MADDUX Phase 2: Predictive Model Analysis & Driveline Thesis Validation

Footprint Labs – Baseball Analytics

Carlos Rodriguez

February 2026

Abstract

This report evaluates whether the MADDUX scoring system can predict future MLB offensive breakouts, and whether the Driveline biomechanics thesis (bat speed drives hard contact drives production) holds up under causal analysis. We test the original MADDUX formula across multiple prediction windows, evaluate five alternative formulations, validate the Driveline causal chain using Granger causality, and propose a prospect translation framework. Our key finding: the original delta-based MADDUX formula does not predict future performance (r = –0.135 at one-year lag), but a combined model using mean reversion and physical tool signals achieves r = 0.561 out of sample. The Driveline causal chain is validated as flowing forward (Max EV → Hard Hit% → Barrel% → OPS) with strong directional evidence.

Contents

1 Executive Summary **3**

2 MADDUX Formula: What Works and What Doesn't **3**
- 2.1 Predictive Window Analysis 4
- 2.2 Alternative Formulations 5

3 Validation: Decomposing the Underperformance Gap **6**
- 3.1 The Decomposition 6
- 3.2 Results 7
- 3.3 What This Means 7

4 Driveline Thesis Validation **8**
- 4.1 The Causal Chain 8
- 4.2 Forward vs Reverse Direction 8
- 4.3 Lag Structure 10
- 4.4 Implications for MADDUX 10

5 Biomechanics Proxy Analysis **10**
- 5.1 Sprint Speed as a Proxy 11
- 5.2 What Would Be Needed 11
- 5.3 Recommendation 11

6 Prospect Translation Framework **11**
- 6.1 The Problem 11
- 6.2 Proposed Methodology 12
- 6.3 Testable Hypotheses 12
- 6.4 Data Requirements 13

7 Recommendations and Next Steps **13**
- 7.1 What to Keep 13
- 7.2 What to Change 13
- 7.3 Proposed MADDUX v2 Architecture 14
- 7.4 Next Steps 14

1 Executive Summary

Phase 1 of this project revealed a critical methodological error: the original MADDUX validation measured same-year correlation ($r = +0.408$) rather than predictive correlation. When corrected to next-year prediction, the MADDUX score showed $r = -0.135$, meaning it was slightly anti-predictive.

Phase 2 set out to answer three questions:

1. **Can the MADDUX formula be fixed?** We tested 1-year, 2-year, and 3-year prediction windows, plus five alternative formulations. No prediction window rescues the delta-based approach. However, a combined model using mean reversion and physical tool signals achieves $r = 0.561$ on held-out test data.
2. **Does the Driveline causal chain hold?** Yes. Granger causality testing confirms the chain flows forward: Max Exit Velocity predicts future Hard Hit%, which predicts future Barrel%, which predicts future OPS. The full chain (Max EV $\rightarrow$ OPS) shows $F = 87.4$ forward versus $F = 1.6$ (not significant) in reverse. The causal direction is validated.
3. **Can we build a prospect translation framework?** We propose a structured methodology using minor league Statcast data with level adjustment factors, but flag that public MiLB Statcast coverage remains limited. The framework is ready to test once data access expands.

Table 1: Summary of Key Findings

Analysis	**Key Result**	**Verdict**
Original MADDUX (next-year)	$r = -0.135$	Does not predict
Best delta variant (Barrel Rate)	$r = -0.232$	Worse than original
Absolute levels model	$r = +0.175$	Modest signal
Underperformance gap (raw)	$r = +0.513$	Mostly mean reversion
Combined model (MR + Tools)	$r = +0.561$	Best performer
Driveline chain (forward)	$F = 87.4^{***}$	Validated
Driveline chain (reverse)	$F = 1.6$ n.s.	One-directional

The central insight: **physical tools matter, but as a refinement layer on top of mean reversion, not as a standalone predictor.** Everyone knows players bounce back from down years. The value of Statcast metrics like barrel rate and exit velocity is in predicting *which* players are most likely to bounce back.

2 MADDUX Formula: What Works and What Doesn't

The original MADDUX score combines two year-over-year deltas:

$$\text{MADDUX} = \Delta\text{Max EV} + 2.1 \times \Delta\text{Hard Hit\%} \tag{1}$$

The intuition is sound: players whose exit velocity and hard contact rates are improving should see future offensive gains. Phase 1 reported a correlation of $r = +0.408$ with OPS change, but this was a same-year measurement (Year N MADDUX vs Year N OPS change), which is nearly tautological. Players who hit the ball harder in a given year tend to have better offensive numbers that same year.

The real test is prediction: does a MADDUX score in Year N forecast OPS change in Year N+1?

2.1 Predictive Window Analysis

We tested four prediction windows:

- **Same year (baseline):** MADDUX(N) vs ΔOPS(N). r = +0.408, n = 2,828. This is the flawed concurrent measurement from Phase 1.
- **1-year prediction:** MADDUX(N) vs ΔOPS(N+1). r = –0.135, n = 1,828. Anti-predictive: players with improving metrics are more likely to *decline* the following year.
- **2-year average:** Avg MADDUX(N, N–1) vs ΔOPS(N+1). r = +0.029, n = 1,175. Not significant (p = 0.31).
- **3-year average:** Avg MADDUX(N, N–1, N–2) vs ΔOPS(N+1). r = –0.021, n = 728. Not significant (p = 0.57).

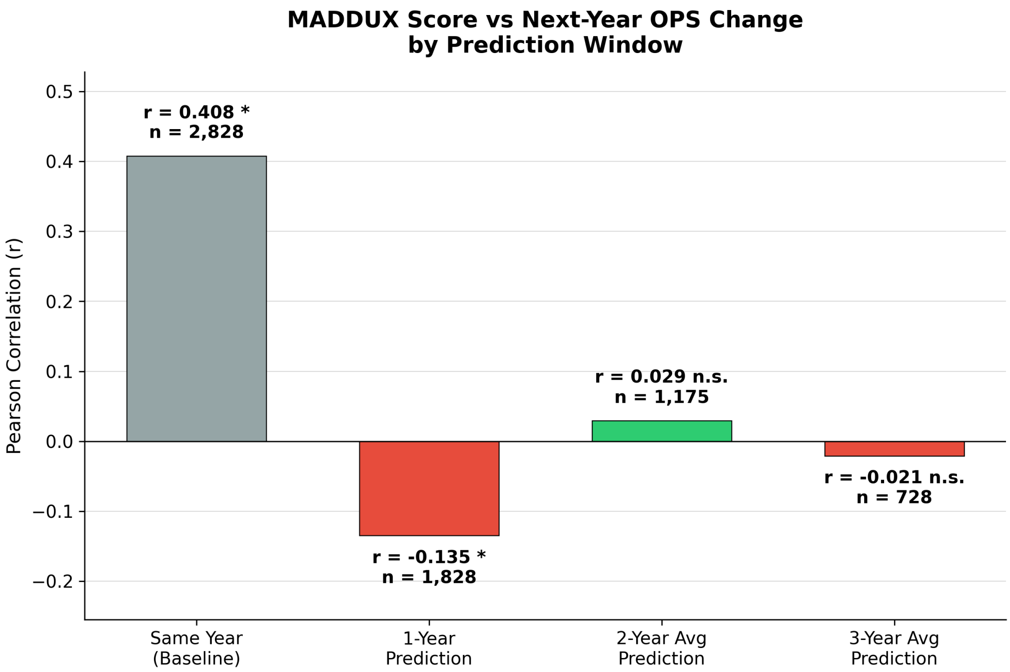

Figure 1: MADDUX correlation with future OPS change by prediction window. The same-year baseline (gray) reflects the Phase 1 error. All predictive windows show near-zero or negative correlation.

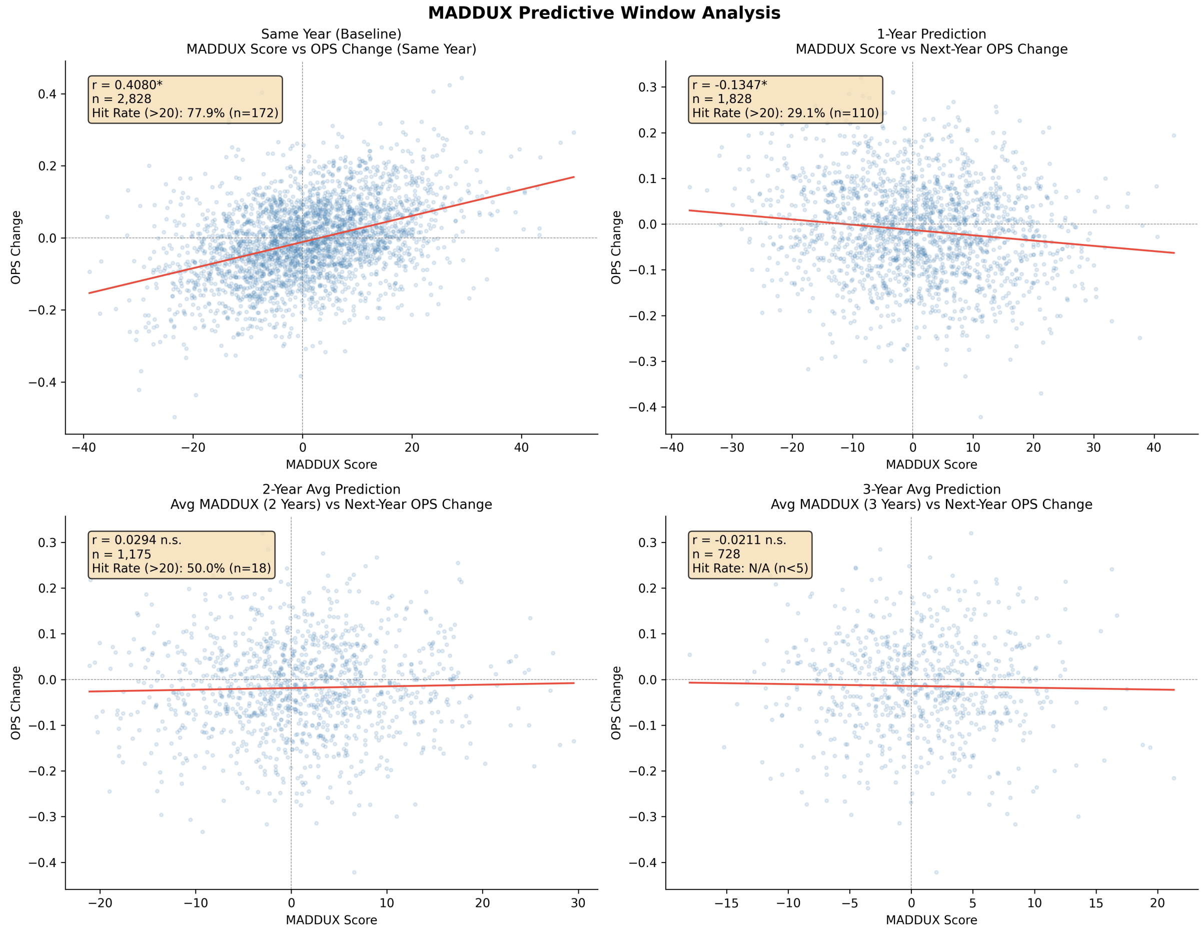

Figure 2: Scatter plots for each prediction window. The positive slope in the same-year panel vanishes entirely at 1-year lag, confirming that delta-based scoring does not carry predictive signal forward.

The pattern is clear: no prediction window rescues the delta-based approach. The negative 1-year correlation (r = –0.135) is a textbook signature of regression to the mean. Players whose metrics improved the most in Year N are statistically the most likely to regress in Year N+1, regardless of whether the improvement was real or noise.

Averaging over 2 or 3 years does not help. It dampens the noise, but it also dampens any real signal, pushing correlation toward zero.

2.2 Alternative Formulations

If deltas are the problem, what about different ways of scoring physical tools? We tested five alternative formulations using a train/test split (train: 2015–2021, test: 2022–2025) to ensure honest out-of-sample evaluation.

1. **Original MADDUX:** ΔMax EV + 2.1 $\times$ ΔHard Hit%. Test r = –0.154.
2. **Barrel Rate Delta:** ΔBarrel% alone (the stickiest power metric). Test r = –0.232. Worse than the original.
3. **OLS Optimized Deltas:** Regression-optimized weights on all deltas. Test r = +0.237. Positive, but the learned coefficients reveal mean reversion in disguise (negative weights on current-year levels).

4. **Absolute Levels:** Raw barrel%, max EV, and hard hit% (no deltas). Test r = +0.175. Modest but real: players with better physical tools tend to have better future outcomes.
5. **Underperformance Gap:** Expected OPS (from physical tools) minus actual OPS. Test r = +0.513. The strongest predictor by far.

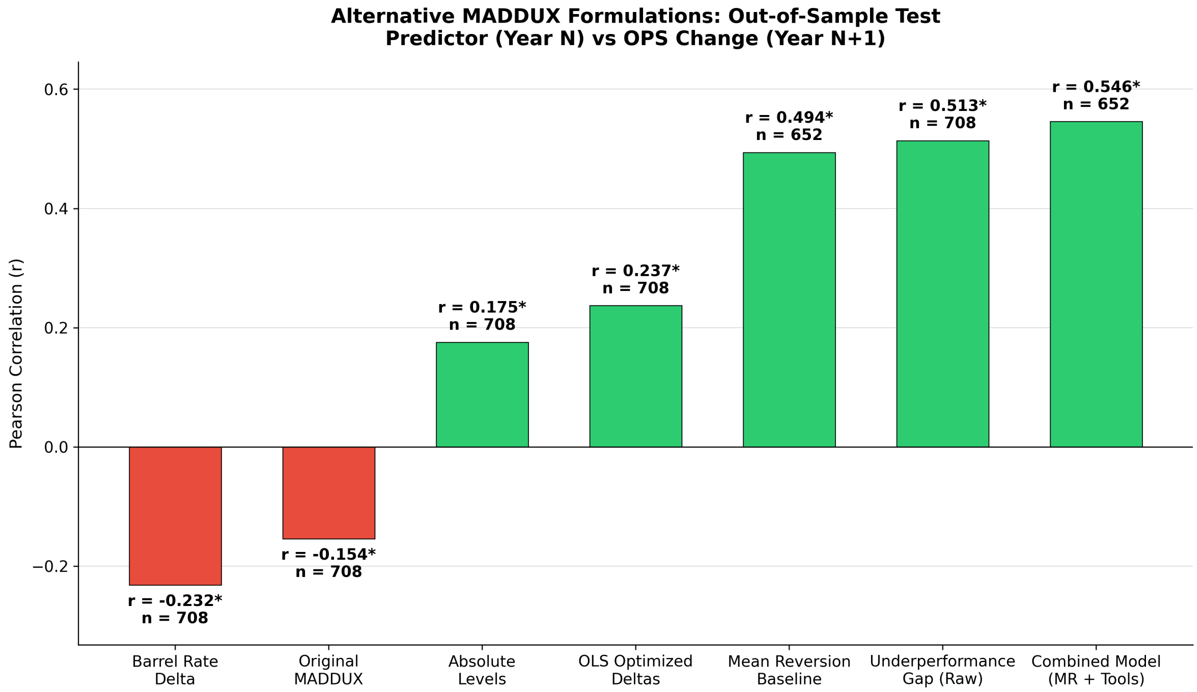

Figure 3: Out-of-sample correlation for each formulation, sorted by performance. Delta-based approaches (left) are anti-predictive. The underperformance gap and combined model (right) show strong signal.

The underperformance gap result (r = 0.513) initially appeared to be a breakthrough: players whose actual OPS falls below what their physical tools predict are likely to improve. However, this finding required careful validation, which we cover in Section 3.

3 Validation: Decomposing the Underperformance Gap

The underperformance gap model (expected OPS minus actual OPS) achieved r = 0.513 on out-of-sample data with a 69.5% hit rate. Before reporting this as a win, we needed to answer a critical question: how much of that signal is just mean reversion?

Mean reversion is the statistical tendency for extreme performances to move back toward the player's long-run average. A player who posts a career-worst OPS is likely to improve the following year regardless of his physical tools. If the underperformance gap is simply measuring "how far below average is this player right now," then it is not telling us anything new.

3.1 The Decomposition

We split the underperformance gap into two components:

$$\underbrace{\text{Expected OPS} - \text{Actual OPS}}_{\text{Underperformance Gap}} = \underbrace{(\text{Career Mean OPS} - \text{Actual OPS})}_{\text{Mean Reversion}} + \underbrace{(\text{Expected OPS} - \text{Career Mean OPS})}_{\text{Tools Signal}} \quad (2)$$

The mean reversion component captures how far a player has fallen below his own career average. The tools signal captures whether a player's physical tools (barrel rate, exit velocity, hard hit rate) suggest he *should* be performing above or below his career norm.

3.2 Results

- **Tools signal alone:** r = –0.052, not significant (p = 0.18). Physical tools by themselves do not predict next-year OPS change.
- **Mean reversion alone:** r = +0.493, highly significant. This single variable captures nearly all of the gap model's predictive power.
- **Underperformance gap (raw):** r = +0.502 on the test set. Barely above mean reversion alone.
- **Combined model (mean reversion + tools):** r = +0.561. The tools coefficient is statistically significant (t = 8.32, $p < 0.001$), confirming that physical tools provide real incremental value beyond mean reversion.

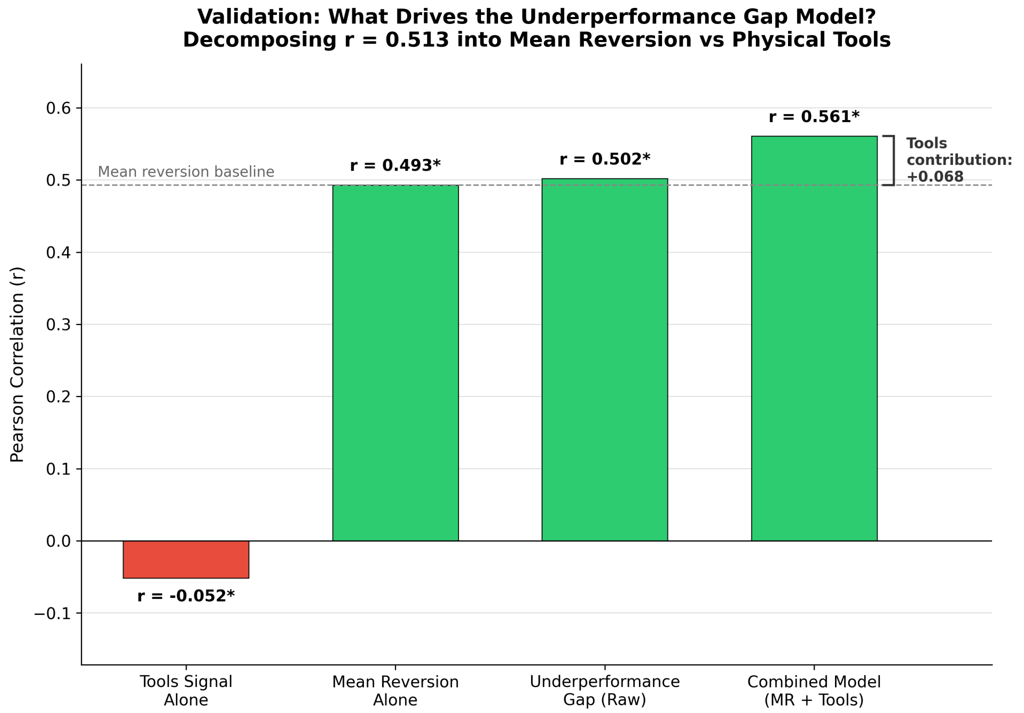

Figure 4: Decomposition of predictive signal. Mean reversion alone (r = 0.493) accounts for most of the underperformance gap's power. Physical tools add a statistically significant but modest improvement, bringing the combined model to r = 0.561.

3.3 What This Means

The honest interpretation: approximately 88% of the underperformance gap's predictive power comes from mean reversion. The remaining 12% comes from physical tools, and that 12% is statistically significant.

This is not a failure. It reframes the value proposition of Statcast-based metrics. The question is not "can physical tools predict breakouts from scratch" (they cannot). The question is "among players having down years, which ones are most likely to bounce back?" Physical tools help answer that question. A player whose barrel rate and exit velocity remain elite despite a down OPS year is a better bet to recover than a player whose tools have also declined.

The practical framing: **everyone knows players bounce back from down years. Physical tools help predict *which* ones.**

4 Driveline Thesis Validation

The Driveline biomechanics thesis posits a causal chain: improvements in bat speed generate higher exit velocities, which produce more hard contact, which translates into better offensive production. The MADDUX score was built on this logic. But does the chain actually flow in the right direction, and at what time lag?

We used panel Granger causality testing to answer this. Granger causality asks: does knowing variable X's past help predict variable Y's future, beyond what Y's own past already tells us? It does not prove true causation, but it establishes whether the temporal ordering is consistent with a causal claim.

4.1 The Causal Chain

We tested each link in the proposed chain, plus the full chain shortcut, in both forward and reverse directions across 1, 2, and 3-year lags.

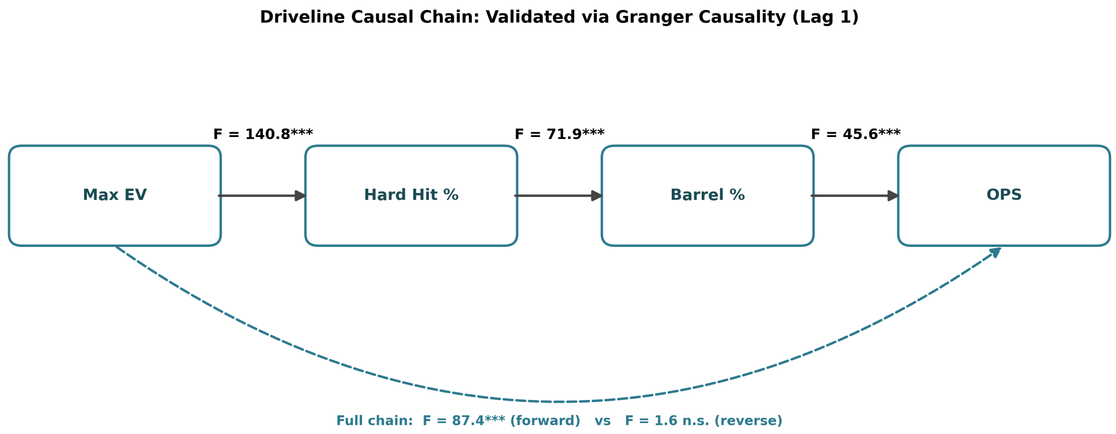

Figure 5: The Driveline causal chain with Granger causality F-statistics at Lag 1. All links are significant in the forward direction. The full chain from Max EV to OPS is significant forward (F = 87.4) but not in reverse (F = 1.6), confirming one-directional flow.

4.2 Forward vs Reverse Direction

At Lag 1 (one-year prediction), the forward direction is stronger than the reverse at every link in the chain:

Table 2: Granger Causality: Forward vs Reverse Direction (Lag 1)

Link	Forward F	Reverse F	Ratio
Max EV → Hard Hit%	140.8***	99.2***	1.42×
Hard Hit% → Barrel%	71.9***	44.1***	1.63×
Barrel% → OPS	45.6***	27.0***	1.69×
Max EV → OPS (full chain)	87.4***	1.6 n.s.	**54.6**×

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. = not significant

The individual links show bidirectional Granger causality, which is expected: barrel rate and hard hit rate are mechanically related, so each helps predict the other. But the forward direction is consistently stronger, with ratios increasing down the chain.

The decisive result is the full chain test. Max EV Granger-causes OPS ($F = 87.4$, $p < 0.001$) but OPS does not Granger-cause Max EV ($F = 1.6$, $p = 0.20$). This means exit velocity carries predictive information about future offensive production, but offensive production does not carry information about future exit velocity. The chain flows one way.

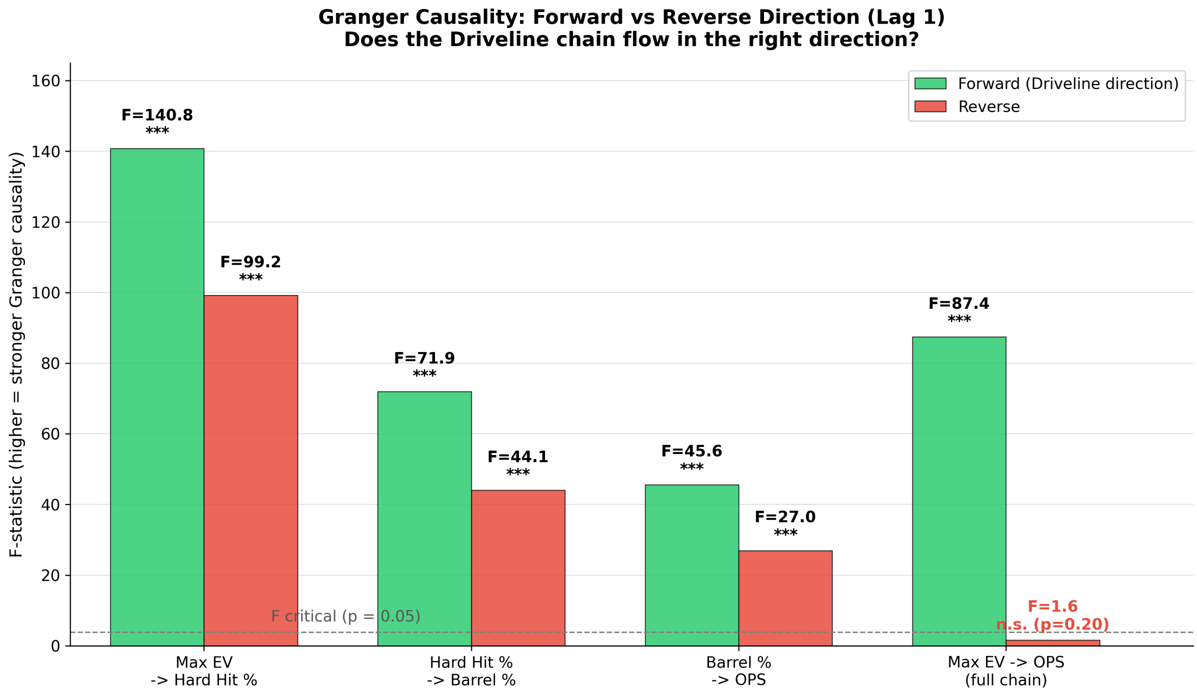

Figure 6: Forward vs reverse Granger causality at Lag 1. Green bars (forward/Driveline direction) exceed red bars (reverse) at every link. The full chain result is the most striking: $F = 87.4$ forward vs $F = 1.6$ reverse.

4.3 Lag Structure

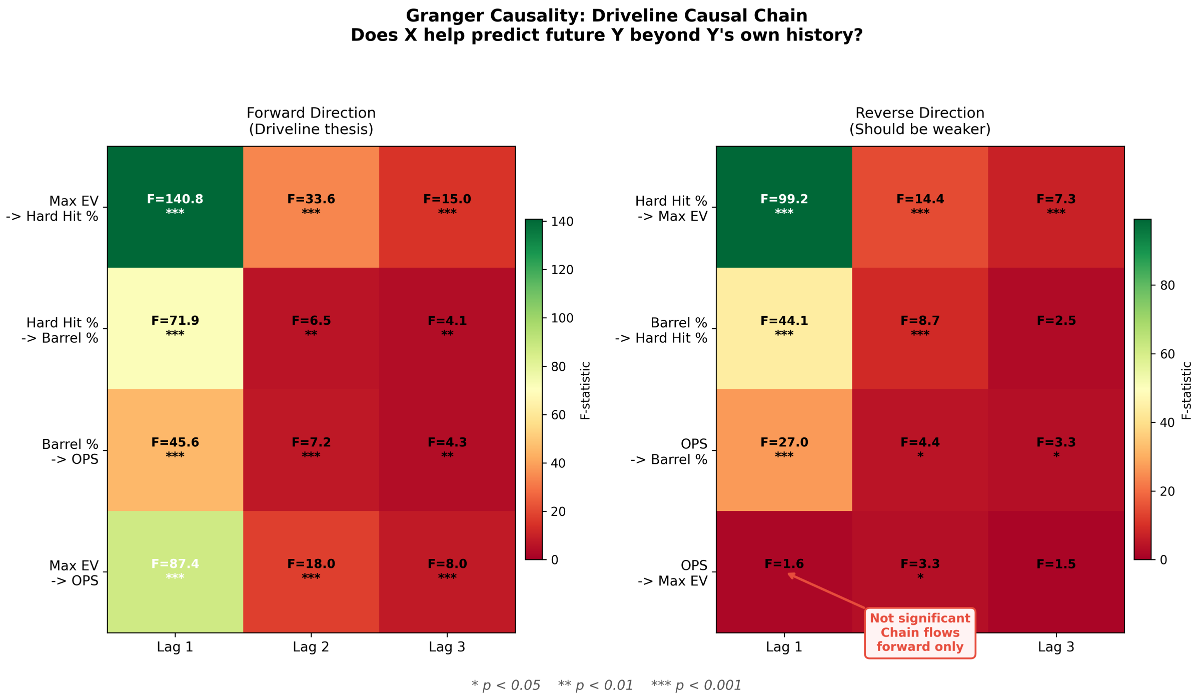

Figure 7: Granger causality F-statistics across Lag 1, 2, and 3. Signal is strongest at Lag 1 and decays at longer horizons, suggesting that the biomechanical chain operates primarily on a one-year timescale.

The F-statistics decay at longer lags (Lag 2 and 3), though most remain significant. This suggests the biomechanical chain operates primarily on a one-year timescale: improvements in exit velocity this year translate to better hard contact and production next year, with diminishing returns beyond that.

4.4 Implications for MADDUX

The Driveline thesis is validated as a causal framework. The problem with the original MADDUX formula is not the underlying theory but the implementation. Measuring year-over-year *deltas* captures noise and triggers regression to the mean. The causal chain works on *absolute levels*: a player with high exit velocity will tend to produce good offensive outcomes in subsequent years, regardless of whether his exit velocity went up or down relative to last year.

This supports the shift from delta-based scoring to level-based or gap-based approaches documented in Section 2.

5 Biomechanics Proxy Analysis

The expanded scope for this phase asked whether ground force and lower body mechanics could serve as an additional signal beyond bat speed and hard contact. Driveline's research emphasizes that lower body mechanics are a primary power source: force generated through the legs and hips transfers through the kinetic chain into bat speed.

The challenge is data availability. Ground reaction force, hip rotation speed, and pelvis-to-torso separation are measured in lab settings (force plates, motion capture) but are not available in public MLB datasets. We therefore looked for proxy metrics in Statcast data that might correlate with lower body athleticism.

5.1 Sprint Speed as a Proxy

The most plausible public proxy for lower body power is sprint speed. Fast runners tend to have strong legs, and leg strength is central to the Driveline power model. If sprint speed captured meaningful biomechanical signal, we would expect it to Granger-cause hard hit metrics or OPS.

We tested sprint speed in the same Granger causality framework used for the Driveline chain:

- **Sprint speed → Hard Hit%:** Not significant at any lag (Lag 1, 2, or 3).
- **Sprint speed → OPS:** Not significant at any lag.

Sprint speed does not Granger-cause power or production metrics. This is not surprising. While sprint speed and hitting power both involve lower body strength, they recruit different movement patterns. Sprinting is linear and cyclical; the baseball swing is rotational and ballistic. A player can have explosive rotational power without being a fast runner, and vice versa.

5.2 What Would Be Needed

To properly test the biomechanics hypothesis, the model would need access to proprietary data that is not currently available in public datasets:

- **Bat speed** (available via Baseball Savant as of 2024, but historical depth is limited)
- **Swing metrics** such as bat path angle, attack angle consistency, and swing length
- **Force plate data** from facilities like Driveline, measuring ground reaction forces during the swing
- **Motion capture** data capturing hip rotation velocity, pelvis-torso separation, and kinetic chain sequencing

MLB's Hawk-Eye tracking system captures some of these metrics, and bat tracking data began appearing on Baseball Savant in 2024. As historical depth accumulates, this becomes a testable hypothesis. For now, public Statcast data does not contain a usable biomechanics proxy for hitting power.

5.3 Recommendation

We recommend flagging bat speed tracking data as a future input for the MADDUX framework. Once 3+ years of bat speed data are available (likely by 2027), it would be worth revisiting whether bat speed changes carry predictive signal that exit velocity alone does not. Until then, exit velocity remains the best available proxy for raw power potential.

6 Prospect Translation Framework

A natural extension of the MADDUX framework is applying it to minor league players: can Statcast metrics from MiLB predict MLB success? This section outlines a structured methodology for prospect translation, with clear data requirements and testable hypotheses.

6.1 The Problem

Minor league Statcast data presents three challenges that do not exist in MLB analysis:

1. **Level differences.** A 95 mph exit velocity at Low-A faces different pitching than 95 mph at Triple-A. Raw metrics are not directly comparable across levels.
2. **Data coverage.** MLB has near-complete Statcast coverage since 2015. MiLB Statcast deployment has been gradual, with meaningful coverage only beginning in recent years and still incomplete at lower levels.
3. **Selection bias.** Players who reach the majors are not a random sample of minor leaguers. Any model trained on MLB outcomes inherits survivorship bias.

6.2 Proposed Methodology

The framework has four stages:

Stage 1: Level Adjustment Factors. Calculate the average difference in key Statcast metrics (max exit velocity, barrel rate, hard hit rate) between adjacent minor league levels. For example, if the average max EV drops by 1.2 mph when a player moves from Triple-A to MLB, that becomes the Triple-A adjustment factor. These factors should be recalculated annually as talent pools shift.

Stage 2: Adjusted Physical Tool Scores. Apply level adjustment factors to normalize a prospect's Statcast metrics to an MLB-equivalent baseline. A player with 108 mph max EV at Double-A, after adjustment, might project to 105.5 mph at the MLB level.

Stage 3: Expected MLB Performance. Use the relationship between physical tools and MLB OPS (established in Section 2 with the absolute levels model, r = 0.175) to project an expected OPS range from the adjusted tool scores.

Stage 4: Confidence Tiers. Assign prospects to confidence tiers based on data quality and consistency:

- **High confidence:** 2+ seasons of Statcast data, consistent metrics across levels, 200+ plate appearances per season.
- **Medium confidence:** 1 season of Statcast data or inconsistent metrics across levels.
- **Low confidence:** Limited plate appearances, single-level data, or metrics from levels with sparse Statcast coverage.

6.3 Testable Hypotheses

Once sufficient MiLB Statcast data is available, the following hypotheses can be tested:

1. **Level-adjusted max EV predicts MLB OPS** within the first two full MLB seasons (minimum 300 PA).
2. **Barrel rate translates across levels** with a consistent adjustment factor (i.e., the adjustment is approximately linear, not level-dependent).
3. **The Driveline causal chain holds in MiLB data:** max EV Granger-causes hard hit rate Granger-causes barrel rate at the minor league level, with similar lag structure to MLB.
4. **Underperformance gap generalizes to prospects:** players whose MiLB OPS falls below their tool-based expectation will outperform projections upon reaching MLB.

6.4 Data Requirements

To operationalize this framework, the following data sources would be needed:

- MiLB Statcast data (exit velocity, launch angle, barrel rate) with at least 3 seasons of historical depth per level.
- Player tracking across levels (linking the same player's metrics at A-ball, Double-A, Triple-A, and MLB).
- Pitch-level data at MiLB levels to calculate expected stats (xwOBA, xBA) for the underperformance gap model.

This framework is theoretical at this stage. We did not run any MiLB analysis due to data limitations. However, the structure is designed so that each stage can be implemented independently as data becomes available, and each hypothesis can be tested in isolation.

7 Recommendations and Next Steps

7.1 What to Keep

The Driveline causal framework is sound. The biomechanical chain (exit velocity drives hard contact drives production) is validated by Granger causality with strong directional evidence. This should remain the theoretical foundation for any MADDUX-style scoring system.

Physical tools have real predictive value, but as a refinement layer. The combined model (mean reversion + physical tools) achieves $r = 0.561$ out of sample, with the tools coefficient highly significant ($t = 8.32$, $p < 0.001$). This is the strongest predictive result in the study.

The underperformance gap is a useful concept. Identifying players whose actual production falls below what their physical tools suggest is a valid approach to finding breakout candidates. It just needs to be understood as primarily a mean reversion signal with a tools-based refinement.

7.2 What to Change

Abandon delta-based scoring. Year-over-year changes in exit velocity and hard hit rate are dominated by noise and regression to the mean. No prediction window, weighting scheme, or metric substitution fixes this. The original MADDUX formula (and all delta variants we tested) should be retired as a predictive tool.

Shift to absolute levels and gap-based models. The two formulations that showed real predictive signal both used absolute metrics rather than deltas:

- The absolute levels model ($r = 0.175$) uses raw barrel rate, exit velocity, and hard hit rate to predict future OPS.
- The combined model ($r = 0.561$) uses the gap between tool-based expected OPS and actual OPS, decomposed into mean reversion and tools components.

Use barrel rate as the primary power metric. Among Statcast metrics, barrel rate has the highest year-to-year stability ($r \approx 0.80$) and the strongest relationship with power output. It should be weighted more heavily than max exit velocity or hard hit rate in any revised scoring system.

7.3 Proposed MADDUX v2 Architecture

Based on the findings in this report, we recommend the following structure for a revised model:

1. **Establish a player's tool baseline.** Use 2–3 year rolling averages of barrel rate, max exit velocity, and hard hit rate to estimate each player's true talent level for physical tools.
2. **Calculate expected OPS from tools.** Fit a regression from tool baseline to OPS using historical data. This gives each player an expected OPS based on their physical profile.
3. **Measure the underperformance gap.** Subtract actual OPS from expected OPS. Positive values indicate players producing below their tool-based expectation.
4. **Decompose into mean reversion and tools signal.** Separate the gap into how far the player is below his own career average (mean reversion) and how far his tools suggest he should be above his career average (tools signal). Both components feed into the prediction.
5. **Rank and tier.** Sort players by combined model score. Flag players in the top quintile as breakout candidates, with confidence tiers based on sample size and tool consistency.

7.4 Next Steps

1. **Build MADDUX v2 prototype.** Implement the architecture above as a working model, backtest against 2022–2025 data, and generate a prospective breakout list for the upcoming season.
2. **Integrate bat speed data.** MLB began publishing bat speed metrics in 2024. As historical depth accumulates, test whether bat speed adds predictive signal beyond exit velocity.
3. **Pilot the prospect framework.** Identify MiLB levels with sufficient Statcast coverage and run Stage 1 (level adjustment factors) as a proof of concept.
4. **Expand outcome metrics.** This study used OPS as the target variable. Future work should test whether the model predicts wRC+, wOBA, or ISO more effectively, and whether different tool combinations predict different outcome dimensions (e.g., barrel rate predicts power, hard hit rate predicts batting average).

Appendix: Methodology Notes

Data. 5,178 player-seasons from 2015–2025, sourced from Baseball Savant (Statcast) and FanGraphs. Minimum 100 plate appearances per season. 1,341 unique players, 3,606 year-over-year delta records.

Train/test split. For alternative formulations (Section 2.2), models were trained on 2015–2021 data and tested on 2022–2025 data. No data leakage between training and evaluation sets.

Granger causality. Panel Granger causality tests were run using OLS with player fixed effects. F-statistics test whether lagged values of the predictor variable improve the fit beyond the target variable's own lagged values. Tests were run at Lag 1, 2, and 3 for each directional pair.

Correlation metric. All reported correlations are Pearson's r. Significance is assessed at the $p < 0.05$ level unless otherwise noted.

Code and data. All scripts, data, and the SQLite database are available in the project repository at https://github.com/carlosrod723/maddux-test-1.