A Statistical and Temporal Analysis of Gacha Mechanics in Wuthering Waves Using Featured Resonator Convene Records: An Exploratory Single-Account Analysis

Carpio, Lee-an Mar Anicken P.
College of Computing and Information Technologies, National University, Manila
*carpiolp@students.national-u.edu.ph

Abstract. This study presents a statistical and temporal analysis of gacha mechanics in Wuthering Waves using Featured Resonator convene records from a single-account dataset. A mechanics-aware pipeline was applied to 582 chronologically ordered pulls (9 five-star, 69 four-star), including data cleaning, pity reconstruction, 50/50 and guarantee-state labeling, exploratory profiling, exact inference, and Monte Carlo robustness checks. Rather than treating pulls as independent Bernoulli trials, the analysis modeled state-dependent behavior induced by pity and guarantee rules through cycle-level and sequence-aware features. Results show a highly imbalanced rarity structure (5-star rate = 1.55%), late-cycle concentration of most five-star outcomes (primarily high-60s to mid-70s pity), and a validated pity sequence of 67, 72, 70, 75, 71, 68, 69, 70, 19, with no hard-pity exceedance. Exact binomial testing and interval estimation indicate that observed outcomes are plausible under tested reference assumptions, while simulation confirms directional stability across alternative rate settings. Overall, the study demonstrates that account-level pull histories can support rigorous mechanics-aware inference when preprocessing, feature engineering, and validation are explicitly aligned with gacha system design.

Keywords: Gacha mechanics · Pity system · Rare-event inference · Monte Carlo simulation · Temporal analysis · Wuthering Waves

1 Introduction

Gacha systems in live-service games are commonly communicated through base drop probabilities, but actual outcomes are shaped by rule-based mechanics such as pity and guarantee states. Because probability changes with pull history, high-rarity acquisition is sequence-dependent rather than a purely independent-trial process. In practice, players interpret outcomes through metrics such as average pity, pull ratio, and 50/50 performance, yet these indicators are often presented descriptively without formal statistical validation. This creates a gap between player-facing “luck” summaries and mechanics-aware inference.

Recent access to pull-history exports and tracker-oriented analytics makes it possible to study gacha behavior at event level. Timestamped pull logs provide enough structure to reconstruct cycle boundaries, evaluate streak patterns, and examine temporal clustering. However, many analyses still rely on aggregate rates alone, which can obscure the non-stationary behavior introduced by pity progression. A more rigorous approach should combine sequence-aware modeling with exact rare-event statistics and robustness checks.

This paper addresses that need through an exploratory analysis of Wuthering Waves Featured Resonator Convene records from a single account. The dataset contains 582 pulls, with derived metrics for total currency expenditure, rarity outcomes, pity intervals, and temporal activity. The method integrates mechanics-aware pity reconstruction, descriptive and temporal analysis, exact binomial testing, confidence intervals, and simulation-based robustness. Hypothesis tests evaluate whether the observed five-star rate is consistent with a base reference (0.8%) and a pity-adjusted reference (1.6%).

The contribution of this study is a reproducible workflow that links event-sequence structure to formal inference in a transparent manner. Rather than treating gacha outcomes as stationary randomness, the analysis models pull history as a progression process with resets and threshold effects. The remainder of the paper is organized as follows: Section 2 reviews related work; Section 3 states hypotheses; Section 4 details data and methods; Section 5 presents results; and Sections 6–8 discuss implications, limitations, and conclusions.

2 Related Literature

Gacha reward systems are increasingly studied as mechanics-dependent processes rather than fixed-probability draws. In modern banners, pity and guarantee rules alter event likelihood as pulls accumulate, making outcomes sequence-dependent and often non-stationary [1], [2]. This shift has moved recent analytics from descriptive rarity summaries toward mechanics-aware modeling, rare-event inference, and temporal sequence analysis.

2.1 Mechanics-Aware Gacha Modeling

Recent formal work argues that gacha outcomes are jointly shaped by probability and system design. Chen and Fang [1] model gacha systems with strategic and stochastic structure, showing that reward architecture and progression rules can influence both expected outcomes and player behavior. This supports the view that pity state should be treated as part of the probability process, not as a post hoc explanatory variable.

Technical analyses of gacha probability computation further show how effective event likelihood changes with pull index under pity mechanics [2]. In these formulations, rare-event acquisition naturally defines cycle boundaries, and each cycle reset changes subsequent event dynamics. For empirical account-level studies, this directly motivates pity-interval reconstruction and progression-aware features such as pulls since last five-star.

2.2 Empirical Pull-Data Patterns and Threshold Effects

Empirical and technical analyses consistently report right-skewed rare-event behavior with concentration near pity thresholds [1], [3], [4]. In practical terms, pull histories often contain long low-rarity sequences followed by high-rarity resets, with occasional early interruptions. This pattern is difficult to capture with aggregate rarity percentages alone and is better represented through cycle-level and sequence-level summaries.

Game-specific mechanic resources [3], [4] are also relevant to applied analytics because they define operational parameters used in interpretation (e.g., pity structure and banner context). Although such resources differ from peer-reviewed studies in evidentiary strength, they provide implementation detail needed for reproducible rule-aligned analysis of real pull records.

2.3 Rare-Event Inference for Gacha Outcomes

A central methodological issue in account-level gacha analysis is low event count for high-rarity outcomes. Under these conditions, exact statistical methods are generally preferred over asymptotic approximations. Current practice supports exact binomial testing and exact confidence intervals for low-frequency outcomes [5], [6], improving inferential reliability when sample sizes are modest.

Another key issue is null-model selection. In gacha contexts, conclusions can differ depending on whether observed rates are tested against base advertised probabilities or pity-adjusted effective references. Transparent reporting of null assumptions is therefore essential. This motivates the present study's use of explicit two-sided tests under multiple reference rates, rather than relying only on descriptive "luck" indicators.

3 Methods

3.1 Participant and Data Collection

This study used a single-subject observational design (n = 1) based on the researcher's own Wuthering Waves account records. The participant is a college student enrolled in the Bachelor of Computer Science program with specialization in Machine Learning at National University - Philippines. Data were collected from exported convene history logs and processed as de-identified pull-level records for academic analysis, with voluntary participation and informed consent.

The analyzed dataset contains 582 Featured Resonator pulls. Based on pull timestamps, the observation window spans 97 calendar days (from 2025-11-08 to 2026-02-12, inclusive), with activity concentrated across 24 distinct pull days. Each record includes banner type, rarity (qualityLevel), item name, and timestamp; records were chronologically ordered and transformed into mechanics-aware features (pity intervals, 50/50 outcomes, guarantee states, and temporal indicators).

3.2 Feature Extraction

The pull-history dataset was then processed, and a set of mechanics-aware features was extracted from banner records, rarity outcomes, and timestamps following the analytical pipeline described in this study. For this section, we prioritized features directly related to gacha probability dynamics, pity progression, and temporal pull behavior. Features were selected based on their relevance to the study hypotheses on rare-event frequency, pity-threshold behavior, and 50/50 guarantee transitions. Specifically, we extracted pull-order features (pull_number), pity-cycle features (pity, pity_at_5star), outcome-state indicators (is_5050_win, is_5050_loss, is_guaranteed_win, is_featured_win), temporal markers (hour_of_pull, day_of_week), and summary metrics (total pulls, astrites spent, 5-star/4-star totals, and pull ratios). These engineered variables form the basis of the descriptive statistics, inferential tests, streak visualizations, and simulation analyses in later sections.

Table 1. List of extracted features used in the gacha mechanics analysis.

Feature Category	Feature
Sequence	pull_number
Pity Mechanics	pity, pity_at_5star, pity_counter
5-star Outcome State	is_featured_win, is_5050_win, is_5050_loss, is_guaranteed_win
Temporal	hour_of_pull, day_of_week
Rolling/Local Dynamics	rolling_4star_count, pulls_since_last_5
Categorical Encodings	banner_encoded, resource_type_encoded
Core KPIs	total pulls, total astrites, total 5-star pulls, total 4-star pulls
Luck Metrics	average 5-star pity, 5-star pull ratio, 50/50 win rate, 4-star pity metrics

To improve reproducibility, featured 5-star labeling followed a deterministic rule hierarchy implemented in the experiment notebook. First, when banner text in cardPoolType contained a recognized featured banner key, the pull was matched using explicit banner-to-featured mappings. Second, if explicit calendar metadata is available, timestamp-based banner assignment can be applied. Third, when banner metadata is incomplete, a fallback rule classifies a 5-star as featured if its name is outside the standard-pool set (Encore, Calcharo, Jianxin, Lingyang, Verina), and stores the applied source tag for auditability (card_pool_text, calendar, or standard_pool_fallback).

3.3 Operational Definitons

Table 2 provides a detailed breakdown of the variables used in this study, including their operational definitions and corresponding measurement procedures. These operational definitions standardize pull-history records from the convene log for mechanics-aware analysis and support consistent interpretation of pity behavior, 50/50 outcomes, and temporal pull dynamics.

Table 2. Definition of Variables

Feature Category	Feature

Pull Number	Sequential index assigned to each pull after chronological sorting (oldest to newest).
Time	Timestamp of each pull event as recorded in the exported convene history.
Card Pool Type	Banner identifier used to determine whether a pull belongs to the Featured Resonator scope.
Quality Level	Outcome rarity tier per pull (3, 4, or 5 stars).
Name	Pulled item identity (resonator or weapon) recorded for each event.
Featured Resonator Dataset	Analysis subset containing Featured Resonator pulls, excluding weapon-banner entries.
Pity	Number of featured pulls since the last 5-star; increments per pull and resets after each 5-star event.
Pulls-to-5-star (pity_at_5star)	Pity-cycle length observed at each 5-star event; used as the primary rare-event interval metric.
50/50 Win	Indicator that a featured 5-star was obtained under non-guaranteed conditions.
50/50 Loss	Indicator that a non-featured 5-star occurred under non-guaranteed conditions.
Guaranteed Win	Indicator that a featured 5-star occurred immediately after a prior 50/50 loss.
Featured Win	Indicator that the 5-star result matches the active featured target within the banner interval.
5-star Pull Ratio	Percentage of featured pulls resulting in 5-star outcomes.
4-star Pull Ratio	Percentage of featured pulls resulting in 4-star outcomes.
Average Pity	Mean pulls required to obtain a 5-star across completed pity cycles.
Astrites Spent	Estimated currency cost computed as total pulls × 160 astrites.
Hour of Pull	Hour extracted from timestamp for intraday temporal analysis.
Day of Week	Weekday extracted from timestamp for weekly temporal analysis.
Rolling 4-star Count	Local-window count of recent 4-star outcomes used for streak and density behavior checks.

3.4 Data Cleaning and Preprocessing

Raw convene-history data were imported from wuwa_gacha_records.csv (N = 582 pull records) and validated against the required schema (cardPoolType, resourceId, qualityLevel, resourceType, name, count, time), with no missing required columns detected. Timestamps were parsed using day-first datetime format (DD/MM/YYYY HH:MM), yielding a 100% parse success rate (582/582, 0 failures). Records were then sorted in ascending chronological order and assigned a sequential pull index (pull_number) to preserve event-order integrity for pity-cycle and streak analyses.

After normalization, records were filtered to the Featured Resonator analytical scope by retaining Resonator entries and excluding weapon-banner entries; all valid rows remained in scope (582 rows). Duplicate events were preserved when present because identical timestamps can occur in multi-pull sessions and still represent distinct outcomes. The final cleaned dataset spans 2025-11-08 20:24:00 to 2026-02-12 08:56:00 across 24 active pull days, with rarity composition of 504 3-star pulls, 69 4-star pulls, and 9 5-star pulls. These standardized records were used for downstream feature extraction, inferential testing, temporal visualization, and simulation-based robustness checks.

3.5 Exploratory Data Analysis (EDA)

Exploratory data analysis was conducted to profile rarity composition, pity-cycle behavior, and temporal pull activity before formal inference. This step was used to verify distributional shape, identify potential threshold effects, and confirm that exact rare-event methods were appropriate for downstream testing.

3.5.1 EDA Summary Statistics

The EDA stage computed descriptive statistics (mean, median, standard deviation, minimum, maximum, and distribution shape diagnostics) and generated visual diagnostics for rarity distribution, pulls-to-5-star intervals, and pity streak trajectories. These outputs were used to guide model assumptions and hypothesis-test selection.

3.6 Statistical Analysis Procedures

This study applies a mechanics-aware statistical framework to evaluate gacha outcomes in the Featured Resonator banner. Analyses were designed to align with the study hypotheses on rarity frequency, pity-cycle behavior, and guarantee-state transitions. Because 5-star events are rare in account-level data, the pipeline combines descriptive statistics with exact inferential methods and simulation-based robustness checks to avoid over-reliance on asymptotic assumptions.

Descriptive statistics were first computed to characterize pull distributions, including total pulls, rarity counts, pull ratios, average pity, dispersion, and distribution shape indicators (e.g., standard deviation, skewness, and kurtosis). Temporal summaries were also produced from timestamp-derived features to describe pull activity across time windows. These summaries establish the empirical baseline used for hypothesis testing and interpretation.

For inferential analysis, exact and proportion-based methods were prioritized. The observed 5-star rate was compared against reference expectations using exact binomial procedures, and uncertainty bounds were reported through exact confidence intervals

(Clopper-Pearson). For categorical outcome comparisons (e.g., 50/50 win-loss structure), contingency-based tests were used when applicable. All tests were evaluated at $\alpha=0.05$, with two-sided interpretation unless otherwise stated.

To evaluate model stability and practical significance under limited sample size, robustness checks were incorporated using Monte Carlo simulation. Simulated pull sequences under baseline and pity-adjusted assumptions were generated to contextualize observed totals and pity-cycle outcomes. This allowed the study to assess whether empirical results were plausible under alternative stochastic settings and to identify sensitivity to parameter assumptions.

All analyses were implemented in Python using reproducible notebook procedures, with deterministic preprocessing and explicit feature definitions to ensure traceability from raw pull logs to final statistical outputs.

3.7 Simulation and Robustness Checks

To complement exact inference, Monte Carlo simulation was used to evaluate whether observed 5-star totals were plausible under baseline and pity-aware assumptions. Synthetic pull sequences were generated at the same sequence length as observed data, and the resulting 5-star count distributions were compared with empirical outcomes.

Sensitivity checks varied key probability assumptions to test directional stability of conclusions. Robustness was inferred when interpretation remained consistent across plausible parameter ranges.

3.8 External/Internal Validation

Validation was implemented at both internal and benchmark levels. Internally, pity-cycle consistency was cross-checked by aligning reconstructed 5-star pity anchors with sequence-based features and visual outputs. Externally, aggregate metrics (total pulls, rarity totals, pull ratio, average pity, and pity sequence ordering) were compared against reference tracker snapshot values in the notebook's validation block.

This dual validation step ensured consistency between engineered features, reported metrics, and benchmark expectations.

3.9 Reproducibility and Ethics

All analyses were implemented in a reproducible Python notebook pipeline (pandas, numpy, matplotlib, seaborn, scipy). Preprocessing and feature-engineering rules were deterministic and explicitly documented. The study is limited to a single-subject account and therefore emphasizes internal validity over population generalization.

The dataset was analyzed in de-identified form, and usage was limited to academic research with participant consent.

4 Results

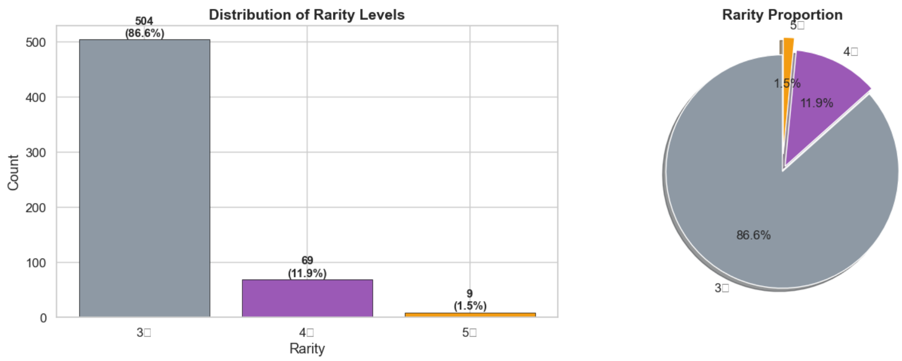
4.1 Descriptive Analysis

Table 3 Reports the core pull metrics for the Featured Resonator dataset. The profile shows a strongly imbalanced rarity structure, with low-rarity outcomes dominating and five-star outcomes remaining sparse. This distributional baseline motivates exact-inference procedures and mechanics-aware interpretation in subsequent analyses.

Table 3. Core Pull Metrics

Metric	Value
Total Pulls	582
Total Astrites Spent (160 per pull)	93, 120
Total 5 Star	9
Total 4 Star	69
5 Star Ratio	1.55%
4 Star Ratio	11.86%

Figure 1. Shows the rarity distribution is heavily concentrated in 3-star outcomes, which account for 504 of 582 pulls (86.6%). In contrast, 4-star pulls occur at a much lower frequency (69 pulls, 11.9%), while 5-star pulls are rare (9 pulls, 1.5%). This pattern confirms a highly imbalanced, rare-event structure that is consistent with pity-based gacha systems and supports the use of exact and simulation-based analyses.



4.2 Pity-Cycle Outcomes

Figure 2. Using the validated pity reconstruction, the observed pulls-to-5-star sequence (oldest to newest) is: 67, 72, 70, 75, 71, 68, 69, 70, 19. The average pity is 64.56 pulls,

with median 70, standard deviation 16.26, minimum 19, and maximum 75. Most 5-star outcomes cluster in the late-cycle region (high-60s to mid-70s), while one early event (19) introduces a left-tail exception. No 5-star event exceeded the hard pity threshold of 80 pulls.

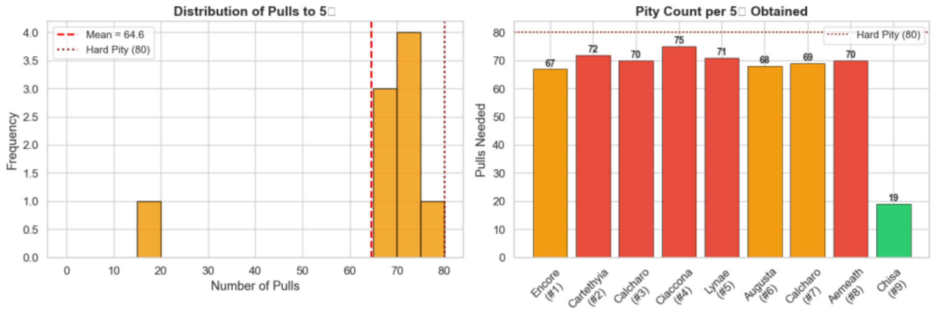


Table 4. Summarizes the reconstructed five-star pity intervals across completed cycles. The distribution centers near late-cycle values (median = 70), with one early-cycle outlier (minimum = 19), consistent with threshold-concentrated but stochastic outcomes.

Table 4. Five-Star Pity Summary

Statistic	**Value**
Pity Sequence (oldest to newest)	67, 72, 70, 75, 71, 68, 69, 70, 19
Mean	64.56
Median	70
Standard Deviation	16.26
Minimum	19
Maximum	75

Cycle-level diagnostics showed 9 completed 5-star cycles and 1 unfinished cycle at the end of the observation window, yielding a completion proxy score of 0.900. Using pity-index reach counts, only 1 win occurred before pity 60 across 491 reaches, indicating very low pre-threshold win frequency. A smoothed empirical pre-60 hazard estimate of approximately 0.0041 further supports late-cycle concentration behavior and is consistent with threshold-shaped pity dynamics rather than stationary outcome rates.

Although the primary focus is 5-star inference, 4-star behavior provides useful context on banner reward cadence. Under the same Featured Resonator scope (n = 582), the experiment observed 69 four-star pulls (11.86%), with average pulls-to-4-star of 7.42 under a reset rule triggered by any qualityLevel >= 4. The within-10-pull attainment rate for four-star outcomes was 78.26%, and current 4-star pity at dataset end was 1.

Table 5. Four-Star Summary Metrics

Metric	**Value**

Total Pulls	582
Total Astrites Spent (160 per pull)	93, 120
Total 4 Star	69
4 Star Ratio	11.86%

4.3 Streak and Temporal Pattern Analysis

Figure 3. Shows the pity-streak trajectory shows repeated linear build-up and reset behavior across cycles, with each reset aligned to a 5-star event. The corrected streak visualization confirms that annotated 5-star points are anchored to validated pity values; for example, Augusta is plotted at 68 pulls, consistent with the reconstructed sequence and below hard pity. This supports internal consistency between feature engineering (pity) and visualization output. Temporal pull behavior also shows clustered activity periods rather than uniform daily pulling, consistent with event-based or resource-based pulling sessions.

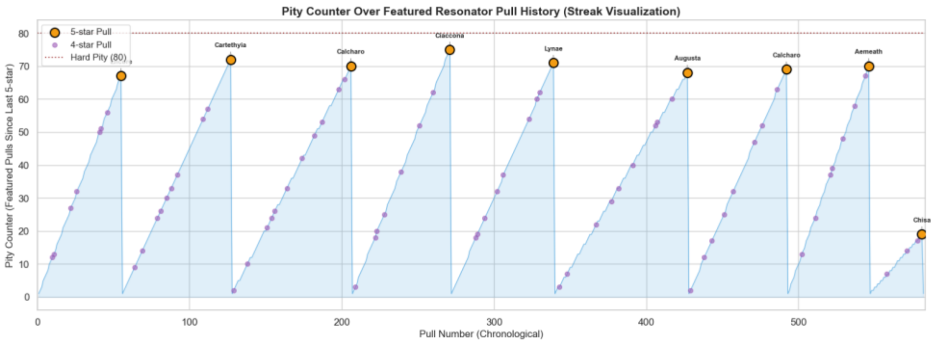
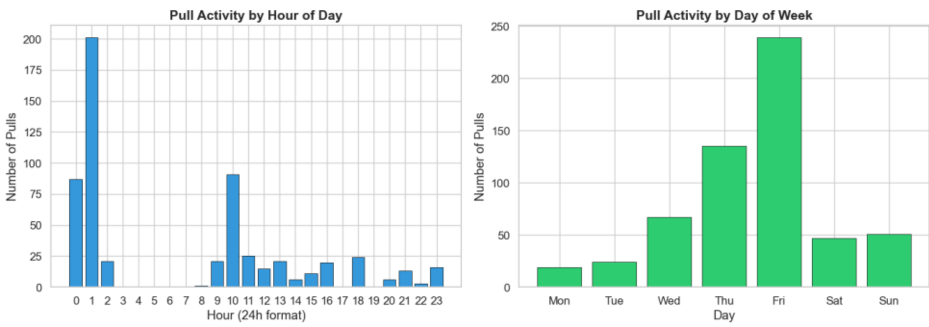


Figure 4. Shows temporal activity was concentrated in specific periods of the day and week. The highest-volume pull hour was 01:00 (n = 201), followed by 10:00 (n = 91) and 00:00 (n = 87), while day-level concentration was strongest on Friday (n = 239) and Thursday (n = 135). The observation window covered 24 distinct pull days across 97 calendar days, indicating burst-like pull behavior rather than uniform day-to-day activity.



Temporal activity was non-uniform within the observation window, with higher pull concentration in selected hour and weekday bins. These temporal outputs are interpreted as descriptive concentration patterns for this single account and are not used for causal claims.

4.4 Hypothesis Testing

Table 3. Maps each hypothesis to its corresponding analytical target, statistical method, and expected reporting output to ensure methodological traceability. Exact procedures are prioritized for rare-event inference, including exact binomial testing for observed 5-star frequency and Clopper-Pearson confidence intervals for rate precision. The table also links categorical outcome analysis (50/50 and guarantee states) and Monte Carlo simulation to robustness evaluation, allowing conclusions to be interpreted through both inferential significance and distribution-based plausibility.

Exact binomial testing was used to evaluate observed 5-star frequency under two explicit reference rates: a base-rate model ($p = 0.008$) and a pity-adjusted effective-rate model ($p = 0.016$). For the base-rate null, the observed 5-star rate (1.546%) yielded $p = 0.0566$, indicating failure to reject at $\alpha = 0.05$. For the pity-adjusted null, the observed rate yielded $p = 1.0000$, also failing to reject. The 95% Clopper-Pearson exact confidence interval for the observed 5-star rate was [0.709%, 2.915%].

For H3 (50/50 and guarantee-state structure), the observed five-star state outcomes were balanced in non-guaranteed events (3/6, 50/50 wins; 50.00%) and complete in recorded guaranteed conversions (3/3, 100.00%). Exact interval reporting in the experiment yielded a featured-win rate of 6/9 with 95% exact CI [29.93%, 92.51%], and guarantee conversion among losses with 95% exact CI [29.24%, 100.00%]. Given sparse event counts, these estimates are interpreted as exploratory uncertainty-bounded evidence rather than population-level rate claims.

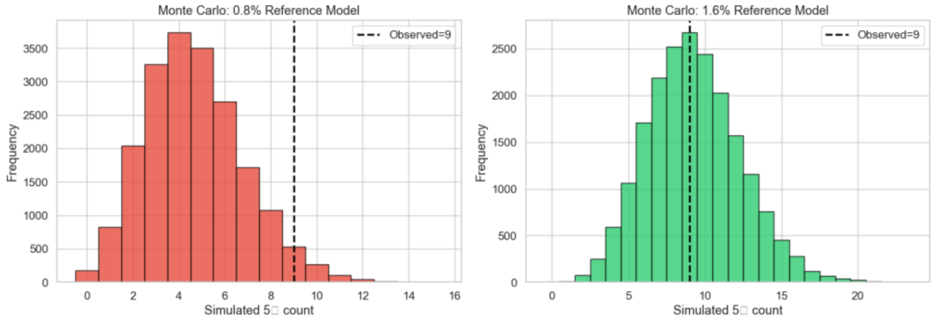
To preserve interpretability for sparse rare-event counts, exact inference is treated as the primary decision framework, while asymptotic outputs (if shown) are secondary for comparability only.

Hypothesis	Analytical Target	Statistical Procedure	Output
H1	Observed 5-star frequency vs reference rate	Exact binomial test	p-value, decision at $\alpha = 0.05$.
H2	Precision of 5-star rate estimate	Clopper-Pearson exact CI	lower/upper confidence bounds
H3	50/50 and guarantee-state outcome structure	Proportion/contingency analysis (exact or chi-square as applicable)	effect direction, p-value
H4	Stability of observed totals under stochastic assumptions	Monte Carlo simulation	empirical percentile/rank, simulation distribution

4.5 Simulation and Robustness Analysis

Figure 4. Presents Monte Carlo reference-rate simulation to contextualize the observed total of 9 five-star outcomes. Synthetic pull totals were generated at the same sample size ($n = 582$) under two fixed-rate models ($p = 0.008$ and $p = 0.016$). This procedure is interpreted as a pity-adjusted reference-rate check, not as a full state-dependent pity-process simulator.
Robustness was further evaluated through (1) rate-sensitivity tests over plausible effective rates (1.4%–1.8%) and (2) an incomplete-cycle exclusion check. Across these checks, the directional interpretation remained stable: observed outcomes are not extreme under the tested assumptions.

Sensitivity analysis across pity-adjusted reference rates (1.4%-1.8%) showed stable exact-binomial conclusions, with p-values ranging from 0.7221 to 1.0000 and small absolute rate differences (approximately -0.254 to +0.146 percentage points). Incomplete-cycle exclusion in the experiment removed 3 trailing pulls, producing 579 completed-only pulls, 9 five-stars, a 1.554% completed-only rate, and exact p-value 1.0000 versus $p = 0.016$. These checks support directional robustness under plausible reference assumptions.



5 Discussion

This section interprets the empirical findings of the Featured Resonator pull-history analysis in relation to gacha mechanics and rare-event modeling literature. It explains key observed patterns, evaluates how the results align with expectations from pity- and guarantee-based systems, and identifies practical implications for account-level statistical analysis. The section also acknowledges methodological constraints and outlines directions for future research that can improve external validity while preserving mechanics-aware rigor.

5.1 Interpretation of Results

The analysis reveals a strongly imbalanced rarity structure in which low-rarity outcomes dominate and 5-star events remain rare, consistent with the probabilistic design of gacha systems. Across 582 Featured Resonator pulls, only 9 were 5-star outcomes (1.55%), while most 5-star events clustered in late pity ranges (primarily high-60s to mid-70s), indicating threshold-driven reward behavior rather than uniform random occurrence. The reconstructed pity sequence (67, 72, 70, 75, 71, 68, 69, 70, 19) further shows mostly late-cycle completions with one early interruption, suggesting that while pity concentration is strong, stochastic variation still permits occasional early 5-star outcomes.

The corrected streak visualization supports this interpretation by showing repeated cycle build-up and reset dynamics, with 5-star anchors consistent with validated pity values and no event exceeding hard pity. In addition, 50/50 state outputs indicate balanced observed outcomes within this account window (3 wins, 3 losses among non-guaranteed 5-star states), with guarantee transitions functioning as expected under mechanics-aware labeling. Taken together, these findings show that account-level pull behavior is structured, state-dependent, and appropriately modeled using sequence-aware features rather than stationary Bernoulli assumptions alone.

5.2 Comparison to Related Work

These results are consistent with prior literature arguing that gacha outcomes should be modeled as state-dependent stochastic processes due to pity and guarantee mechanics. The observed late-cycle concentration of 5-star events supports threshold-effect interpretations reported in empirical gacha studies, while the presence of a rare early 5-star event reflects the expected residual randomness in non-deterministic reward systems. Methodologically, the use of exact inference and simulation for sparse high-rarity outcomes aligns with recommendations for rare-event analysis under limited sample conditions.

More broadly, this study contributes an account-level case demonstrating that mechanics-aware feature engineering (pity reconstruction, guarantee-state coding, and sequence-validated plotting) can produce internally coherent and interpretable results even without large-scale platform data. The findings therefore support the view that small but well-structured pull histories can yield meaningful insight when analysis is explicitly aligned with game mechanics.

5.3 Limitations

Numerous limitations ought to be emphasized:

- Single-subject scope: Findings are account-specific and not directly generalizable to broader player populations.
- Rare-event sample size: Only 9 observed 5-star outcomes constrain inferential precision and sensitivity to small changes.
- Assumption dependence: Some robustness outputs depend on reference-rate and pity-adjustment assumptions used in simulation.
- Behavioral context gap: Pull logs do not capture motivation, spending intent, event urgency, or psychological drivers.
- Temporal coverage: The analysis window captures one bounded period and may not represent longer-term behavioral shifts.
- System-specific rules: Mechanics interpretations are tied to the modeled banner logic and may vary with future system updates.

5.4 Recommendations and Future Work

Future work should prioritize multi-account replication to improve external validity while preserving the same mechanics-aware pipeline. Expanding data across players, banner periods, and patch cycles would enable stronger comparative inference on pity behavior, guarantee transitions, and temporal pull strategies. Longitudinal tracking over longer horizons can also distinguish persistent pull tendencies from short-term event effects.

Methodologically, future studies may integrate hierarchical or Bayesian rare-event models to better estimate uncertainty under low 5-star counts. Additional extensions include player-cluster analysis of pity-cycle behaviors, sensitivity profiling across alternative pity-rule assumptions, and mixed-method designs that combine pull logs with survey/interview context to connect statistical outcomes with player decision-making. These directions can convert account-level evidence into broader behavioral models of gacha interaction.

6 Conclusion

This study developed and applied a mechanics-aware statistical framework for analyzing Featured Resonator gacha behavior in Wuthering Waves using pull-history records. The results confirm a highly imbalanced rarity distribution and a clear pity-threshold structure, with most 5-star outcomes occurring in late-cycle ranges and no hard-pity exceedance. Exact inference and simulation-based checks indicate that the observed 5-star outcomes are statistically plausible under tested reference assumptions, while internal validation confirms consistency between reconstructed pity features and visualization outputs.

The study demonstrates that account-level gacha datasets, although limited in scope, can support rigorous and reproducible analysis when preprocessing, feature extraction, and inferential methods are explicitly aligned with game mechanics. By combining pity reconstruction, guarantee-state modeling, exact rare-event statistics, and Monte Carlo robustness checks, this work provides a practical template for student-led quantitative research in digital game systems. Future expansions using multi-account and longitudinal data can build on this framework to improve generalizability and deepen behavioral interpretation of gacha mechanics.

References

1. C. Chen and Z. Fang, Gacha Game Analysis and Design (2023), Proceedings of the ACM on Measurement and Analysis of Computing Systems 7 (1), Article 6.
2. K. Chen, Algorithms for calculating gacha probabilities (2025), available at: https://kylechen.net/writing/gacha-probability/
3. Prydwen Institute, Wuthering Waves Gacha System Guide (2025), available at: https://www.prydwen.gg/wuthering-waves/guides/gacha/
4. Mana Wiki, Wuthering Waves Convene Tracker (2025), available at: https://wuwa.mana.wiki/convene-tracker
5. S. M. Ross, Simulation (5th ed.) (2024), Academic Press.
6. A. Thavamuni et al. 2025. Inherent Addiction Mechanisms in Video Games' Gacha. Information 16, 10, 890. https://www.mdpi.com/2078-2489/16/10/890
7. M. S. El-Nasr, A. Drachen, and A. Canossa, Game Analytics: Maximizing the Value of Player Data (2023), Springer.
8. C. Ma, J. Shao, and P. Li. 2025. Monetization mechanisms in gacha games: The behavioral triad of pricing strategies, pity systems, and belief of luck. Entertainment Computing 55, 101044. https://www.sciencedirect.com/science/article/pii/S1875952125001247