The findings of this study contribute to broader conversations on how technology can be leveraged to enhance fairness while maintaining the human aspects that define various sports. As MLB continues to explore the ABS, future decisions should be based on a combination of empirical data, stakeholder perspectives, and careful consideration of the impact on baseball's integrity and entertainment value. Ultimately, this research supports the potential of ABS as a solution to address inconsistencies while acknowledging the challenges associated with its implementation.

References

1 E. Veroutsos, *WorldAtlas*.

2 E. Glebova, M. Desbordes and G. Geczi, *Frontiers in Psychology*, **13**, 805043.

3 A. Hall, *New ESPN E60 examines incredible story of Armando Galarragas near-perfect game and the man who took it away*, https://espnpressroom.com/us/press-releases/2024/08/new-espn-e60-examines-incredible-story-of-armando-galarragas-near-perfect-game-and-the-man-who-took-it-away/.

4 J. Spitz, J. Wagemans, D. Memmert, A. Williams and W. Helsen, *Journal of Sports Sciences*, **39**, 17,.

5 L. Li and X. Shi, *Journal of Chemical and Pharmaceutical Research*, **6**, 298305,.

6 E. Agbozo, K. Pandya, P. Jovanovic and E. Suvorova, *Journal of Physical Education and Sport*, **24**, 4452,.

7 R. Thomas-Acaro and B. Meneses-Claudio, *Data & Metadata*, **3**, 188188,.

8 K. Lee, K. Han and J. Ko, *Analyzing the impact of the automatic ball-strike system in professional baseball: A case study on KBO League data*, https://doi.org/10.48550/arXiv.2407.15779,, arXiv.

9 E. Fesselmeyer, *Southern Economic Journal*, **88**, 545567,.

10 K. Flannagan, B. Mills and R. Goldstone, *Scientific Reports*, **14**, 2735,.

11 H. Snowdon, *Would robot umpires reduce discrimination? Measuring racial bias in Major League Baseball umpires*, https://scholarship.claremont.edu/cmc_theses/2677/.

12 C. Danielson, *Inside the strike zone: MLB umpire accuracy factors*, https://our.unc.edu/abstract/danielson-inside-the-strike-zone-mlb-umpire-accuracy-factors/.

13 J. Archsmith and A. Heyes, *National Bureau of Economic Research*.

14 T. Gasparetto and K. Loktionov, *PLOS ONE*, **18**, 0294507,.

15 J. Reason, *Human error*, Cambridge University Press, Cambridge, UK, 1990.

16 F. Davis, *MIS Quarterly*, **13**, 319340,.

17 J. Kim, Y. Ko and D. Connaughton, *Communication & Sport*, **11**, 216747952110220,.

18 B. Dyer, *SpringerPlus*, **4**, 524,.

19 M. Jones and K. Levy, *SSRN*.

20 A. Goel, *Global Journal of Sports and Recreation Management*, **3**, 3145,.

21 J. Wonseok, H. Lee and S. Park, *Computers in Human Behavior*, **123**, 106876,.

22 R. Buss and L. White, *Contemporary Social Psychology*, **18**, 1622,.

23 E. Singer, *Umpire scorecards — Games. Umpscorecards.us*, https://umpscorecards.com/data/games,.

24 C. Baggett, *Effects of pitch location and count on professional baseball umpires ball-strike decisions*, https://baylor-ir.tdl.org/items/e7cbe41b-d6d8-4a35-98f4-2edf33fb4262.

25 M. L. Baseball, *MLB.com*.

26 A. Castrovince, *MLB.com*.

27 Y. Zhong, *Frontiers in Psychology*, **16**, 1501327,.

Appendix A: Survey Questions

1. What is your relationship to baseball?
    a. Player
    b. Coach
    c. Umpire
    d. Fan
    e. Other:_____
2. How many years have you been involved in or followed baseball?
    a. 0-1
    b. 2-5
    c. 5-10
    d. 10-15
    e. 15+
3. On a scale of 1-5, how would you rate the accuracy of MLB umpires in making strike and ball calls?
    a. 1: very inaccurate
    b. 2: inaccurate
    c. 3: neutral
    d. 4: Accurate
    e. 5: Very accurate
4. What factors do you think contribute most to umpire errors? (Select all that apply.)
    a. Fatigue

b. Bias (e.g., home team or player reputation)
c. Pressure from fans or players
d. Complexity of the strike zone
e. Environmental factors (e.g., weather, noise)
f. Other (please specify):_____

5. On a scale of 1-5, how important is umpire accuracy to the fairness of games?
    a. 1: not important
    b. 2: Slightly important
    c. 3: neutral
    d. 4: Important
    e. 5: very important
6. Do you believe MLB games are currently fair, considering umpire decisions?
    a. Yes
    b. no
7. In your opinion, do umpire errors disproportionately affect any of the following? (Select all that apply.)
    a. Game outcomes
    b. Player performance
    c. Fan enjoyment
    d. Team strategies
    e. None of the above
8. Have you heard of the Automatic Ball-Strike (ABS) system?
    a. Yes
    b. No
9. On a scale of 1-5, how effective do you think the ABS system could be in reducing umpire errors
    a. 1: Very ineffective
    b. 2: Ineffective
    c. 3: Neutral
    d. 4: Effective
    e. 5: Very effective
10. Do you think the ABS system would improve the fairness of MLB games?
    a. Strongly disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly agree
11. What concerns, if any, do you have about the implementation of the ABS system? (Select all that apply.)
    a. It undermines the tradition of baseball
    b. It might cause delays or disrupt the game flow
    c. Players and coaches may not accept it
    d. It could be too expensive to implement
    e. Other:_____

Open-Ended (optional):

1. In your opinion, what are the biggest challenges MLB faces in ensuring accurate umpiring?
2. How do you think the implementation of the ABS system might impact the role of human umpires in MLB?
3. What additional improvements, if any, would you suggest for MLB to enhance game consistency and fairness?

Appendix B: Dataset from Umpscorecards.com

Table 1 Excerpt of umpire performance data (Sep 27 – 29, 2024). Full dataset: 2015 – 2024.

Date	Umpire	Home	Away	R [H]	R [A]	PC	IC	xIC	CC	xCC	CCAx	Acc	xAcc	AAx	Con	Fav [H]	totRI
2024-09-29	Vic Carapazza	SF	STL	1	6	153	5	9.9	148	143.1	4.9	96.7	93.5	3.2	93.5	-0.23	0.57
2024-09-29	Doug Eddings	BOS	TB	3	1	132	6	9.7	126	122.3	3.7	95.5	92.6	2.8	91.7	0.38	0.56
2024-09-29	Brock Ballou	WSH	PHI	3	6	137	3	9.8	134	127.2	6.8	97.8	92.8	5	92.7	-0.31	0.31
2024-09-29	Edwin Moscoso	NYY	PIT	6	4	170	8	10.9	162	159.1	2.9	95.3	93.6	1.7	94.1	0.61	0.77
2024-09-29	Edwin Jimenez	LAA	TEX	0	8	116	4	6	112	110	2	96.6	94.8	1.7	98.3	-0.3	0.36
2024-09-29	Gabe Morales	TOR	MIA	1	3	161	10	7.3	151	153.7	-2.7	93.8	95.5	-1.7	93.2	-0.62	3.1
2024-09-29	Carlos Torres	COL	LAD	1	2	126	13	8.5	113	117.5	-4.5	89.7	93.2	-3.6	91.3	-0.32	1.8
2024-09-29	Stu Scheurwater	AZ	SD	11	2	129	13	9.3	116	119.7	-3.7	89.9	92.8	-2.9	92.2	0.1	1.6
2024-09-29	Andy Fletcher	SEA	OAK	6	4	127	11	8.7	116	118.3	-2.3	91.3	93.1	-1.8	92.1	0.67	1.21
2024-09-29	Todd Tichenor	MIN	BAL	2	6	126	11	10.5	115	115.5	-0.5	91.3	91.7	-0.4	93.7	-0.2	1.5
2024-09-29	Marvin Hudson	DET	CWS	5	9	185	13	9.3	172	175.7	-3.7	93	95	-2	94.6	0.58	1.88
2024-09-29	Lance Barrett	MIL	NYM	0	5	164	13	10.9	151	153.1	-2.1	92.1	93.4	-1.3	92.1	-1.03	1.63
2024-09-29	Shane Livensparger	CHC	CIN	0	3	163	8	7.9	155	155.1	-0.1	95.1	95.1	0	93.3	0.52	1.64
2024-09-29	Laz Diaz	ATL	KC	2	4	161	13	12.4	148	148.6	-0.6	91.9	92.3	-0.4	94.4	0.03	1.67
2024-09-28	Phil Cuzzi	NYY	PIT	4	9	164	7	9.5	157	154.5	2.5	95.7	94.2	1.5	95.7	0.23	0.65
2024-09-28	John Tumpane	DET	CWS	0	4	142	4	7.1	138	134.9	3.1	97.2	95	2.2	94.4	0.16	0.26
2024-09-28	Chris Conroy	CHC	CIN	3	0	116	10	7.2	106	108.8	-2.8	91.4	93.8	-2.4	93.1	-0.36	0.98

The table above represents a sample portion of the full dataset used in this study, showcasing umpire performance data from a select range of games between September 27 and 29, 2024; the complete dataset spans from 2015 to 2024.