

LAB

Controlled
Standardized
Clean data

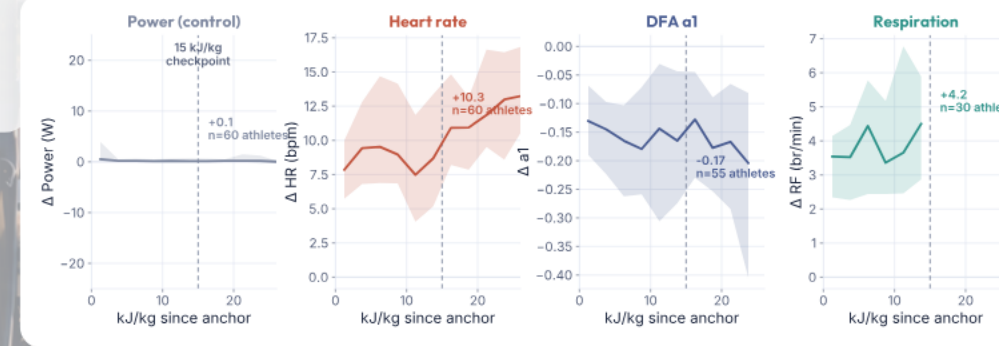
THE WILD

Variable
Unpredictable
Real performance

Measuring Durability in the Wild

Quantifying physiological drift from real-world cycling data using heart rate, respiration frequency, and DFA $\alpha 1$

Markus Rummel · AI Endurance, Hamilton, Canada
Science in Cycling · Bromont 2026



1 · BACKGROUND

Durability has become central to endurance profiling

- **Durability**: onset and magnitude of deterioration during prolonged exercise, studied at 15–35 kJ/kg of accumulated work ($\approx 1,200$ – $2,700$ kJ for a 77 kg rider)

Maunder et al. 2021

- Two constructs: **performance** durability (maximal power after work) and **physiological** durability (internal-external decoupling) Hunter et al. 2025 · Schwalm et al. 2026
- Power **after** accumulated work (not fresh power) **separates competition levels**, and **repeatably so** Mateo-March et al. 2022, 2025
- **Intensity** of prior work, not just volume, drives the downshift Spragg et al. 2024
- HR, RF and DFA a1 drift: repeatable **internal** durability markers

Rogers et al. 2025 · Rothschild et al. 2025

1 • BACKGROUND

The gap: protocols, professionals, single snapshots

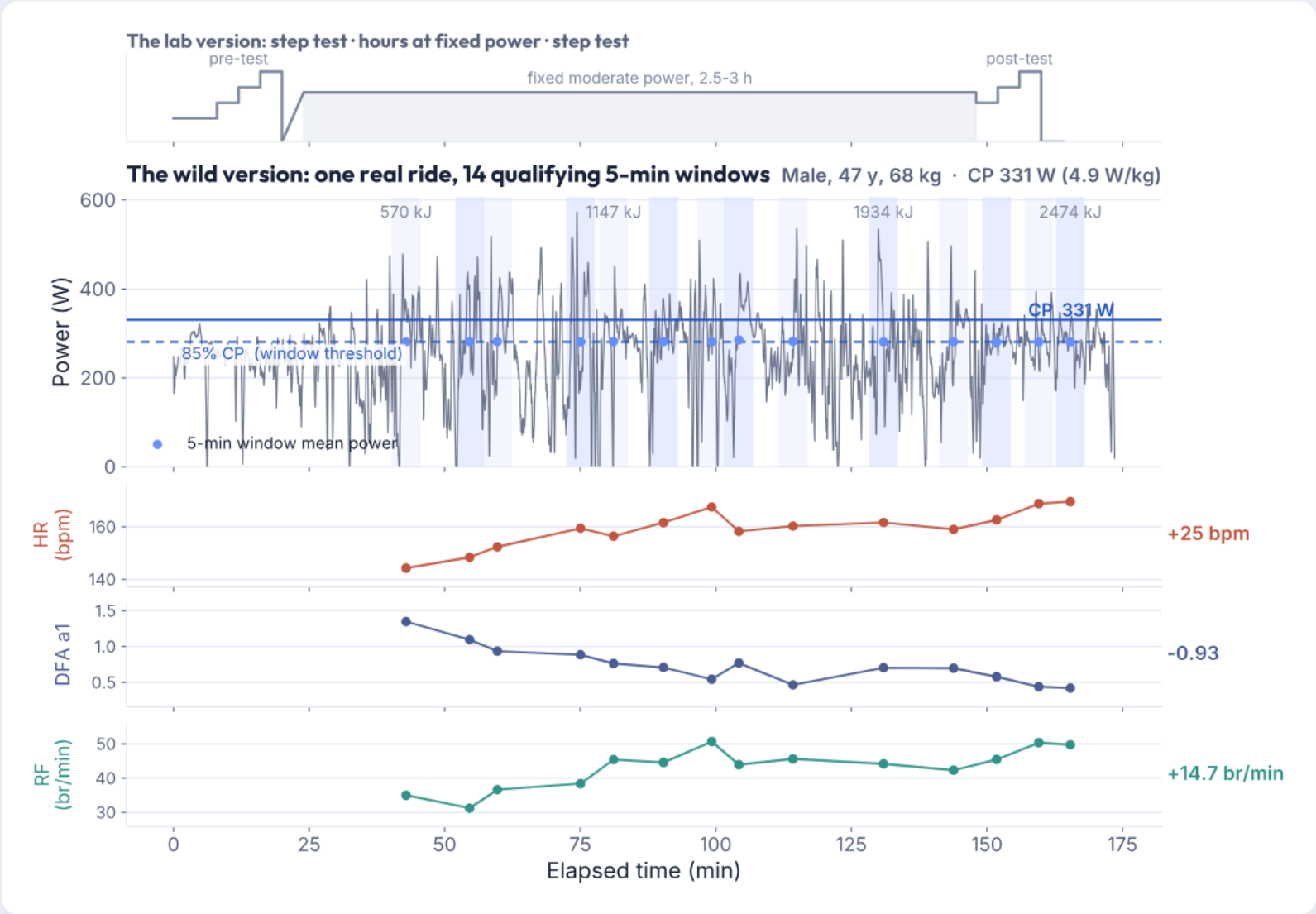
- The evidence so far: [lab protocols](#) and pro race files
- Measured durability depends on the protocol chosen: no standard protocol yet
Spragg et al. 2024 · Hunter et al. 2025
- Protocols aren't how athletes [actually train](#), and no one repeats them every two weeks for a year

Can durability be measured from routine training data, at the same workloads the protocols use (15–35 kJ/kg, $\approx$ 1,200–2,700 kJ)?

2 · THE INSTRUMENT

The unit of measurement: one ride's qualifying windows

Builds on our validated field pipeline: power-DFA a1 from everyday workout data · Andriolo, Rummel & Gronwald 2024



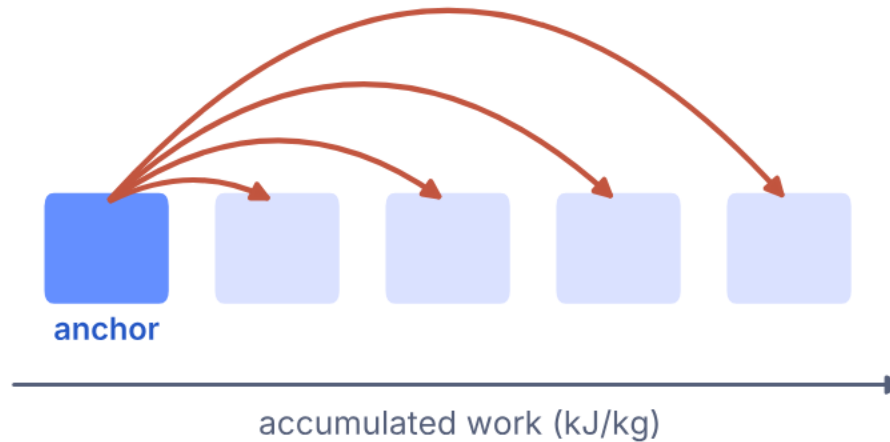
2 · THE INSTRUMENT

Two complementary analyses, adapted from the literature

Two complementary readings of the same windows · internal drift at matched power

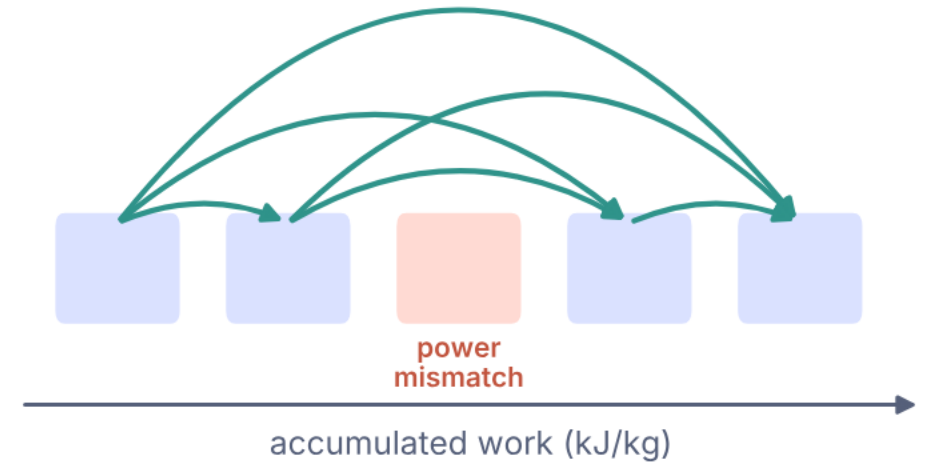
(i) Anchor-based drift

every later window vs the first,
power within $\pm 10\%$ of the anchor



(ii) Within-ride matched pairs

every pair with $|\Delta \text{ power}| \leq 10\%$,
 $\Delta \text{ drift}$ vs $\Delta \text{ accumulated work}$

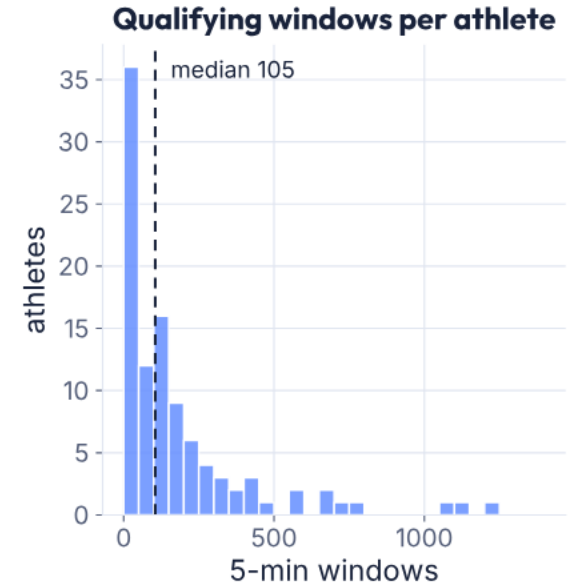
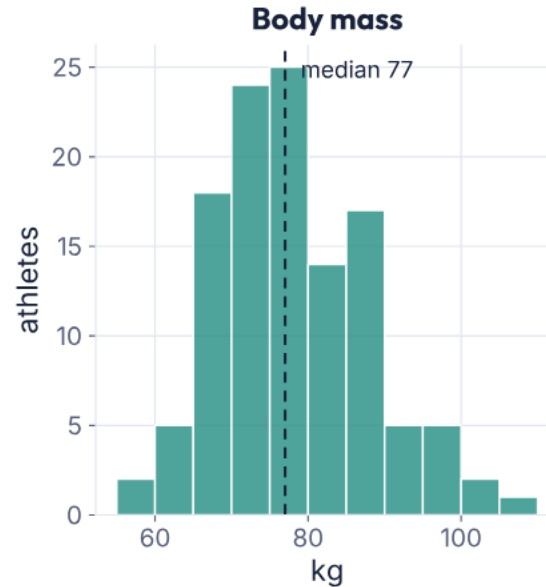
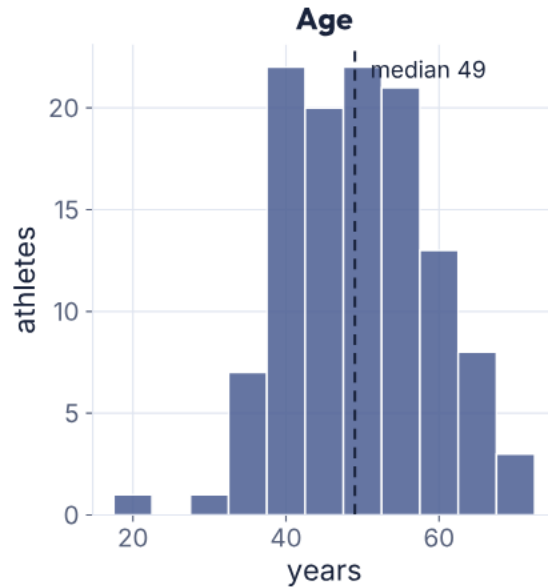


- (i) anchor-based drift: Rogers / Rothschild framing
- (ii) matched paired efforts: Mateo-March / Spragg design, read on [internal](#) measures instead of external power

2 · THE INSTRUMENT

Participants

Participants: 118 eligible (115 M / 2 F / 1 n.r.), 105 contributing windows · amateur, consented, chest-strap RR data



Eligibility (all required)

- informed consent to research use of training data
- ≥ 20 rides with real power-meter data
- ≥ 20 rides with clean RR data (artifact rate $< 5\%$, chest strap: Polar H9/H10, Suunto)
- HR-verified critical-power history (time-varying CP)

Resulting cohort

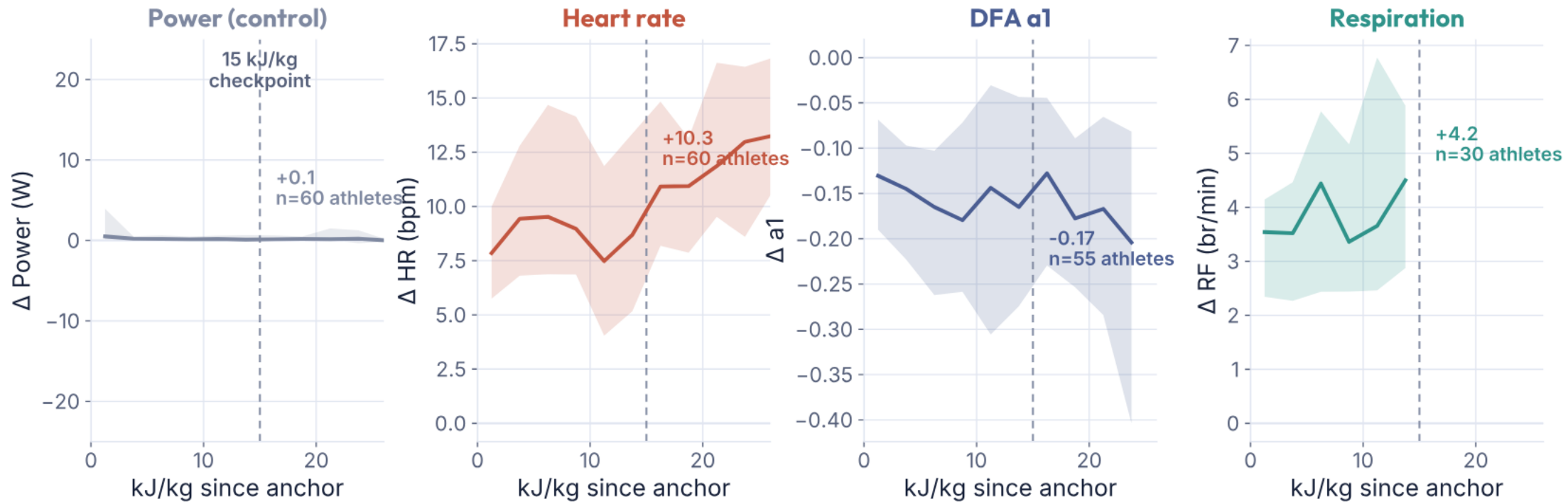
- 4,850 sessions · 26,073 qualifying windows
- **Amateur** athletes: the population protocols can never study at scale
- nearly all male (a platform-usage limitation)

3 • RESULTS

Power matched, internal cost drifts

method (i) • anchor-based drift

Intensity-matched drift vs accumulated work (median + IQR across athletes)



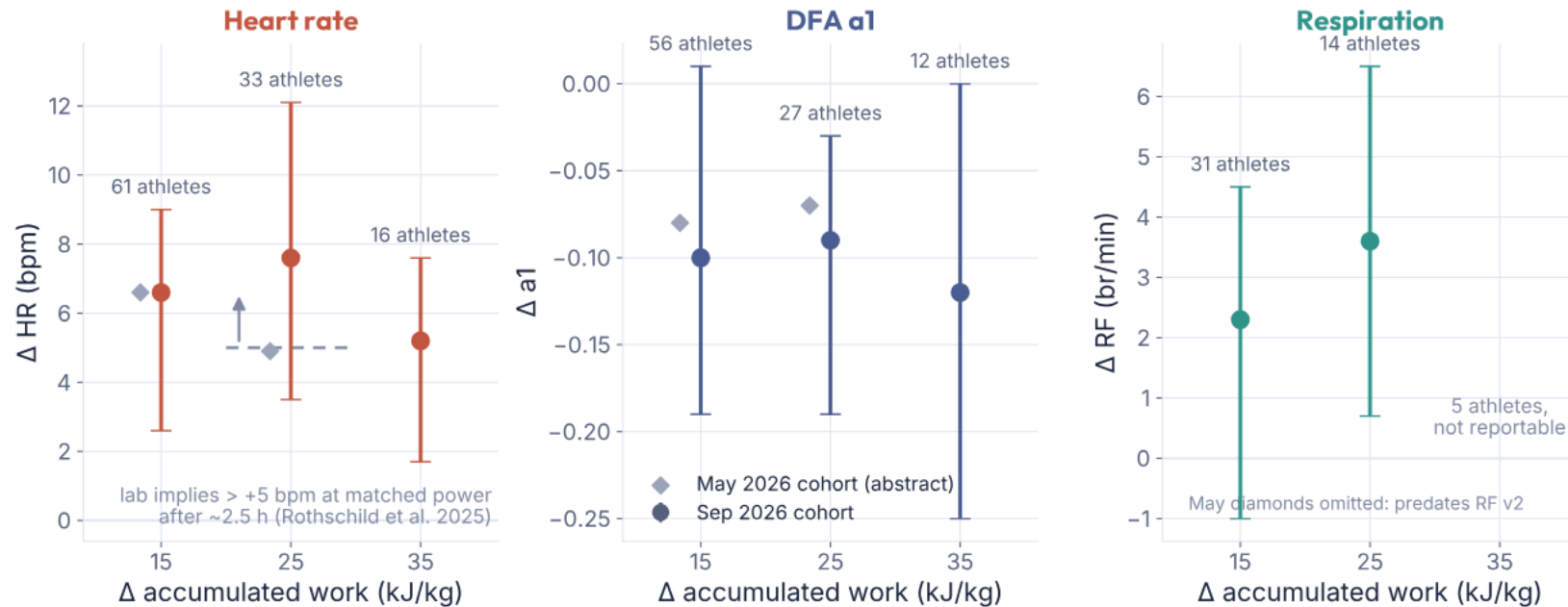
- Same power, rising internal cost: physiological durability, visible in everyday rides

3 · RESULTS

Paired efforts at protocol checkpoints

method (ii) · within-ride matched pairs

Matched-pairs drift at literature checkpoints (median + IQR across athletes)

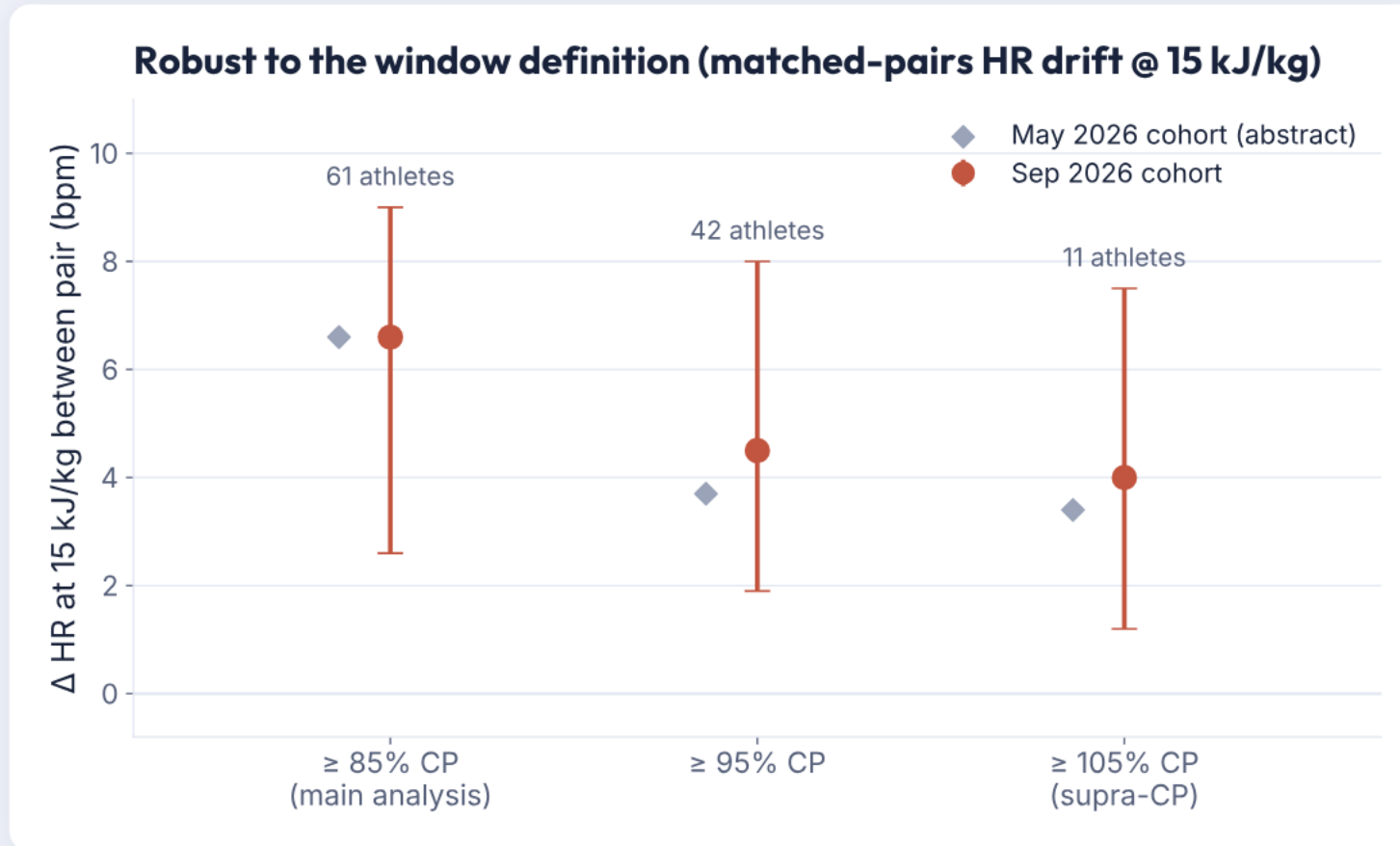


- 77,108 matched pairs · Rothschild: HR at VT1 rose ~5 bpm while VT1 power fell, implying > +5 bpm at matched power after ~2.5 h; our +7.6 bpm [IQR +3.5 to +12.1] at 25 kJ/kg is consistent
- a1 / RF directions consistent with Rogers et al. 2025 (their protocol reaches only ~8-9 kJ/kg); external-power designs (Mateo-March et al. 2022 · Spragg et al. 2024) carry no internal number

4 • WHAT SCALE ADDS

Robust to the window definition

method (ii) • within-ride matched pairs

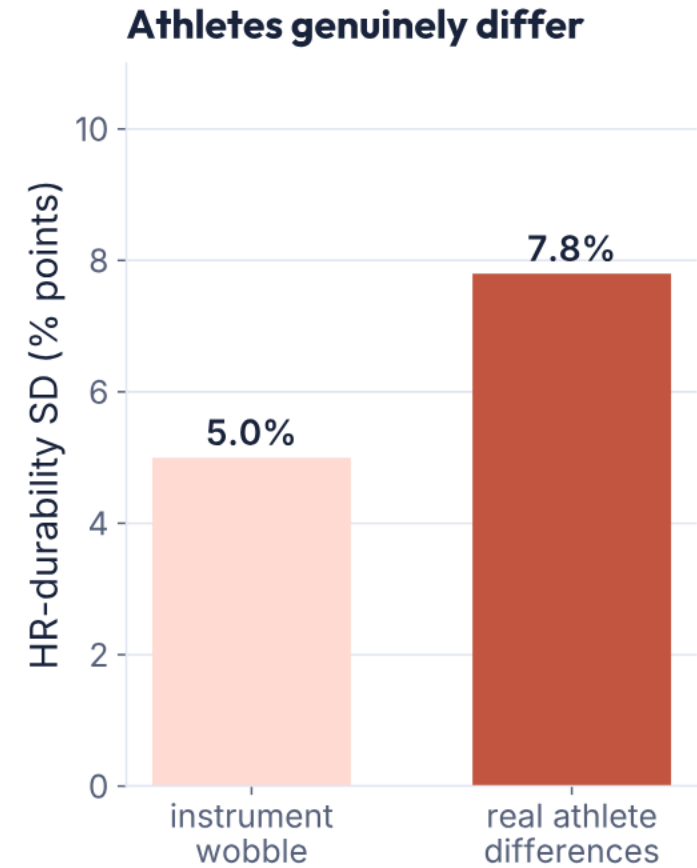
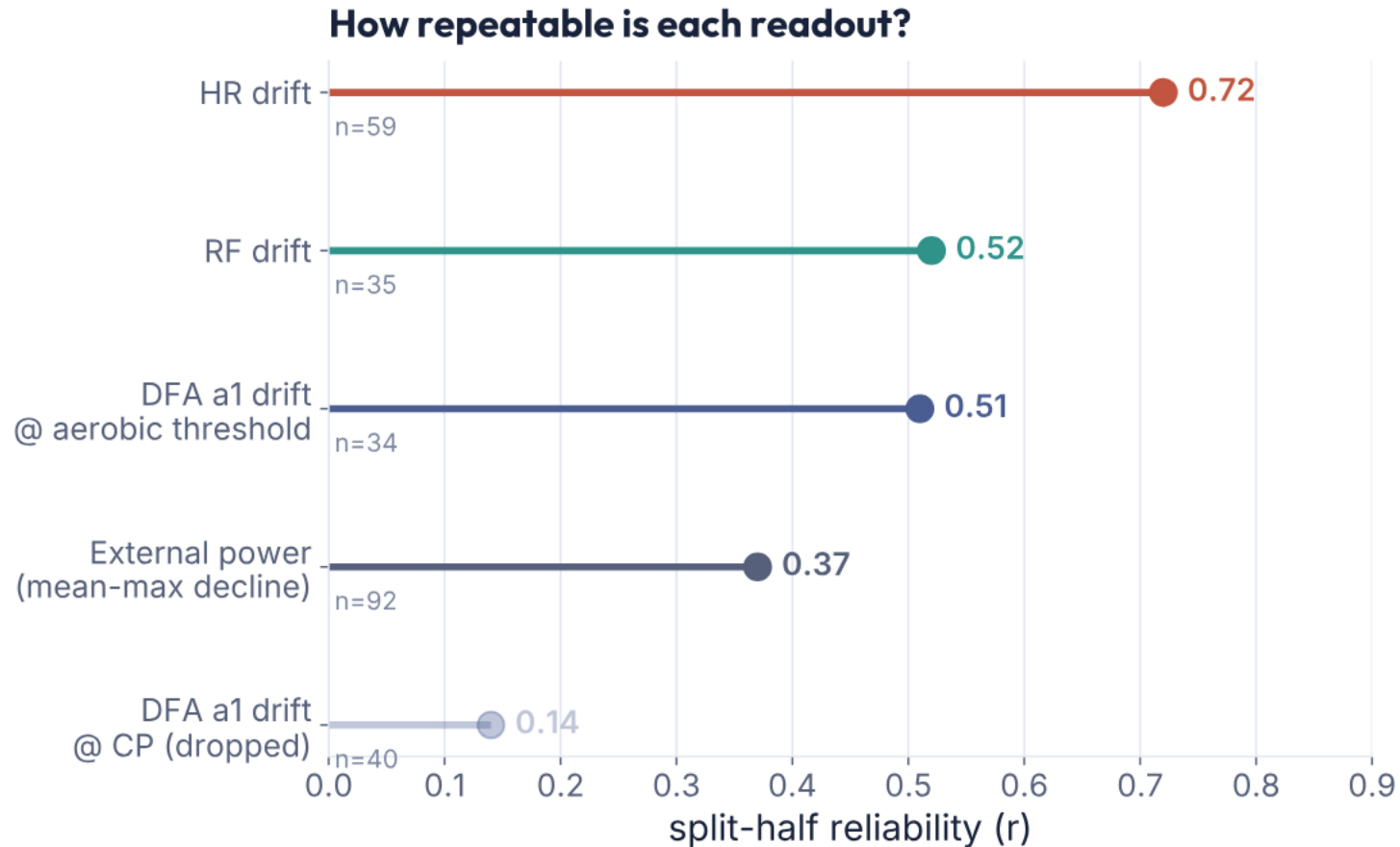


- The drift is a property of the effort, not of our window threshold

4 · WHAT SCALE ADDS

At scale: reliable measurement, real athlete differences

Cohort validation (n = 113 consented athletes, split-half by date, @ 1000 kJ)



4 • WHAT SCALE ADDS

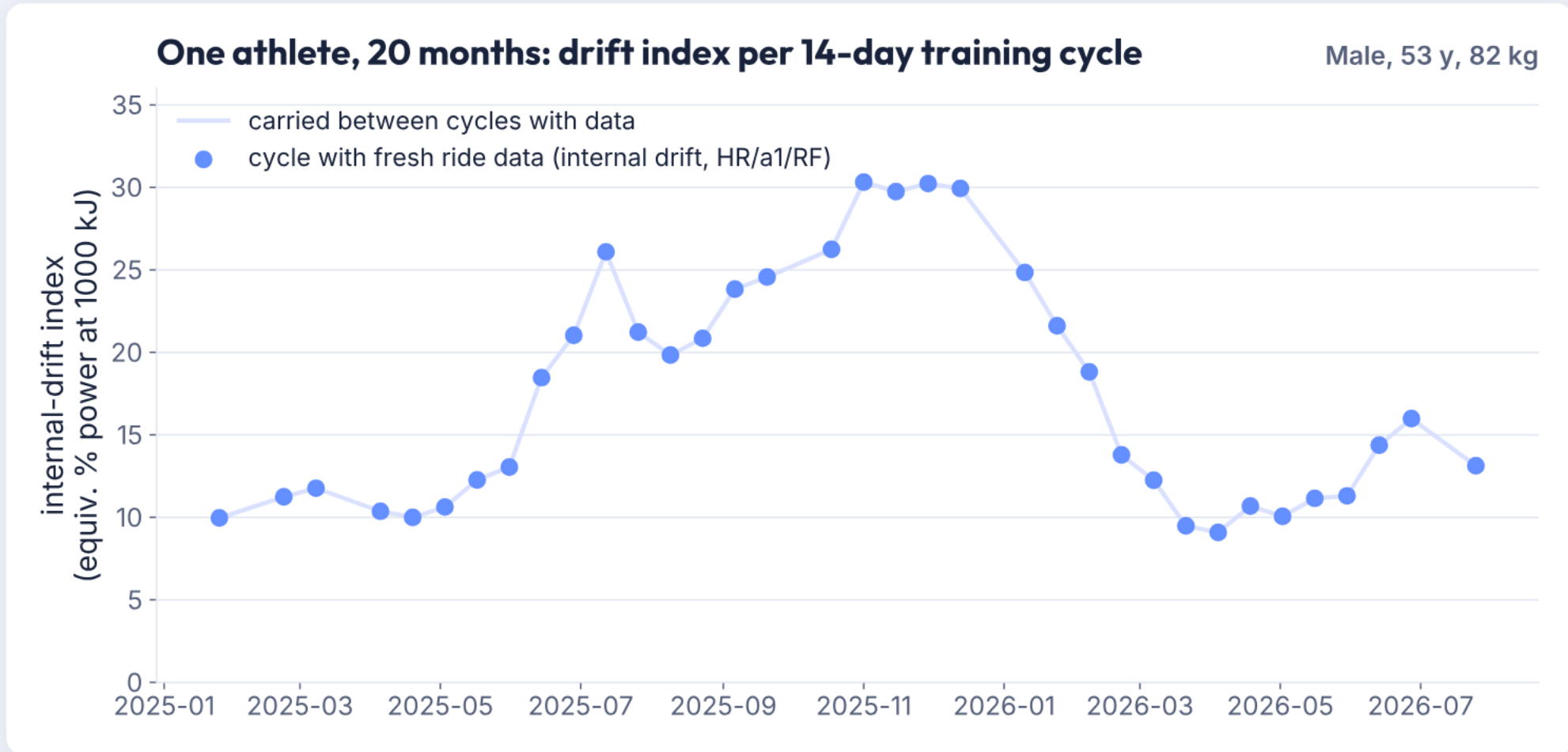
Why physiological durability? Performance durability is nearly unmeasurable in amateur training data

- Mean-max power decline: reliability tops out near $r \approx 0.45$, and only at year-long windows that smooth the signal away HR drift: 0.72
- Late in rides, amateurs rarely go to the limit: soft-pedaling reads as fake "loss", while true decline goes unobserved; no quality gate can rescue efforts that never happened
half the athletes have no late hard efforts
- The two performance-durability readouts, best-power decline (mean-max curve) and matched-effort decline, don't even agree with each other $r \approx 0.0$
- Pro race data is different: racing supplies the maximal efforts training lacks, and fatigued maximal power is repeatable there, incl. across a season Mateo-March et al. 2025 • Spragg et al. 2023

For amateurs, physiological durability is not the convenient instrument but the feasible one

5 • LONGITUDINAL

What tracking looks like: one athlete, twenty months



- **Model-estimated:** internal drift converted to equivalent power units via a fitted conversion, not directly observed power loss; 14 days is the update cadence, and adjacent points share overlapping fitting windows

6 • LIMITATIONS

What this method cannot (yet) do

- Intervening work collapsed to **total kJ/kg**, not stratified by intensity composition
cf. Spragg et al. 2024
- RF shares the RR stream with HR and **a1** dedicated wearables ahead: Tymewear, BreezeLabs
- Cohort thins at 35 kJ/kg and is **nearly all male**
- Needs clean RR data (<5% artifacts): a quality chest strap Polar H9/H10, Suunto
- Within an athlete, **heat and acute load** can shift drift values seasonally; core-temperature sensors are a modeling option

CONCLUSION

Durability, measured where athletes actually ride

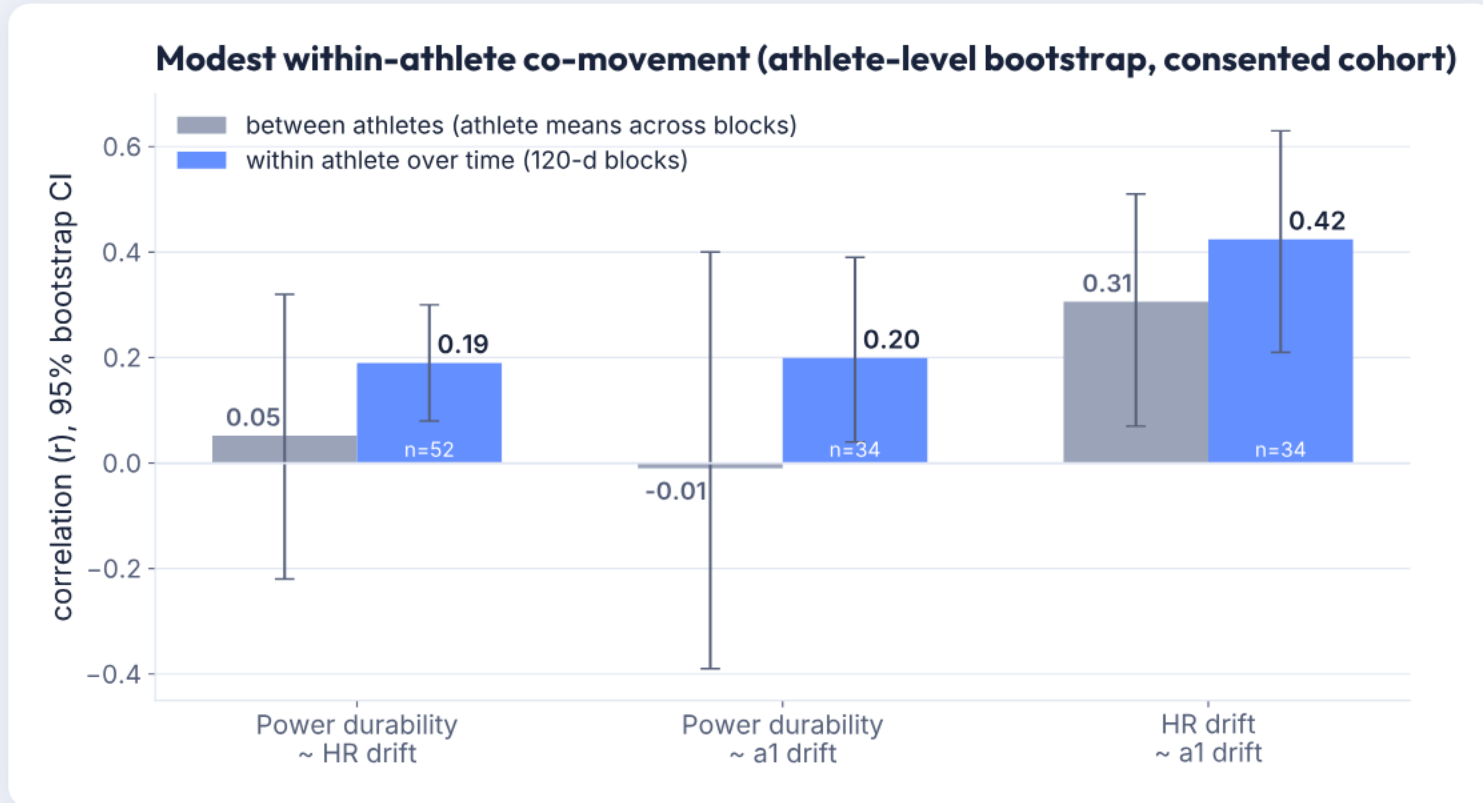
- Physiological durability is measurable from routine training **without dedicated testing**, at protocol workloads, stable under data refresh and window definitions
- HR drift **consistent with the lab's decoupling data**; a1 and RF drift directions consistent with the protocol literature
- Physiological durability shows **modest within-athlete co-movement** with performance durability: early evidence linking the two constructs
- The payoff is **longitudinal tracking** from data that already exists, for the athletes race files cannot measure

Protocols told us durability matters.
Routine training data lets every athlete measure theirs, continuously.

Looking to partner on a lab-criterion study: field physiological durability vs protocol performance durability

BACKUP

Does physiological durability track performance durability? Modest within-athlete co-movement



- **Why it matters:** reliability alone could mean "reliably measuring something meaningless"; this is the field's only validity evidence, and it is early evidence, not proof
- **Power durability** = decline of the mean-max power curve after 1000 kJ, per 120-day block; drift per block the same way; "between" = athlete means across blocks

BACKUP

Does it matter HOW the work was accumulated? (Spragg et al. 2024)

- Spragg et al.: prior **high-intensity** work (despite slightly LESS total work) caused larger declines in short-duration power (1-15 s); CP and 12-min power were unchanged between conditions
- Our pair design collapses intervening work to one number (Δ kJ/kg): composition **not yet stratified**
- Why not yet:
 - • thin strata: 33 athletes at 25 kJ/kg; composition splits fall below reporting thresholds
 - • little natural contrast: protocols manufacture what amateur riding rarely provides
 - • confounding: composition co-varies with ride type, terrain, heat, fueling
- Note the domain: their composition effect concerns **sprint-duration** power; CP-domain performance (where our threshold-intensity drift lives) showed no condition difference
- **Testable retrospectively**: full per-second power streams are stored for every ride, so the intervening work can be stratified by intensity; the missing ingredient is cohort size, not data

BACKUP

HR-drift confounders: what the design does and does not remove

- We measure durability **as expressed under real training conditions**: progressive thermal, hydration and fueling strain is part of expressed deterioration, not contamination of it

Maunder et al. 2021 · Rothschild et al. 2025

- Pairing **removes** between-day baselines (starting hydration, caffeine dose, sleep); it **cannot remove** what accrues with duration: dehydration, thermal strain, depletion travel with the drift
- $r = 0.72$ (HR-drift split-half reliability) certifies a **stable expressed tendency**: it rules out random noise, not stable habits or climate
- Protected use: the athlete against **their own baseline**; cross-athlete or cross-season comparisons need environmental and nutritional context as Hunter et al. 2025 recommend
- Supporting: a1 and RF drift consistently alongside HR, though all three derive from one RR stream, so this is coherence, not independent validation; core-temperature sensors are the natural covariate ahead