PERFORMANCE INTELLIGENCE · TECHNICAL REPORT

SportPlan Performance Engine: Scientific Review and Current-Engine Simulation

Technical report · Version: 4.0
Review date: 7 August 2026
Scope: endurance performance training, adaptive prescription, tapering, load governance, strength, adherence, and escalation
Status: architecture and simulation evidence; no claim of prospective clinical or performance efficacy

1 Technical summary

Verdict. SportPlan now represents a **state-of-the-art performance-training architecture**: it combines athlete-specific response modelling, multi-signal readiness, event-aware tapering, structured intensity constraints, strength as a first-class training domain, explicit missingness, auditable rules, and human escalation. That verdict concerns the design and implemented decision boundaries. It is **not** a claim that SportPlan has already outperformed coaches or reduced injury in real athletes.

Three proof layers support the conclusion. First, a publication-scale synthetic season trial contains **38 sport × level cohorts, 9,994 paired athlete profiles per arm, six arms, and 59,964 arm simulations**. Second, the current code passed **13 of 13 deterministic boundary scenarios** covering corroboration, taper geometry, individual response, interval programming, adherence missingness, and escalation. Third, a project-local rapid synthesis contains **466 full-text review documents and 16,703 extracted outcome rows**; only **94 reviews (20.2%)** meet the moderate/high-quality restriction, preventing weak domains from being presented as settled science.

What is genuinely advanced. The engine does not reduce training to a single readiness score or a fixed “optimal” ratio. It preserves uncertainty, separates observed non-completion from silence, requires corroboration before a contested perception changes a taper, keeps intensity and frequency intact when taper duration is reduced, and withholds a prescription when medical or training-context conflicts require a human. This is consistent with current evidence that performance response is heterogeneous and that no single training-intensity distribution, monitoring signal, or workload ratio should be treated as universally optimal.[1,2,3,4,5,6,7]

What remains unproven. The synthetic trial predicts that the fully integrated adaptive + life-fit configuration can improve modelled performance, completion, and plan retention. These are hypotheses generated by a shared mathematical world, not causal estimates. A prospective, preregistered comparison using objective performance outcomes and a real coach comparator is still required.

STATE-OF-THE-ART CONFORMANCE

Contemporary performance-training requirement	SportPlan implementation	Evidence status
Individual response, not population average alone	Personal fitness-decay estimate may shift taper length within bounded limits	Implemented boundary; athlete efficacy untested
Multi-source monitoring	Completed load, readiness, perceived effort, soreness, life constraints, adherence, and event context remain separate inputs	Implemented; no single signal is treated as truth
Periodized intensity with sport/context heterogeneity	Interval-block envelope and protected race/taper commitments	Validator implemented; production-path coverage is not universal
Taper preserves quality while shedding volume	Duration-only cuts; intensity/frequency protected; final 48 h protected	Evidence-shaped design plus engine-policy bounds
Strength integrated with endurance	Strength exposure is modelled as a training domain, not an optional footnote	Evidence-supported construct; dose translation remains modelled
Uncertainty and safety governance	Missingness lowers confidence; conflicts withhold and route to coach or clinician	Implemented fail-closed behavior
Traceability	156 registered rules with provenance labels and policy notes	Audit inventory, not external validation

Claim boundary · state-of-the-art architecture; prospective athlete efficacy remains unproven.

2 System and study design

ENGINE UNDER REVIEW

The current engine is a bounded decision system, not a black-box predictor. It estimates completed training dose and individual response, interprets readiness without allowing absent data to become a negative score, applies event and taper constraints, and records why it adjusted, preserved, withheld, or escalated a decision. The rule registry currently contains **156 entries**: 7 evidence-anchored, 21 literature-cited, 121 engine-policy, 6 practitioner, and 1 hypothesis. These labels are deliberately non-interchangeable: a literature citation can justify the shape of a rule without validating its numerical threshold.

TWO-LANE SIMULATION

The review uses two separate estimands:

1. **Season lane.** **A paired, seed-42 synthetic trial compares A adaptive, B authored static, C generic static, D adaptive + life-fit, E self-coached, and F conventional-coach-shaped weekly review. Every arm shares the same latent athlete profile and outcome model. Arm F models a cadence and decision shape; it is not observed human coaching.**
2. **Boundary lane.** **Thirteen frozen scenarios call pure functions from the current engine. A pass means the returned action matches the declared contract. It does not prove benefit, production reachability, or external validity.**

Primary season outcomes are the arbitrary-origin truth-performance index, completion percentage, plan abandonment or training cessation, unavailable illness/injury days, and overload weeks. Only point differences are interpreted for the performance index. Arm E cessation and plan-arm abandonment are related but non-identical constructs.

3 Season-scale simulation

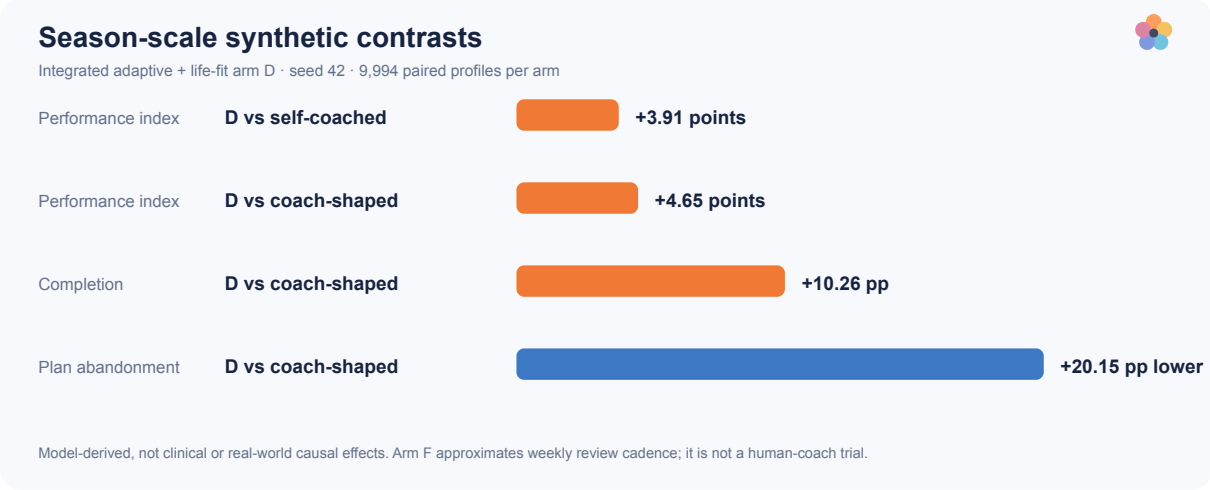

Figure 1 | Integrated-engine contrasts in the publication-scale synthetic season. Arm D is the adaptive engine plus life-fit scheduling. Positive bars favor D except plan abandonment, which is shown as percentage points lower. Values are model-derived contrasts, not real-world effects.

MAIN RESULTS

Outcome	D integrated	E self-coached	F coach-shaped	D vs E	D vs F
Truth-performance index	135.57	131.66	130.92	+3.91 points	+4.65 points
Completion	63.98%	NA	53.72%	NA	+10.26 pp
Plan abandonment / cessation	43.02%	55.17%†	63.17%	−12.15 pp†	−20.15 pp
Unavailable illness/injury days	1.97	1.93	1.84	+0.04	+0.13
Overload weeks	2.24	1.88	1.97	+0.36	+0.27

†Arm E measures cessation, while D measures plan abandonment; the contrast is descriptive only.

The integrated configuration ranks first on modelled performance and completion and lowest on plan abandonment among the plan arms. The result is not explained by a universal “more adaptation is always better” rule: arm D also delivers more completed training and shows slightly more overload weeks and unavailable days than E or F. Therefore the simulation supports a **performance-and-engagement hypothesis with an explicit load trade-off**. No injury-prevention or clinical-safety superiority is inferred.

The season trial predates several new boundary modules and does not exercise every new rule end to end. It remains the best available scale test of the integrated adaptive/life-fit configuration, while the boundary battery below is the current-code proof for newer individualization and governance behavior. Keeping the lanes separate avoids presenting unit-level conformance as season-level efficacy.

4 Current-engine boundary simulation

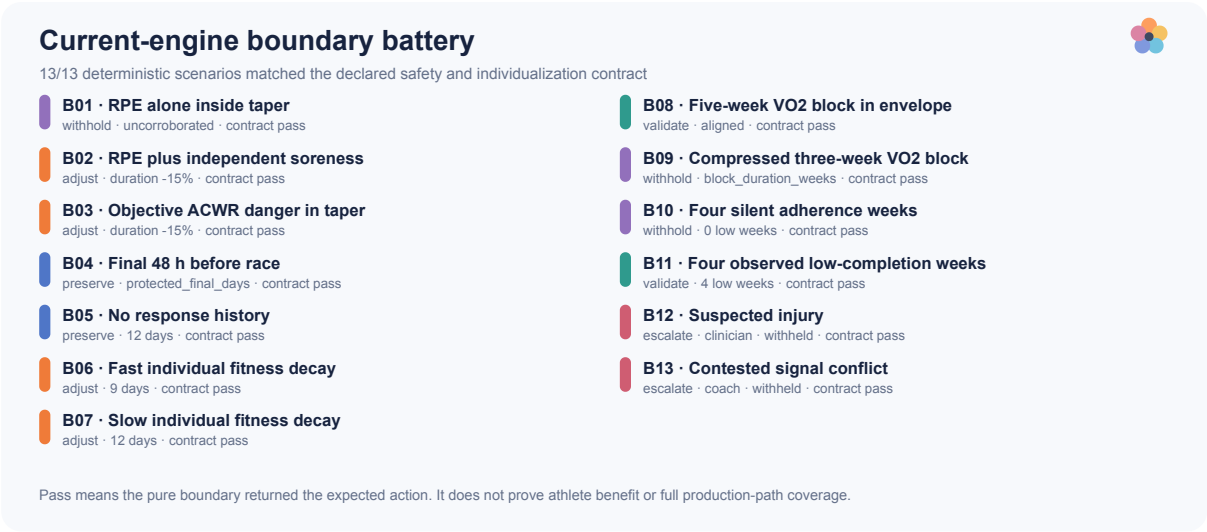

Figure 2 | Deterministic boundary battery executed against the current engine. All 13 scenarios matched their prespecified action. "Withhold" is a successful safety behavior, not a failure.

WHAT THE BATTERY PROVES

Boundary family	Direct current-code result	Interpretation
Contested effort	RPE 9 alone: no taper change; RPE 9 + independent soreness: duration −15%	Corroboration is operational, not merely documented
Objective load in taper	ACWR 1.62: duration −15%; intensity is structurally absent from the output	Objective guardrail survives the taper without breaking taper quality
Protected final days	Even ACWR 2.0 cannot alter race−1	Race preparation cannot be destabilized at the last moment
Individual taper kinetics	No history: table stays 12 d; fast decay: 9 d; slow decay: 12 d	Personal data may nudge, never replace, the authored anchor
Interval programming	Five-week envelope passes; compressed three-week block is rejected	The validator reports evidence-envelope misses instead of silently compressing
Adherence missingness	Four silent weeks: zero low-consistency weeks; four observed low weeks: four	Absence of data becomes uncertainty, not a negative athlete label
Human authority	Suspected injury → clinician; contested training signals → coach; prescription withheld	The engine has a named "I am not deciding this" state

This battery demonstrates that the new engine's most important epistemic rules are executable: **corroborate, bound, preserve, abstain, and escalate**. That is a stronger standard than a dashboard that merely displays evidence language while allowing an uncorroborated value to change training. The remaining gap is integration coverage: the interval validator and rule registry, for example, do not by themselves prove every production prescription passes through them.

5 External evidence benchmark

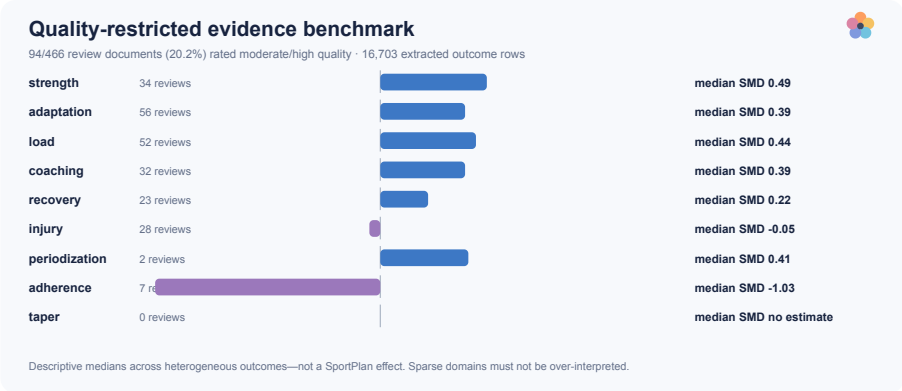

Figure 3 | Quality-restricted domain benchmark from the project-local full-text rapid synthesis. Median standardized effects summarize heterogeneous outcomes within each domain; they are not SportPlan effects and are not pooled estimates for one intervention.

After quality restriction, strength (34 trustworthy reviews; median SMD 0.49), load (52; 0.44), adaptation (56; 0.39), and coaching (32; 0.39) retain positive descriptive medians. Recovery is smaller (23; 0.22). Injury collapses to –0.05 across 28 trustworthy reviews, while taper has zero trustworthy reviews in this local rating. Periodization and adherence have only 2 and 7 trustworthy reviews respectively—too sparse for product-level superiority claims.

This pattern is consistent with recent primary syntheses. Strength training can improve endurance performance determinants, but certainty and transfer vary by sport and protocol.[4,8] Taper evidence supports reducing volume while maintaining intensity and frequency, not arbitrary last-minute session deletion.[1] The newest network meta-analyses do not establish one universally superior intensity distribution and identify training status, sport, duration, and measurement method as moderators.[2,3] HRV-guided training shows uncertain or small average performance advantages, supporting its use as one input rather than a command signal.[5] A randomized ACWR-guided intervention did not reduce health problems, so SportPlan treats ACWR as a conservative load guardrail—not an injury predictor.[6]

6 Interpretation, limits, and next validation

SCIENTIFIC INTERPRETATION

Defensible claim: SportPlan is state-of-the-art **by architecture** because it operationalizes current performance-training principles while exposing uncertainty, policy ownership, and human authority. The combination matters: personalization without bounds can overfit noise; monitoring without missingness semantics can punish silence; tapering without protected intensity can erase quality; automation without escalation can conceal unsupported decisions.

Not yet defensible: SportPlan improves race performance, prevents injury, or outperforms a skilled coach in real athletes. The season model shares assumptions across arms, cannot capture the full quality of human coaching, and uses a performance index with an arbitrary origin. The boundary battery checks contract behavior, not effect size. The local evidence corpus is broad and quality-stratified, but its domain medians mix outcomes and designs and only 20.2% of review documents pass the moderate/high-quality restriction.

REQUIRED PROSPECTIVE TEST

The next decisive study should preregister a minimum 12–16 week, multi-sport pragmatic trial with SportPlan adaptive + life-fit versus (a) static structured planning and (b) real coach-led programming. Primary outcomes should be objective event-appropriate performance change and completion of planned training; secondary outcomes should include validated wellbeing/readiness, dose actually delivered, adverse events, and retention. Analysis must be intention-to-treat, stratified by baseline level and sport, with blinded outcome adjudication where feasible. The engine version, rule registry hash, decision logs, and every abstention/escalation must be frozen before enrollment.

Release threshold for an efficacy claim: a prospectively declared, practically meaningful benefit with uncertainty bounds that exclude trivial effects, no evidence of unacceptable adverse-event trade-off, and independent replication or an external audit. Until then, the correct public statement is: **advanced, evidence-shaped adaptive training software with promising synthetic results—not clinically or competitively proven superiority.**

7 References

1. Wang Z, Wang YT, Gao W, Zhong Y. Effects of tapering on performance in endurance athletes: a systematic review and meta-analysis. *PLoS One*. 2023;18:e0282838. doi:10.1371/journal.pone.0282838.
2. Li H, Yang Q, Wang B. Effects of different training-intensity distribution models on maximal oxygen uptake and time-trial performance in endurance athletes: a Bayesian network meta-analysis. *J Strength Cond Res*. 2026;40:e755–e764. doi:10.1519/JSC.0000000000005415.
3. Walsh JA et al. Which training intensity distribution intervention will produce the greatest improvements in maximal oxygen uptake and time-trial performance in endurance athletes? An individual-participant-data network meta-analysis. *Sports Medicine*. 2025. PMID 39888556.
4. Blagrove RC, Howatson G, Hayes PR. Effects of strength training on the physiological determinants of middle- and long-distance running performance: a systematic review. *Int J Sports Physiol Perform*. 2018;13:1117–1149. doi:10.1123/ijspp.2017-0032.
5. Javaloyes A et al. Heart-rate-variability-guided training for enhancing aerobic fitness and endurance performance: systematic review and meta-analysis. 2021. PMID 34639599.
6. Fanchini M et al. Does load management using the acute:chronic workload ratio prevent health problems? A cluster randomised trial of 482 elite youth footballers. *Br J Sports Med*. 2021;55:108–114.
7. Schwellnus M et al. How much is too much? International Olympic Committee consensus statement on load in sport and risk of illness. *Br J Sports Med*. 2016;50:1043–1052.
8. Rosenblat MA et al. Heavy strength training effects on physiological determinants of endurance cyclist performance: a systematic review with meta-analysis. *Eur J Appl Physiol*. 2025. doi:10.1007/s00421-025-05883-2.