PERFORMANCE INTELLIGENCE
WHITEPAPER SERIES

PAPER FIVE OF SEVEN · TRIATHLON AND ENDURANCE

The Connected Endurance Athlete

How artificial intelligence can integrate training, recovery and race execution in triathlon

Paul Forrest
Founder and Chief AI Officer, Athleet.AI
December 2025

About this series

This is the fifth of seven papers examining how artificial intelligence can be used responsibly in performance sport. Each paper stands alone. Together they set out a single argument, which is that the durable advantage in sport comes from better decisions rather than from more data, and that better decisions require systems designed around the person who is accountable for them.

The series shares one framework, described in Chapter 1 and applied here to a sport that combines three disciplines, a gym, a kitchen and a life. The flagship paper on governance, *The Responsible Performance System*, supplies the assurance model this paper draws on.

Editorial balance. Roughly seven parts in ten of this document is evidence and sector analysis, two parts in ten is original framework, and one part in ten concerns the Athleet.AI platform itself. Readers who want the argument without the product can stop at Chapter 13.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim in this paper carries a tag stating what kind of evidence supports it. The tags are used consistently across the series.

PEER-REVIEWED A primary study in a peer-reviewed journal.

REVIEW A systematic review or meta-analysis.

CONSENSUS A consensus statement, governing-body position or regulator guidance.

COMMENTARY Methodological criticism, which carries weight without being primary data.

PRACTICE Practitioner opinion, including the author's own.

ATHLEET.AI The author's own framework or analysis, proposed rather than proven.

Where the evidence is thin, the paper says so. Where it points the other way, the paper says that too.

Legal, medical and regulatory notice. This paper describes physiological, nutritional and environmental evidence, and in Chapters 8 and 11 it touches on health and on competition regulation. Nothing in this document constitutes medical or legal advice, and nothing in it should be used to diagnose, treat or screen for any condition. The provision of such advice sits outside the terms of any engagement with the author or with Athleet.AI, and the material is presented to support discussion and further review by qualified practitioners and advisers.

Statements about what a source says are attributed to that source. Where the author draws a practical implication the source does not itself state, the text says so. The position described is the position the author was able to verify in December 2025.

ILLUSTRATIVE MATERIAL

Every athlete, squad, programme and scenario in this paper is invented, and none describes a real athlete, coach or organisation. Cases are marked **illustrative scenario** wherever they appear. Published research is cited and numbered, and invented material is never presented as evidence.

CITATION

Forrest, P. (2025). *The Connected Endurance Athlete: How Artificial Intelligence Can Integrate Training, Recovery and Race Execution in Triathlon.* Athleet.AI Performance Intelligence Whitepaper Series, Paper Five. Version 1.0, December 2025.

Comments and corrections are welcome and will be acknowledged in later versions.

ALSO IN THIS SERIES

- **Paper one.** *The Responsible Performance System.* Governance and trust in sports AI.
- **Paper two.** *The Intelligent Track.* AI and event-specific expertise in elite athletics.
- **Paper three.** *Closing the Coaching Gap.* Responsible AI in grassroots athletics.
- **Paper four.** *Casting the Net Wider.* AI as a force multiplier in football pathways.
- **Paper six.** *The Augmented Touchline.* Responsible AI in elite football.
- **Paper seven.** *The Augmented Golfer.* AI across technique, practice and strategy.

Contents

A NOTE FROM THE AUTHOR

Six apps and nobody in charge

The triathlete I have in mind is not a professional. She is 38, she has two children and a job, she trains about nine hours a week and she has been trying to go faster for four years. On her phone are six applications. One holds the swim, one the bike, one the run, one the strength work she does when there is time, one a sleep score she has learned to distrust and one a plan she bought in January and stopped following in March.

None of the six knows about the other five. The plan does not know she rode hard on Tuesday because a meeting moved. The sleep score does not know she was awake with a child. The strength app has never been opened by anybody but her. When she asks her coach why she is slower than last year, the honest answer is that nobody has the record that would let them tell.

That is the problem this paper is about, and it is not a modelling problem. It is a reconciliation problem, and it is unglamorous enough that almost nobody is selling a solution to it.

Triathlon is the best test case in sport for integrated performance intelligence, precisely because everything that makes it hard is visible. Three disciplines that adapt on different timescales. A strength component that most athletes treat as optional and most software treats as absent. Fuelling that decides long-course races. Weather and water temperature that change the event on the morning. And a population that is overwhelmingly self-coached, working around a life that was not designed around the sport.

I have written this paper to be useful to that athlete and to the coach trying to help her, and to be uncomfortable reading for anybody selling a single number that claims to summarise her Tuesday. Several of the recommendations here would rule out product features that are commercially easy to build. That is deliberate, and Chapter 5 explains why.

Test the citations rather than the author. Every one of them is listed, and every one has been checked twice.

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

Triathlon does not have a shortage of measurement. It has a shortage of coordination. Three disciplines, a gym, recovery, fuelling, equipment, environment and the rest of an athlete's life have to be reconciled into one week, and no product on the market holds all of it in a form a coach can reason about.

This paper argues that the honest opportunity in endurance sport is reconciliation and explanation rather than prescription. A system earns its place by holding the whole week in one record, naming what it cannot see, stating the error on the numbers it reports and helping a coach and an athlete make a trade-off they can both defend. It argues equally firmly that composite readiness scores, automatic plan rewriting and any inference about energy availability belong outside the product.

1

The interference effect is narrower than the folklore. Across 43 studies, concurrent training cost nothing measurable in maximal strength or muscle size. The cost fell on explosive strength, and it appeared only when both modes were done in the same session.

2

Strength work changes economy, not the engine. Concurrent training improved running economy by about 3.9 per cent, and explosive work by 4.8 per cent, while a recent meta-analysis found no improvement in maximal oxygen uptake or the maximum metabolic steady state.

3

No commercial readiness score has been validated. A search of the peer-reviewed literature found no validation of any proprietary composite recovery or readiness score against a physiological or performance criterion. The evidence that exists is qualitative.

4

The instruments are worse than the interfaces suggest. Wrist heart rate carries limits of agreement from minus 24 to plus 18 beats per minute. Sleep-stage specificity against laboratory measurement ranged from 0.18 to 0.54, and was worst on disturbed nights.

5

How the bike is ridden changes the run. At identical average power, an hour of variable cycling made the following 9.3 kilometre run 42 seconds slower, and 35 of those seconds appeared in the first half.

6

The intensity debate is softer than it sounds. A network meta-analysis of individual participant data found no difference between polarised and pyramidal distributions, and the same training classified two defensible ways differs by a factor of three.

What follows for a coach or an athlete

Hold one record, not six. Reconcile the disciplines against one athlete identity and one clock, and treat the gym and the rest of life as part of the record rather than as context somebody remembers. Report error alongside every derived number. Refuse the single score. Measure completed planned weeks, which is countable, auditable and the measure most closely linked to hitting a performance goal in the published prospective work.

Refuse four things outright. No composite readiness score presented as a fact. No automatic rewriting of a plan without a person approving it. No inference about energy availability, which the governing consensus reserves for a physician and for which no single validated diagnostic test exists. And no claim to have measured what the device only estimated.

None of that requires a better model, which is the part that tends to disappoint. The changes that would help most are administrative rather than analytical, and the largest of them costs one decision about which source wins when two disagree. What a system adds is coherence, memory and an honest account of its own uncertainty, and it adds those only once somebody has decided what the record is going to say.

The gains are also unevenly distributed, and worth aiming deliberately. An elite programme already has staff to reconcile its own data by hand. A self-coached age-group athlete has nobody, and is the person a coherent record would help most. That is an argument for building for the second and letting the first benefit, which is the opposite of how endurance software has been priced.

Chapter 13 proposes a connected athlete pilot for a coach, squad, federation or technology partner, with an independent evaluation partner and measures on completed planned weeks, reconciliation accuracy and the proportion of recommendations a coach changed before acting. A reader who wants the argument without the product can stop at Chapter 13.

PART ONE

Three sports, one athlete

Four chapters on what makes triathlon the hardest coordination problem in endurance sport, and on what a system would have to hold to help.

CHAPTER ONE

Why triathlon is the hardest test

If integrated performance intelligence works anywhere, it has to work here. Triathlon presents every coordination problem in endurance sport at once.

Most endurance sports ask a coach to reconcile two things, which are training and life. Triathlon asks them to reconcile seven. Three disciplines that improve on different timescales and interfere with one another in ways the literature is still arguing about. A strength component that is genuinely useful and almost never recorded in the same place as the endurance work. Recovery measured by instruments whose individual error is rarely stated. Fuelling that decides long races and is trained by almost nobody. Equipment and environment that change the event on the morning. And a life that was arranged before the sport arrived.

The scale of the sport in England is worth stating early, because it shapes what is economically possible. National activity data put adult triathlon participation at 24,200 people, a rate of 0.05 per cent, unchanged from the 2015 baseline and unchanged over the preceding year.[52] SURVEY Running and swimming sit at 6.5 million and 4.2 million on the same measure. Triathlon is at the edge of the survey's resolution, which is a statistical fact with a commercial consequence. This is not a market that will be served by a product built only for professionals.

1.1 What makes it hard is not what people think

Ask a coach what makes triathlon difficult and they will usually say volume. My own view, after some years of watching people fail at it, is that volume is the easy part. The hard part is that every decision in a triathlon week is a trade-off against another discipline, and almost none of the information required to make that trade-off well is in one place.

An hour is spent. It could have been swum, ridden, run or lifted. Whichever it was, the other three did not happen. In a single-discipline sport the counterfactual is rest. In triathlon the counterfactual is a different adaptation, and choosing between them requires knowing what the athlete has already done, what they respond to, what their life will permit next week and what the race actually asks of them. That is a reasoning problem, not a prediction problem, and it is the reason this paper spends more time on reconciliation than on models.

0.05%

ENGLAND, ADULTS

Triathlon participation on the national activity survey. Unchanged from the baseline and unchanged over the last year, and small enough that the confidence interval is wide.

THE ARGUMENT IN ONE PARAGRAPH

Triathlon's problem is coordination, not measurement. The useful system is the one that holds three disciplines, a gym, a recovery record and a life in a single reconciled account, states what it does not know, and helps a coach and an athlete choose between competing adaptations with the trade-off visible. Everything that follows in this paper is either evidence about what that system should hold, or a limit on what it should be allowed to say.

1.2 The self-coached majority

The other structural feature is that most triathletes coach themselves. That changes the design problem entirely. A system speaking to a coach can be terse and technical, because a professional will interrogate it. A system speaking to a self-coached athlete at ten o'clock at night is the only voice in the room, and it will be believed. Chapter 10 sets out what follows from that, and it is the chapter I would most like suppliers to read.

CHAPTER TWO

The fragmented endurance stack

Six records, six timestamps, six vocabularies. The gap is not in what is measured. It is in what nothing is measuring.

Before proposing what a system should do, it is worth being precise about what the current one does not. The honest inventory is uncomfortable, because the parts that are missing are the parts that most often explain a disappointing season.

THE FRAGMENTED ENDURANCE STACK

DISCIPLINE	WHAT USUALLY RECORDS IT	WHAT NO SYSTEM SEES
Swim	Pool clock, lap counter, occasional video	Stroke mechanics under fatigue, open-water sighting, water temperature effects
Bike	Power meter, head unit, indoor platform	Position under fatigue, road conditions, what the ride cost the run
Run	Watch, foot pod, optical heart rate	Mechanics late in a session, surface, the accumulated cost of the other two
Gym	A notebook, if anything	Everything. Strength work is almost never in the same record as the endurance work
Recovery	Sleep score, readiness score, resting heart rate	Whether any of it is measuring what it claims to measure
Life	Nothing	Work, travel, family, illness, the four things that actually move the plan

Six records, six timestamps, six vocabularies and no single place where a coach can see the week the athlete actually had.

FIGURE 1 What records each part of a triathlete's week, and what nothing records at all. The third column is where most explanatory power sits.

Athleet.AI analysis. The measurement gaps in the third column are supported by the evidence in Chapters 5 and 6. ATHLEET.AI

2.1 Where the record breaks

Three breaks matter more than the rest. The first is the gym. Strength work in triathlon is genuinely useful, as Chapter 7 shows, and it is almost never in the same record as the endurance work, which means it is invisible to any analysis of the week. The second is life. Work, travel, family and illness are the four things that most often move a plan, and no consumer product holds any of them. The third is reconciliation. When two devices disagree, and they will, nothing decides what the record says.

That third break is the one I would fix first. A watch, a head unit and an indoor platform will produce three different accounts of the same ride, and an athlete management platform will usually take whichever arrived last. Nobody notices until a coach tries to explain a training block six months later and finds that the record contains two Tuesdays.

2.2 Why the market has not fixed it

Reconciliation is unglamorous, hard to demonstrate and impossible to put in a screenshot. Every incentive in the consumer market pushes toward a number on a watch face. My own view, offered as practice rather than evidence, is that this is why the endurance software category has improved its charts for a decade without improving the decisions those charts inform.

ILLUSTRATIVE SCENARIO

The fortnight that disappeared

An invented age-group athlete has a poor race and asks why. The coach opens the record. The bike file for the key long ride two weeks earlier is missing, because the head unit did not sync and the watch recorded the same ride at a different duration. The strength work is in a notebook. The week of a work trip shows three sessions, which is what the plan said, rather than the one that actually happened. The sleep score for the fortnight is complete and confident and measured by an instrument whose specificity for detecting wakefulness has never exceeded 0.54. The only truthful sentence available to the coach is that they do not know, and the only reason they do not know is that six systems each did their own job perfectly.

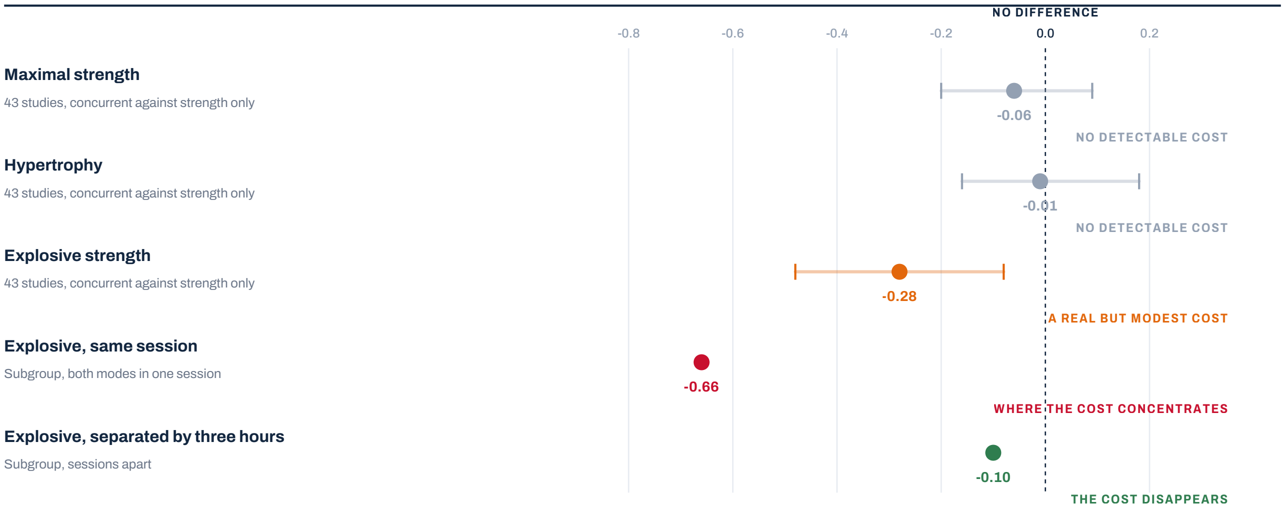
CHAPTER THREE

One adaptive system, three stimuli

The interference effect is real, narrow and widely misdescribed. Knowing which of those three is which changes how a week is built.

Triathletes are told that endurance training blunts strength adaptation and that strength training compromises endurance. Both claims have a basis, and both are stated far more broadly than the evidence supports. Getting this right matters because it determines whether the gym goes in the week at all.

THE INTERFERENCE EFFECT, AS THE META-ANALYSES REPORT IT



The interference effect is narrow, not general

It touches explosive strength in trained athletes, and it is removed by separating the sessions. Cycling and running did not differ.

FIGURE 2 Standardised mean differences for concurrent training against strength training alone, with confidence intervals where the analysis reports them.

Reference 1, including its subgroup analyses. REVIEW

3.1 What the meta-analyses actually found

An updated systematic review and meta-analysis of 43 studies compared concurrent training against strength training alone. For maximal strength the standardised mean difference was minus 0.06, with a confidence interval spanning zero. For muscle size it was minus 0.01, again spanning zero. Only explosive strength showed a real effect, at minus 0.28. The attenuation of explosive strength was significant when both modes were performed in the same session and not when they were separated by three hours or more, and there was no moderating effect of aerobic modality, frequency, training status or age.[1] REVIEW

A separate meta-analysis of 27 studies looking at lower-body maximal strength found a decrement in trained participants of minus 0.35, with the effect confined to same-session training at minus 0.66 against minus 0.10 when the sessions were separated.[2] REVIEW The two agree on the mechanism that matters practically, which is proximity rather than presence.

The modality claim deserves a paragraph of its own because it is so widely repeated. An influential 2012 meta-analysis reported that resistance training concurrent with running, but not cycling, produced significant decrements.[3] REVIEW A later analysis of concurrent high-intensity interval and resistance training found the opposite direction, with the negative trend in cycling rather than running.[4] REVIEW The largest and most recent synthesis found no moderating effect of modality at all.[1] REVIEW Anybody stating confidently that running interferes and cycling does not is repeating one paper and ignoring two.

THREE HOURS

The separation that removed the explosive-strength cost in the largest meta-analysis. Not a rest day. Three hours.

3.2 What this means for a triathlon week

Three things follow directly, and one follows from my own reading rather than from the papers. The first is that the gym belongs in the week for almost every triathlete, because the measured cost is confined to a quality most triathletes are not chasing. The second is that session ordering and spacing matter more than the presence or absence of strength work. The third is that a system reporting the week should treat proximity as a variable, because a strength session three hours after a hard run is a different stimulus from the same session the next morning.

The fourth, which is mine, is that the mechanism remains unsettled. A mechanistic review states that chronic training studies provide robust evidence that endurance exercise can attenuate hypertrophy and strength, while the underpinning of the interference effect is unknown and probably involves multiple integrated rather than isolated effectors.[5] COMMENTARY An unsettled mechanism is a reason for a system to present the trade-off rather than resolve it.

WHAT THIS RULES OUT

No system should remove strength work from a triathlon plan on the grounds of interference. The measured effect does not support it. A system may reasonably flag that a strength session and a hard endurance session sit within three hours of each other, and it should say which body of evidence it is drawing on when it does.

CHAPTER FOUR

Planning across horizons

Season, block, week and today are four different questions. Most software answers the fourth and pretends it has answered the first.

Planning in triathlon happens on four timescales that require different evidence and different degrees of confidence. The season is a set of choices about what to be good at and when. The block is a set of choices about which adaptation to prioritise. The week is a set of choices about ordering and spacing. Today is a set of choices about whether the plan survives contact with the athlete's actual state.

THREE ATHLETES, ONE SYSTEM, DIFFERENT ANSWERS

Elite short course

DRAFT-LEGAL RACING, FULL-TIME

- Swim front pack or lose the race
- Run off the bike is decisive
- Travel and time zones dominate the calendar

Long course age group

NINE HOURS A WEEK, A JOB, TWO CHILDREN

- Volume is capped by life, not by physiology
- Durability matters more than peak fitness
- Fuelling is the limiter nobody trains

Returning athlete

TWELVE WEEKS AFTER AN INTERRUPTION

- Availability is the whole objective
- Load must rise from a base nobody recorded
- Confidence is a training variable

The same session means three different things

A 90-minute ride is a recovery spin, a key aerobic session and an act of faith, depending on which of these three athletes is riding it. A system that cannot tell them apart is a spreadsheet with opinions.

FIGURE 3 Three athletes for whom the same session means three different things. The constraint that binds is different in each case, and it is rarely physiological.

Athleet.AI framework. The three cases are invented and are used throughout the paper. ATHLEET.AI

4.1 The intensity argument, honestly represented

No planning discussion in endurance sport survives long without the question of intensity distribution, and it is worth setting out where the evidence actually sits, because a great deal of software is sold on a slogan.

WHAT THE INTENSITY DISTRIBUTION EVIDENCE ACTUALLY SAYS

1 What athletes do

Descriptive reviews of highly trained and elite distance runners find a pyramidal distribution for most of the year, shifting toward polarised only in the competitive period. Marathon runners were more pyramidal, 1500 metre runners more polarised.

2 What the trials show

A network meta-analysis of individual participant data across 13 studies and 348 athletes found no difference between polarised and pyramidal for maximal oxygen uptake or time-trial performance. No difference was found between polarised and any other distribution.

3 What the measurement does

The same training, classified by time in zone, was 96.1 per cent easy. Classified by session goal, it was 86.6 per cent easy. The conversion factor in the high-intensity range is threefold, which is larger than most of the effects being argued about.

What this rules out

Any system that prescribes an 80/20 split as settled science is overstating the evidence. A distribution target is a defensible planning convention. It is not a finding, and a coach should be told which of the two they are being shown.

The useful question is not which distribution is correct. It is whether the system can tell you how it classified the week, because two defensible methods disagree by a factor of three.

None of this means intensity distribution does not matter. It means the labels are softer than the arguments about them, and an honest system says so.

FIGURE 4 Three findings that a system should be able to state before it recommends a distribution.

References 6, 7 and 8. REVIEW

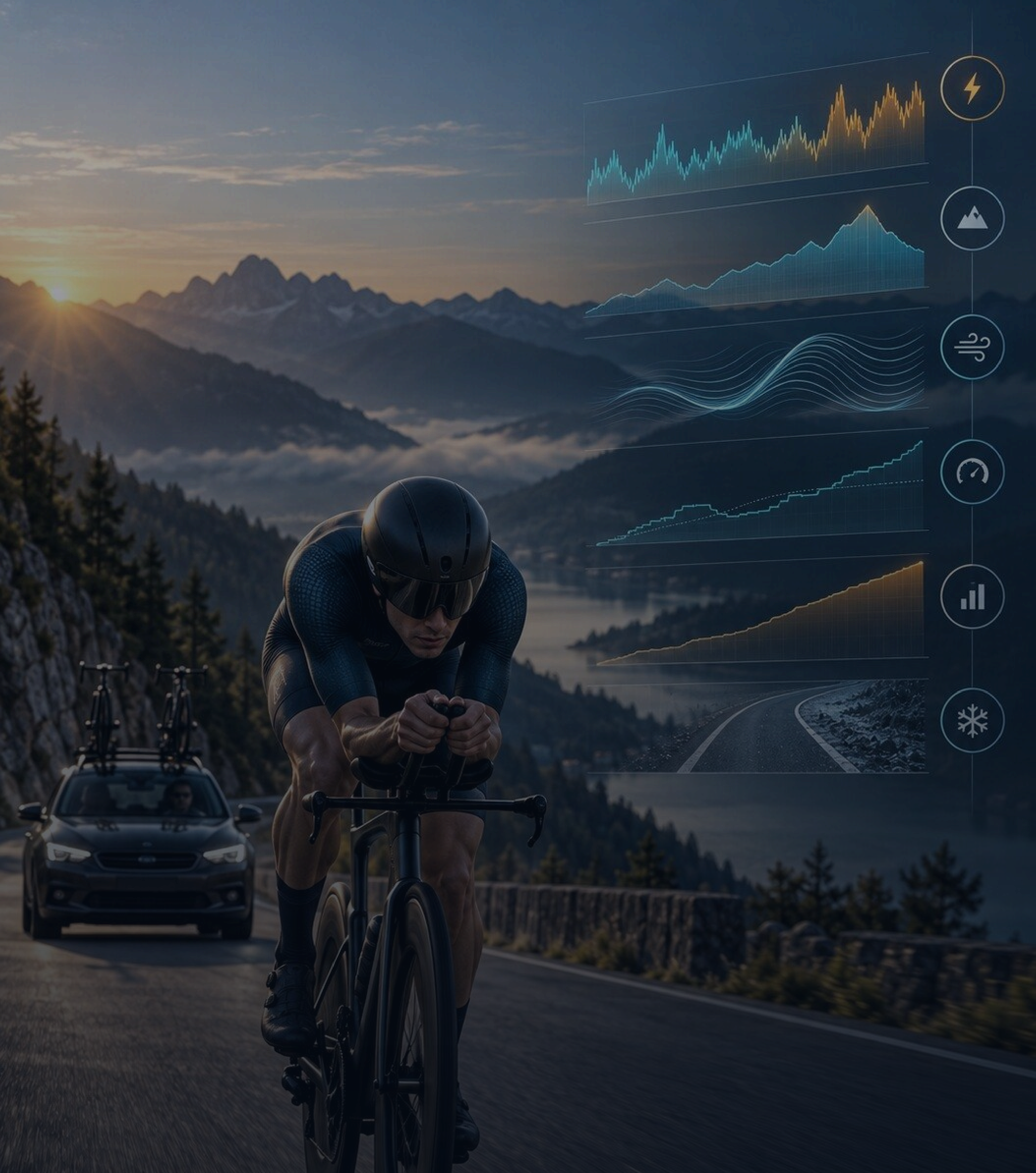
The descriptive literature is clear that pyramidal, not polarised, is what highly trained and elite distance runners actually do for most of the year, shifting toward polarised only in the competitive period, with marathon runners more pyramidal and 1500 metre runners more polarised.[6] REVIEW The interventional literature was pooled in 2025 at the individual participant level across 13 studies and 348 athletes, and found no difference between polarised and pyramidal for maximal oxygen uptake or time-trial performance, and no significant difference between polarised and any other distribution. A subgroup interaction by performance level was significant, suggesting competitive athletes may respond better to polarised and recreational athletes to pyramidal.[7] REVIEW Two of that analysis's co-authors are among the concept's original proponents, which makes the null unusually credible.

The methodological point undercuts both. Across 29 elite cross-country skiers, 570 sessions and one training camp, the same picture emerges twice. Classified by time in zone the training was 96.1 per cent easy. Classified by session goal it was 86.6 per cent easy. The conversion factor in the high-intensity range is threefold.[8] PEER-REVIEWED The same training, two defensible methods, a threefold difference in the quantity everybody is arguing about.

My own view is that a distribution target remains a useful planning convention and should be presented as exactly that. What a system must not do is present it as a finding. If it reports a distribution it should say which method produced it, because the reader cannot otherwise interpret the number.

4.2 What a planning system should actually offer

Four things, in descending order of how well they are supported. A record of what was planned and what was completed, which requires no model. A statement of how the week was classified and by which method. A projection of the competing demands in the block ahead, with the trade-off between them made explicit. And, last and most tentatively, a suggestion, presented with its evidence and its alternative.



PART TWO

What the data can and cannot say

Four chapters on the instruments, their error, and the four things a system should refuse to infer.

CHAPTER FIVE

Understanding readiness

The single readiness score is the most successful unvalidated product in endurance sport. This chapter sets out what would have to be true for it to be trustworthy.

Every major wearable now presents a composite number that claims to summarise whether an athlete is ready to train. It is the feature athletes talk about most and the feature with the least published support. I searched the peer-reviewed literature for validation of any commercial composite readiness or recovery score against a physiological or performance criterion, and found none. What exists is qualitative work on how athletes experience the scores.[9] PEER-REVIEWED One participant in that work put the limitation better than the literature does, observing that you cannot really capture the complexities of a human on a device.

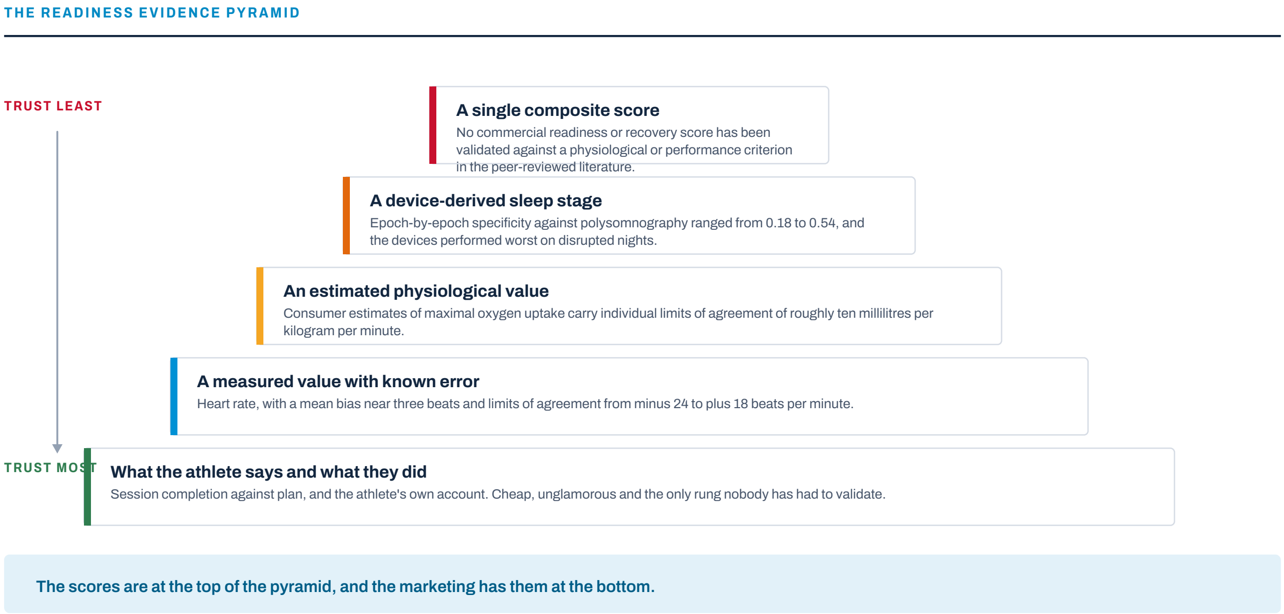


FIGURE 5 Five rungs of readiness evidence, ordered by how much of each has been validated. The scores sit at the top and are marketed as though they sat at the bottom.

Athleet.AI framework, informed by references 9, 10, 11 and 12. ATHLEET.AI

5.1 What the instruments actually do

The components of a readiness score are measured with more error than the interfaces suggest. A meta-analysis of 52 studies found a mean heart rate bias of about three beats per minute against reference measures, which sounds tolerable, and population limits of agreement from minus 24 to plus 18 beats per minute, which does not. For energy expenditure the mean bias was under three kilocalories per minute with limits spanning roughly 20.[10] REVIEW An earlier review across 158 publications and nine brands found heart rate more variable than step counting and concluded that for energy expenditure no brand was accurate.[11] REVIEW

Sleep staging is worse and matters more, because it is the component athletes weight most heavily. Seven consumer devices were compared against polysomnography across three nights including a disrupted-sleep condition. Sensitivity to sleep was at least 0.93 for every device, which is why the marketing looks good.

Specificity to wake ranged from 0.18 to 0.54, and the devices performed worse on nights with poorer or disrupted sleep.[12] PEER-REVIEWED A device that cannot reliably detect wakefulness, and that degrades exactly on the nights an athlete most wants information, is not a sound input to a decision.

Estimated physiological values carry the same pattern. Consumer estimates of maximal oxygen uptake from exercise-based algorithms had a mean bias near zero and individual limits of agreement of roughly ten millilitres per kilogram per minute, which the authors describe as adequate at population level while stating that the individual estimation error is large and needs improvement.[13] REVIEW A ten-unit band spans much of the difference between a competent age-grouper and a good one.

WHAT THE VALIDATION LITERATURE REPORTS

MEASURE	HOW IT LOOKS ON AVERAGE	WHAT IT DOES TO ONE ATHLETE ON ONE DAY
Heart rate, wrist optical	Mean bias near three beats per minute	Limits of agreement from minus 24 to plus 18 beats
Energy expenditure	Mean bias near three kilocalories per minute	No brand was found accurate; limits span 20 kilocalories per minute
Sleep staging	Sensitivity to sleep at least 0.93	Specificity to wake between 0.18 and 0.54, worst on disturbed nights
Maximal oxygen uptake, exercise-based	Mean bias near zero	Individual limits of agreement of about ten millilitres per kilogram per minute
Running power	One device repeatable to within 4.3 per cent	Devices disagree with each other; the construct is not shared
Functional threshold power	Within 1.4 per cent of the steady state on average	Individual agreement of about nine per cent, and contested as a threshold

A mean bias of three beats sounds tolerable until the limits of agreement are read. The average is a statement about a population. The athlete in front of you is not a population.

FIGURE 6 What the validation literature reports for the measures a triathlete sees daily, and what those figures do to a single reading on a single day. References 10 to 13, 20 and 21. Figures as reported by each study. REVIEW

5.2 Heart rate variability, which is the best-supported case

Guiding training by heart rate variability is the most seriously studied version of the readiness idea, and it is instructive because the evidence is real and modest. A methodological systematic review with meta-analysis found a significant effect on vagally mediated variability itself, at 0.50. For the outcomes an athlete cares about, the effects were smaller and their intervals crossed zero, with maximal aerobic capacity at 0.20 and endurance performance at 0.20. The authors' own summary is that if this approach is superior for producing group-level improvements, current data suggest it is only by a small margin.[14] REVIEW

The individual trials show the same shape. In 40 recreational runners, those guided by variability performed fewer hard sessions and improved 3000 metre performance where controls did not, with a small between-group effect.[15] PEER-REVIEWED Two cycling trials reported within-group improvements and explicitly no differences between groups.[16,17] PEER-REVIEWED The strongest positive result came from a composite rule combining nocturnal variability, perceived recovery and a heart-rate-to-speed index, where the individualised group improved a ten kilometre time by 6.2 per cent against 2.9 per cent, with more high responders and no low responders.[18] PEER-REVIEWED That last study is the most encouraging in the area and it is a composite decision rule in 30 runners, not a proprietary score in a watch.

WHAT THIS RULES OUT

No composite readiness score presented as a fact. No automatic modification of a session on the basis of one. Where a system uses such a score it should show the components, state the measurement error on each, and leave the decision with the athlete or the coach. The honest version of this feature is a prompt to look, not an instruction to change.

5.3 What to use instead

The bottom rung of the pyramid is the one nobody has had to validate, because it is not an inference. What the athlete did against what was planned, and what the athlete says about it. Session rating of perceived exertion has been examined more than most people assume, with a review finding 36 of 950 citing studies had examined its validity and reliability and concluding it holds across sports and age groups.[19] REVIEW Work in swimming found individual correlations with heart-rate-based methods from 0.55 to 0.94 and with distance from 0.37 to 0.81, and a significant interaction in which coaches under-rated low-intensity sessions and over-rated high-intensity ones relative to their athletes.[20] PEER-REVIEWED That last finding is worth more than it looks. The coach and the athlete are systematically disagreeing about the week, and a system that holds both is more useful than one that holds either.

CHAPTER SIX

Technique, efficiency and the numbers behind them

Running power is not one thing. Functional threshold power is contested. Both are useful anyway, provided nobody pretends they are measurements.

Endurance athletes have acquired a set of derived numbers that look like measurements and are not. Two deserve examination because they anchor so much training prescription.

6.1 Running power

Five commercial technologies were compared for repeatability and concurrent validity against oxygen consumption in endurance-trained runners across manipulated speed, weight and slope. One device was clearly the most repeatable, with a standard error of measurement at or below 12.5 watts and a coefficient of variation at or below 4.3 per cent, and had the best concurrent validity. The others retained some relationship with oxygen cost, but the authors concluded that their low repeatability questions their suitability.[21] PEER-REVIEWED

The deeper problem is conceptual. These devices disagree with one another, which means running power names several device-specific proprietary estimates rather than one shared construct. A study in well-trained runners found running economy correlated with the best-performing device's power at 0.6, and reported identical mean power at two cadences despite a significant change in the caloric cost of running, concluding that the foot pod may not be sufficiently accurate to estimate differences in running economy in competitive runners.[22] PEER-REVIEWED The device manufacturer supplied the equipment for that study, which is declared and worth knowing.

6.2 Functional threshold power

Cycling's central training anchor is genuinely contested and both sides are worth hearing. One validation study in 15 cyclists found the estimate close to the maximal lactate steady state on average, with a bias of 1.4 per cent, and concluded it is a usable practical estimate. The limits of agreement were 9.2 per cent, which on a rider at 270 watts is roughly plus or minus 25 watts.[23] PEER-REVIEWED A later study had cyclists ride at the estimate and 15 watts above it, found time to task failure of 33.7 against 22.0 minutes with oxygen uptake stabilising at both intensities, and concluded that it should not be considered a threshold marker between the heavy and severe domains.[24] PEER-REVIEWED

Both can be true. A quantity can be a useful practical anchor for prescribing training and a poor physiological threshold. The failure mode is a system that treats it as the second while using it as the first, and then reports zone compliance to two decimal places.

NOT A SHARED UNIT

Two runners quoting running power are not necessarily quoting the same quantity. A system that pools power across devices is adding numbers that do not share a definition.

6.3 Swimming, which is the least instrumented

Swimming is where measurement is thinnest and where most triathletes lose the most time. Systematic reviews of inertial sensor technology for swimming have examined feature detection for starts, turns and free swimming and the consequences of sensor placement.[25,26] REVIEW What does not exist is more interesting. I searched for validated work on the biomechanical or energetic cost of sighting in open water and found none. Everything on sighting is coaching literature. A skill that every open-water triathlete performs dozens of times per race, and that visibly disrupts the stroke, has not been studied.

Wetsuits, by contrast, are well characterised. A systematic review of 26 studies found full-body wetsuits produced velocity increments of 3.2 to 12.9 per cent across distances from 25 to 1500 metres, with higher stroke rate, stroke length and stroke index and lower energy cost.[27] REVIEW That is a large and reliable effect, and it means a swim split without a note of what was worn and what the water temperature was is close to uninterpretable.

CHAPTER SEVEN

Strength, durability and injury

The gym is the most under-recorded part of a triathlete's week and one of the best supported. Both facts have the same cause.

Chapter 3 dealt with whether strength work costs anything. This chapter deals with what it buys, and with the injury problem it is often supposed to solve.

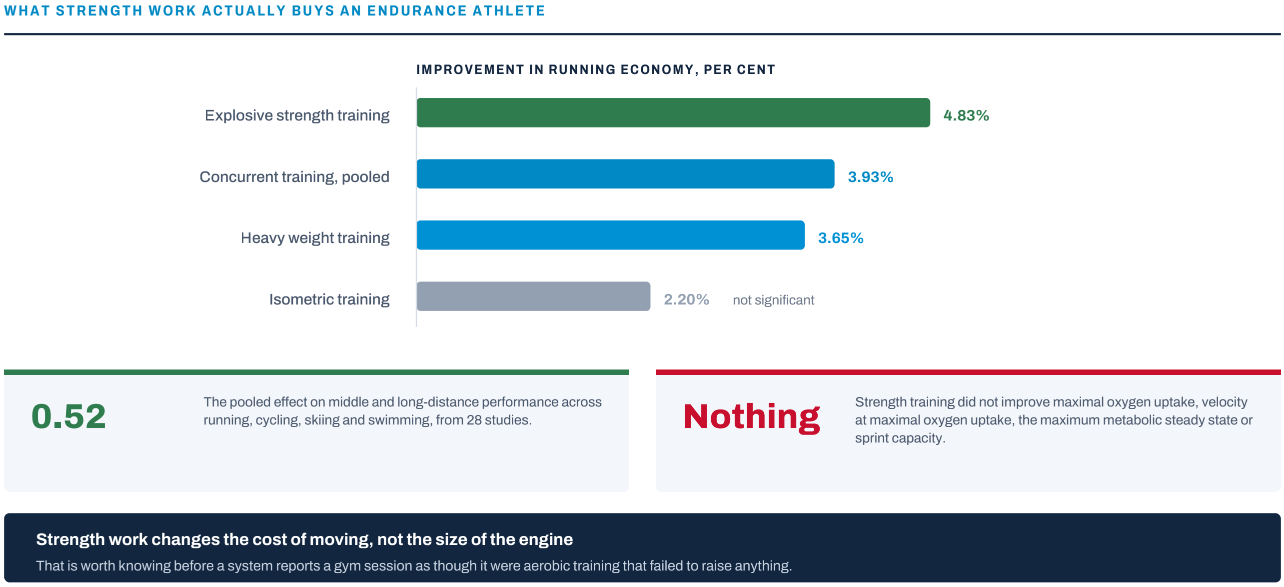


FIGURE 7 Improvements in running economy by training method, with the pooled performance effect and the outcomes that did not move. References 28, 29 and 30. REVIEW

7.1 What it buys

A meta-analysis of 16 studies found concurrent training improved running economy by 3.93 per cent, explosive training by 4.83 per cent and heavy weight training by 3.65 per cent, while isometric training produced no significant change. Longer interventions predicted greater improvement.[28] REVIEW A meta-analysis across running, cycling, cross-country skiing and swimming found a net standardised effect of 0.52 on middle and long-distance performance, with maximal force training outperforming other intensities and benefits consistent irrespective of athlete level.[29] REVIEW

The most recent and most carefully graded synthesis, covering 894 athletes, found moderate effects for high-load training and large effects for combined methods, with plyometric training not reaching significance. Its most useful finding is the negative one. None of the methods improved maximal oxygen uptake, velocity at maximal oxygen uptake, the maximum metabolic steady state or sprint capacity, and the certainty of evidence was graded from very low to moderate.[30] REVIEW

The practical reading is that strength work changes the cost of moving rather than the size of the engine. That is worth knowing before a system reports a gym block as aerobic training that failed to raise anything, which is what happens when the strength work and the endurance work live in different applications with different assumptions.

7.2 The injury picture

Triathlon injury epidemiology is better than it was and still uneven. A 26-week prospective cohort of 174 entrants to a long-distance event found an average prevalence of overuse problems of 56 per cent and substantial overuse problems in 20 per cent, with the knee at 25 per cent and the lower leg and lower back at 23 per cent each. Acute injury incidence was about one per 1,000 hours in both training and competition.[31] PEER-REVIEWED

A four-year prospective study of 50 elite short-course athletes recorded 266 injuries across 46 athletes, with 67.3 per cent causing time loss and 70.7 per cent training-related. Bone stress injuries were the most burdensome type at 31.4 days lost per 365 days, 40 per cent of athletes reported at least one, and female athletes had roughly three times the rate of bone stress injury of their male counterparts. Overall injury incidence did not differ by sex.[32] PEER-REVIEWED That last figure is the one this paper returns to in Chapter 8, and it should shape what any system is willing to say and what it refuses to infer.

A systematic review of long-distance triathlon injuries screened 975 studies and included six, of which two were prospective, and reported an overuse injury incidence range of 37 to 91 per cent.[33] REVIEW A range that wide is close to uninformative, and saying so is more useful than quoting its midpoint. The wider review of triathlon health noted that the literature has failed to consistently report age group, sex, ability level and event-distance specialisation, and that the demands of training and racing are relatively unquantified.[34] REVIEW

SPECIFICATION

What a system should do about injury, and what it should not

- **Record availability, not risk.** Weeks completed against weeks planned, by discipline, is countable and auditable.
- **Show the sequence.** What preceded an interruption is more useful to a coach than a probability attached to the next one.
- **Flag the known patterns.** A sustained rise in running volume in an athlete with a bone stress history is a prompt for a conversation, presented as such.
- **Refuse individual prediction.** The flagship paper in this series sets out why, and the endurance literature offers no exception.

CHAPTER EIGHT

Nutrition, hydration and environment

This chapter contains the hardest line in the paper. Energy availability is not something software may infer, and the reason is in the consensus statement itself.

Fuelling decides long-course races and environment decides some of them outright. Both are areas where useful decision support is possible and where the boundary of what a system may say is unusually clear.

WHAT THE ENVIRONMENT CONSENSUS ACTUALLY SPECIFIES

Heat acclimation

38.5 to 39.8
degrees core temperature, held for at least 60 minutes, is what counts as a session

One to two weeks
of training in the heat improves performance in the heat

Biweekly
exposures maintain the adaptation once it is induced

Hydration

Under 1 to 2
per cent daily body mass change, as the practical guide

Under 1.020
urine specific gravity

6 ml/kg
every two to three hours before individual data exist

Carbohydrate

60 g/h
the oxidation ceiling for a single carbohydrate source

90 g/h
achievable only with multiple transportable carbohydrates

Not by weight
the guidance is independent of body mass and training status

Where a system must stop
Exertional heat illness is a medical emergency with a defined threshold and a defined response. No performance tool should be the thing that decides whether an athlete has one.

FIGURE 8 The practical numbers the consensus documents actually give, and the point at which a performance tool must stop.

References 35, 36, 37 and 38. Figures as stated in each source. CONSENSUS

8.1 Carbohydrate, where the numbers are firm

A single carbohydrate source can be oxidised at rates up to approximately 60 grams per hour, which is the recommendation for exercise of two to three hours. For ultra-endurance events the recommendation is higher at approximately 90 grams per hour, and carbohydrate ingested at such rates must be multiple transportable carbohydrates to allow high oxidation rates and prevent accumulation in the intestine. The advice is independent of body weight and training status, though it should be adjusted downward at low absolute intensity.[36] REVIEW A parallel review puts 30 to 60 grams per hour as the target for sports of longer duration, with events beyond two and a half hours benefiting from intakes up to 90.[37] REVIEW

On whether any of this can be individualised algorithmically at present, the honest answer is not yet. A recent review of carbohydrate research in endurance athletes calls for a much more personalised approach that could be further supported through future research and technological innovation, naming continuous glucose monitoring as an example.[38] REVIEW That is a research aspiration stated as one. A product presenting individualised carbohydrate targets today is ahead of its evidence, and should say so.

8.2 Heat, where the numbers are practical

The environmental consensus is unusually specific and unusually usable. A heat acclimation session is defined as exercise in the heat with core body temperature between 38.5 and 39.8 degrees for at least 60 minutes, with elevated skin temperature, sweating and skin blood flow. Performance in the heat is increased after one or two weeks of training in the heat. Decay is slower than induction, so an athlete unable to acclimate in the fortnight before a race can acclimate about a month earlier and re-acclimate for a few days, maintaining the adaptation with biweekly exposures. For hydration, daily body-mass change should remain under one to two per cent and urine specific gravity under 1.020, with a general guide before individual data exist of six millilitres per kilogram every two to three hours.[35] CONSENSUS A meta-analysis of 96 articles found regimens most commonly lasting seven to 14 days, a moderately beneficial effect on capacity and performance in the heat irrespective of regimen, and greater adaptation from longer regimens.[39] REVIEW

All of that is decision support a system can legitimately provide, because it is a published protocol with stated thresholds. What sits on the other side of the line is illness. Exertional heat stroke is defined as extreme hyperthermia above 40.5 degrees with central nervous system dysfunction, and is a medical emergency requiring immediate whole-body cold-water immersion.[35] CONSENSUS No performance tool should be the thing that decides whether an athlete has one.

8.3 The hard line, and why it is drawn here

Relative energy deficiency in sport is the area where an endurance product is most tempted to be helpful and where it must not be. The 2023 consensus statement introduced updated conceptual models and a clinical assessment tool for detection and diagnosis.[40] CONSENSUS The companion paper describing that tool is explicit that the condition has various risk factors, numerous signs and symptoms, is heavily influenced by environment, and that accordingly there is no singular validated diagnostic test. It specifies a three-step process ending in a physician-led final diagnosis and treatment plan developed with the athlete, coach and the wider health and performance team.[41] CONSENSUS

The measurement side is equally clear. A methodological critique by the consensus's own lead nutrition author sets out that there is no universal protocol for calculating energy availability, that estimating each component consumes resources and carries residual error, and that there is no clear definition of the value considered low.[42] COMMENTARY

Put those together. Energy availability is a ratio of three quantities, each estimated with substantial error in free-living athletes, there is no agreed calculation protocol, there is no agreed threshold, and the issuing body states there is no single validated diagnostic test and reserves the final judgement for a physician. An algorithm operating on wearable-derived expenditure and self-logged intake would be inferring an unmeasurable quantity against an undefined threshold, and substituting itself for the step the consensus reserves for a doctor. That is the argument, and it is made entirely from the consensus's own documents.

WHAT THIS RULES OUT

No system should score, flag, estimate or infer energy availability, low energy availability or relative energy deficiency in sport, for any athlete, at any level. No system should recommend a change in intake on the basis of an estimated deficit. A system may record what an athlete logs, may surface published guidance, and may make it easy to involve a qualified practitioner. That is the whole permitted range, and the female bone stress finding in Chapter 7 is why the line is drawn tightly rather than generously.

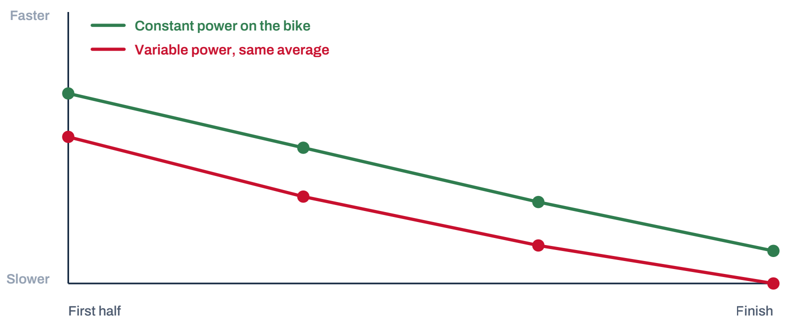
CHAPTER NINE

Race modelling and execution

One finding in triathlon pacing is well quantified and directly actionable, and most software throws away the data that would let you use it.

Race execution in triathlon is the point at which every earlier decision is tested, and it is the area where the evidence is most uneven. Almost all of it concerns Olympic distance. Both a foundational pacing review and a triathlon-specific one state plainly that knowledge of optimal pacing in ultra-endurance events is lacking and that little is understood about the specific physiological, environmental and interdisciplinary effects on pacing in longer races.[43,44] REVIEW Anybody selling a long-course pacing model should be asked which literature it rests on.

WHAT THE BIKE DOES TO THE RUN



42s

slower over the run after an hour of variable cycling at the same average power, in 12 well-trained triathletes.

Thirty-five of those seconds appeared in the first half of the run, before the athlete had settled.

This is the most actionable finding in triathlon pacing

The same average power produces a different race depending on how it was distributed. Any system reporting a ride by its average alone has discarded the part that mattered.

Illustrative shape drawn to represent the measured direction and the reported magnitudes. The underlying study reports total and first-half decrements rather than a continuous curve.

FIGURE 9 What an hour of variable cycling does to the run that follows it, at identical average power. The curve is illustrative; the magnitudes are measured.

Reference 45. The study reports total and first-half decrements rather than a continuous curve, and the shape here represents the reported direction. PEER-REVIEWED

9.1 The finding worth building around

In 12 well-trained male triathletes, two one-hour cycling trials were completed at the same mean power, one with power varying between 40 and 140 per cent of maximal aerobic power and one constant, each followed by a 9.3 kilometre outdoor time-trial run. The run was 42 seconds slower after variable cycling, and 35 of those seconds appeared in the first half. Higher post-cycle lactate and perceived exertion correlated moderately with a larger decrement, and there was no association between a laboratory power-profile test and the size of the decrement.[45] PEER-REVIEWED

Three things follow. The first is that average power is an inadequate summary of a triathlon bike leg, and a system that reports a ride by its average alone has discarded the part that mattered. The second is that variability is a measurable, coachable property of a ride that most athletes never see quantified. The third, and this is the one I would put in front of a supplier, is that the laboratory test did not predict who would suffer. The effect is real at group level and individual susceptibility remains unpredicted, which is a familiar pattern in this paper.

Observational work in draft-legal racing supports the same direction. In 24 elite men, athletes in the bottom half after the swim cycled faster at 13.4 kilometres, speed at that point was inversely correlated with running performance, and run rank correlated with overall result at 0.86.[46] PEER-REVIEWED The bike is where the race is lost and the run is where it is scored.

ILLUSTRATIVE SCENARIO

The rider who did everything right and ran badly

An invented age-group athlete rides a hilly half-distance course at exactly the average power she trained for and runs 11 minutes slower than her standalone half marathon. Her platform reports the ride as a successful execution, because the average matched the target. What the file also contains, and what nothing surfaces, is that she spent 22 minutes above her threshold across 14 separate climbs, and that the same average power on a flat course would have been a different ride entirely. The number she was given to hit was the wrong number, and the system that gave it to her had the data to know.

9.2 What a race model should and should not do

It should hold the course, the expected conditions and the athlete's own history of executing similar demands. It should express a plan in terms of variability as well as average. It should state contingencies in advance, which is where most race-day decisions are actually made, and it should record what was done against what was planned so that the next race starts from evidence rather than memory. It should not present a predicted finish time as a target, and it should not adjust the plan during the race, both because most governing rules restrict communication devices during competition and because the athlete is in no position to interrogate an instruction at that moment.

CHAPTER TEN

Coached and self-coached workflows

The same output is safe in one workflow and unsafe in the other. Most products are built for the workflow that has a professional in it, and sold to the one that does not.

Triathlon's population is unusual in that most of it is self-coached, and the design consequence has been almost entirely ignored. A recommendation shown to a coach enters a process. The coach interrogates it, discounts it against what they saw on the pool deck, and decides. A recommendation shown to a self-coached athlete at ten o'clock at night enters no process at all. It is the only voice in the room.

FIGURE 10 What a system would have to hold to reason about a triathlon week. The bottom row is what determines most weeks and what almost no product records.

Athleet.AI framework. ATHLEET.AI

10.1 Two settings, two designs

My own view, offered as practice, is that a product intended for self-coached athletes should be more conservative than one intended for coaches, and that the industry has it backwards. The coached product can afford to surface a tentative pattern, because a professional will filter it. The self-coached product should surface fewer things, state uncertainty more plainly, and be markedly harder to persuade into a definite instruction.

The escalation route matters more in the self-coached setting for the same reason. A system that has noticed a pattern it cannot interpret should be able to say so and to make it easy to involve somebody qualified, rather than producing a confident answer because a confident answer is what the interface expects.

10.2 What language models can and cannot be trusted with

Large language models are now writing training plans, and the evaluations that exist are worth knowing. A critical evaluation of a leading model asked to produce 30-day exercise programmes for five hypothetical profiles found that the prescriptions lacked precision in addressing individual health conditions and goals, often prioritising excessive safety over the effectiveness of training, and concluded that these technologies are not yet recommended as a substitute for personalised, progressive and condition-specific prescription by qualified professionals.[47] COMMENTARY A comparison of two models asked to design 12-week resistance programmes found all derived programmes objectively similar regardless of sex, and concluded that plans likely require additional refinement before application.[48] COMMENTARY A larger expert-rated evaluation found the overall quality of generated plans moderate, with small differences between models, and stated that expert supervision remains essential.[49] COMMENTARY

Every one of those evaluations concerns resistance training. I searched for a peer-reviewed evaluation of language-model-generated endurance, running, cycling or triathlon plans and found none published before this paper was written. Language models are writing endurance plans for large numbers of athletes on the basis of no published evaluation whatsoever. That is a finding rather than an omission, and it is the single most useful sentence in this paper for anybody assessing a product.

What these models are good at is the part nobody demonstrates. Drafting a session note from a coach's voice memo. Summarising a fortnight into something a busy athlete will read. Translating a plan into the athlete's own language. Retrieving what the coach said in March. All of that is real value, none of it requires the model to be right about physiology, and all of it is auditable.

A GAP WORTH STATING

Every published evaluation of language-model training plans concerns resistance training. None concerns endurance. The plans are being written anyway.

CHAPTER ELEVEN

Data ownership and the rules of racing

The competition rules of this sport contain a data clause that almost nobody in it has read.

Two governance questions bear on endurance technology and one of them is specific to triathlon. The general question is the one the flagship paper in this series addresses, being who holds an athlete's record, who may train models on it and what an athlete receives when they leave. The specific question concerns what the sport's own rules already say.

11.1 What the competition rules say

The international competition rules in force when this paper was written contain an intellectual property clause stating that all personal data from an athlete entered into any event, and the data produced from the event including the results, are the exclusive property of the governing body. The same clause claims rights over digital media captured within the field of play regardless of who owns the capturing device.[50] CONSENSUS I have not seen that clause discussed anywhere in the sports technology literature, and I think athletes and coaches would be interested to know it exists.

My own reading, which an athlete or federation should verify with their own adviser rather than accept from this paper, is that a claim of exclusive property in personal data sits awkwardly alongside data protection rights that cannot be assigned away. Stating that is not an accusation. It is an observation that the sport's commercial instruments and its data protection obligations were written by different people at different times, and that nobody appears to have reconciled them.

The same rules restrict what an athlete may carry. Communication devices may not be used in a distractive manner during competition, on pain of disqualification. Headphones and technical earplugs are prohibited in all three disciplines. Watches are not permitted in elite, under-23 and junior racing at standard distance and shorter, subject to the head referee's discretion, and any device added to the bike is subject to approval before the competition.[50] CONSENSUS Any product proposing in-race guidance should establish which of those rules apply to its users before building the feature rather than after.

11.2 What an athlete should be able to expect

Four things, none of which requires new law and none of which the market currently delivers. A record that travels, so that an athlete changing coach or platform takes their own history with them in a usable form. A stated position on secondary use, so that an athlete knows whether their data trains models they do not own. An explanation on request, so that a recommendation can be interrogated after the fact. And a retention position, so that somebody has decided what happens to a decade of training data when the athlete stops.

The flagship paper sets out why these four are the practical frontier and why the law is not the obstacle. In endurance sport there is an additional argument, which is that the athlete usually bought the devices. An age-group triathlete has typically paid for the watch, the head unit, the power meter and the subscription, and it would surprise most of them to learn how little of the resulting record they can take away.

CHAPTER TWELVE

Integration architecture

Five layers. The market sells the first and the fourth, and the value is in the second, third and fifth.

The architecture that would actually serve a triathlete is not technically difficult and it is commercially unattractive, which is a combination that explains a great deal about the current market.

CONNECTED ENDURANCE DATA ARCHITECTURE

Capture
Watches, head units, power meters, indoor platforms, gym logs, questionnaires, the coach's own notes.

Identity
One athlete, one identifier, across three sports and every device that has ever touched them.

Reconciliation
Where two sources disagree, which they will, and the rule that decides what the record says.

Interpretation
Discipline-aware analysis that states its own uncertainty and names the rung it is standing on.

Decision and record
What changed, who decided, on what evidence, and what happened next.

Almost every product in this market sells the first layer and the fourth. The value sits in the second, third and fifth, which are unglamorous and which nobody demonstrates.

Reconciliation is the layer that decides whether a coach ever trusts the record.

FIGURE 11 Five layers, of which the second, third and fifth determine whether a coach ever trusts the record.

Athleet.AI framework. ATHLEET.AI

12.1 The layer nobody sells

Reconciliation is the decision about what the record says when sources disagree. It sounds like plumbing and it is the difference between a record a coach will reason from and a record a coach will abandon. The rules it requires are unglamorous and specific. Which source wins for duration. Which wins for power. What happens when a session exists in one system and not another. How a manually entered gym session is weighted against a device-captured ride. Whether a corrected file replaces the original or sits beside it.

None of that requires artificial intelligence. All of it requires somebody to decide, and to write the decision down where an athlete can see it. My own view is that a supplier who cannot describe their reconciliation rules in a paragraph has not made them, and that this is the single most diagnostic question a buyer can ask.

12.2 Identity, which is the prerequisite

One athlete, one identifier, across three sports and every device that has ever touched them. That sounds trivial and is where most integrations fail, because devices and platforms each maintain their own idea of who the athlete is and none of them is authoritative. Until identity is settled, every number downstream is an assertion about somebody who may not be the same person.

SPECIFICATION

Five questions for a technology partner

- What are your reconciliation rules when two sources disagree, and where can an athlete read them?
- How is athlete identity established and maintained across devices and platforms?
- Which numbers in the interface are measured, which are derived, and which are estimated by a proprietary model?
- What error is reported alongside each derived number, and where does that error figure come from?
- What does an athlete receive on leaving, in what format, and does it include the derived record?

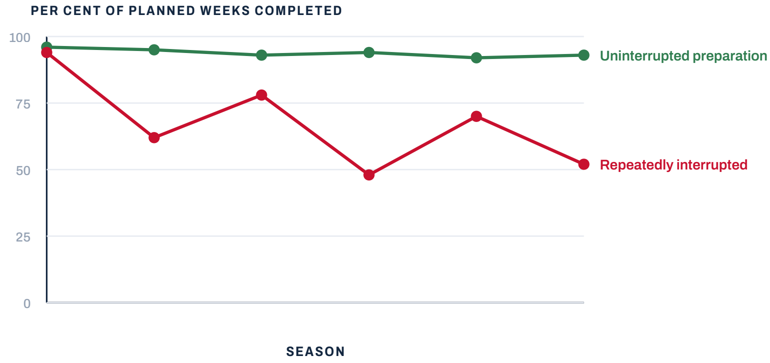
CHAPTER THIRTEEN

The connected performance cycle

What to adopt, in what order, and what to measure to find out whether it worked.

The adoption sequence that follows from this paper is deliberately modest, because the ambitious version does not survive contact with a coach who has 30 athletes and a job.

CONTINUITY IS THE OUTCOME WORTH MEASURING



Who this matters to

The athlete
More weeks of the training they intended to do.

The coach
A record that explains a plateau rather than guessing.

The event
Entrants who reach the start line.

The sponsor
A career long enough to be worth investing in.

The curve is illustrative. The measure is not.

Completed planned weeks is countable from records a coach already keeps, and it is the measure most closely linked to performance goals in the published prospective work.

FIGURE 12 Completed planned weeks across a season, and who benefits when the line stays high. The curve is illustrative; the measure behind it is countable.

Athleet.AI framework. The link between training availability and performance-goal success is established in reference 51. ATHLEET.AI

13.1 The measure

Completed planned weeks is the measure this series keeps returning to, and endurance sport has the cleanest evidence for it. In 33 international track and field athletes monitored weekly across 76 athlete-seasons, those completing more than 80 per cent of planned training weeks were seven times more likely to achieve a performance goal, at an area under the curve of 0.72, and training availability accounted for 86 per cent of successful seasons.[51] PEER-REVIEWED That is a different sport and a different population, and it should be read as an indication rather than a transferred result. It is still the best-supported thing a triathlon coach can put on a page.

13.2 The sequence

First, one record. Reconcile the three disciplines and the gym against one athlete identity, and write down the reconciliation rules. Nothing else works until this does. Second, error alongside every derived number, which mostly means changing what the interface says rather than what it computes. Third, remove the composite score, or demote it to a prompt with its components visible. Fourth, start recording decisions alongside recommendations, so that in a season's time there is something to learn from. Fifth, and only fifth, consider prediction, and expect to conclude that you do not need it.

13.3 What good looks like in three years

An athlete who can see why a block was built the way it was. A coach who can explain a plateau from evidence rather than from memory. A record that survives a change of watch, a change of coach and a change of platform. Fewer weeks lost to interruptions nobody saw coming, and more of the training that was actually intended. And a system that is markedly less confident than the one it replaced, which the athletes will notice and the marketing department will not enjoy.

The useful system is the one that holds the whole week in one record and is honest about what it cannot see.

CHAPTER ONE

END MATTER

Athlete and coach checklist

Eighteen items to work through before adopting anything, or before trusting what you already have.

- [ ] **One athlete identity across every source**
  Established and maintained, not assumed from a matching name.
- [ ] **Reconciliation rules written down**
  Which source wins for duration, for power, and what happens to a corrected file.
- [ ] **Strength work in the same record**
  Not in a separate application and not in a notebook.
- [ ] **Life recorded as part of the week**
  Travel, work pattern and illness, because they move more plans than physiology does.
- [ ] **Measured, derived and estimated distinguished**
  In the interface, not in a support article.
- [ ] **Error reported alongside derived numbers**
  And the source of that error figure available on request.
- [ ] **No composite readiness score presented as fact**
  Components visible, or the score removed.
- [ ] **No automatic plan rewriting**
  A person approves any change to what the athlete is asked to do.
- [ ] **No inference about energy availability**
  None, at any level, for any athlete.
- [ ] **No individual injury risk output**
  Availability against plan instead, which is countable.
- [ ] **Intensity classification method stated**
  Because two defensible methods differ by a factor of three.
- [ ] **Bike variability reported, not only average**
  The finding in Chapter 9 depends on it.
- [ ] **Swim conditions recorded**
  Wetsuit and water temperature, or the split cannot be interpreted.
- [ ] **Competition rules checked before in-race features**
  Devices, watches and communication are restricted in several categories.
- [ ] **Secondary use position stated**
  Whether athlete data trains models the athlete does not own.
- [ ] **Export includes the derived record**
  Not only the raw files, in a documented format.
- [ ] **Escalation route exists**
  The system can say it does not know and make it easy to ask somebody who might.
- [ ] **Completed planned weeks is the headline measure**
  Reported ahead of any score, any estimate and any prediction.

END MATTER

Glossary

Concurrent training
Endurance and strength training performed within the same programme, and the source of the interference argument in Chapter 3.

Critical power
The asymptote of the power-duration relationship, separating intensity domains in which physiological responses can or cannot be stabilised.

Energy availability

Energy intake minus exercise energy expenditure, relative to fat-free mass. Estimated with substantial error and reserved for clinical judgement.

Functional threshold power

A practical cycling anchor derived from a sustained effort. Useful for prescription and contested as a physiological threshold.

Heat acclimation

Adaptation induced by repeated exercise in the heat, with a defined core temperature range and duration in the consensus documents.

Heart rate variability

Beat-to-beat variation used as a marker of autonomic state. The best-studied readiness input and still modest in effect.

Interference effect

The attenuation of strength or endurance adaptation by concurrent training. Real, narrow, and dependent on session spacing.

Limits of agreement

The range within which most individual differences between a device and a reference measure fall. Far wider than mean bias.

Multiple transportable carbohydrates

Combinations using more than one intestinal transporter, allowing oxidation rates above those of a single source.

Polarised and pyramidal

Two descriptions of how training time is distributed across intensity zones. Both defensible, and not distinguishable in the pooled trial evidence.

Reconciliation

The rules deciding what the record says when two sources disagree. The least glamorous and most consequential layer in the architecture.

Session rating of perceived exertion

Perceived effort multiplied by duration. Cheap, examined more than most assume, and systematically different between coach and athlete.

END MATTER

Method and evidence grading

How the sources in this paper were selected, checked and graded, and what the reader should distrust.

Sources were gathered from the peer-reviewed literature, consensus statements, governing-body regulation and official national statistics, with priority given to systematic reviews, meta-analyses and methodological criticism over single primary studies. Every reference in the list that follows was checked against the publishing journal, publisher or issuing body, and then checked a second time by an independent process against the same sources. Where two sources give different page ranges for the same article, the discrepancy was resolved against the issue itself.

What could not be verified

Four things were searched for and not found, and their absence shapes this paper. No systematic review of machine learning applied to triathlon was located. No peer-reviewed evaluation of language-model-generated endurance training plans was located, in contrast to several for resistance training. No validation of any commercial composite readiness or recovery score against a physiological or performance criterion was located. And no peer-reviewed study of sighting in open-water swimming was located, despite it being a skill every open-water athlete performs repeatedly.

Several sources are cited with limitations recorded here. The joint nutrition position stand is cited for its own statement about referral rather than for its intake tables, which the author was unable to open. The heat adaptation meta-analysis is cited for the direction and duration of its findings rather than for effect magnitudes, which its abstract gives only as verbal descriptors. Two inertial sensor reviews in swimming are cited for their scope rather than for pooled accuracy figures. And the national participation figure is for England rather than the United Kingdom, and sits at the edge of that survey's resolution.

What to distrust in this paper

Three things. The readiness pyramid in Chapter 5, the context model in Chapter 10 and the architecture in Chapter 12 are the author's own frameworks, proposed rather than validated, and are published in a form designed to invite disagreement. The pacing curve in Chapter 9 is an illustrative shape drawn to represent measured magnitudes, and the caption says so. And the continuity curve in Chapter 13 is illustrative, with only the underlying measure and the cited relationship taken from published work in a different sport.

Conflict of interest

The author is the founder of Athleet.AI, which builds software in the area this paper examines. The paper recommends against several categories of product feature that would be commercially straightforward to build and sell into endurance sport, including the one endurance athletes ask for most. Readers should weigh the argument accordingly. Test the citations rather than the author.

Three conflicts in the cited literature are declared by their authors and are repeated here because they bear on interpretation. One wearable validation review includes an author employed by a device manufacturer during publication. One running power study was supplied equipment by the manufacturer. One variability meta-analysis includes an author declaring fees from technology companies in the same area.

END MATTER

About the author

Paul Forrest is the founder and Chief AI Officer of Athleet.AI. He works with organisations on the deployment of artificial intelligence, having begun in business process reengineering in the 1990s and moved into natural language and machine learning work from 2014, building small, domain-specific models in preference to general-purpose ones.

He is a qualified athletics coach, holding level three qualifications in sprints, hurdles and endurance, and he coaches as a volunteer at club level. That combination of enterprise systems work and time on a track is the reason this series exists, and it is also the reason the papers decline to recommend several things that would be straightforward to sell.

He is the author of *Tales from AI Development Hell* and *AI Made the Decision: Who Answers When Nobody Owns the Outcome?*, and holds an adjunct professorship teaching digital acumen and digital entrepreneurship.

END MATTER

References

Fifty-two sources, each verified twice against the publishing journal, publisher or issuing body.

Concurrent training and the interference effect

1. **Schumann, M., Feuerbacher, J. F., Sünkeler, M., Freitag, N., Rønnestad, B. R., Doma, K., & Lundberg, T. R.** (2022). Compatibility of concurrent aerobic and strength training for skeletal muscle size and function. An updated systematic review and meta-analysis. *Sports Medicine*, 52(3), 601–612. doi:10.1007/s40279-021-01587-7
2. **Petré, H., Hemmingsson, E., Rosdahl, H., & Psilander, N.** (2021). Development of maximal dynamic strength during concurrent resistance and endurance training in untrained, moderately trained, and trained individuals. A systematic review and meta-analysis. *Sports Medicine*, 51(5), 991–1010. doi:10.1007/s40279-021-01426-9
3. **Wilson, J. M., Marin, P. J., Rhea, M. R., Wilson, S. M. C., Loenneke, J. P., & Anderson, J. C.** (2012). Concurrent training. A meta-analysis examining interference of aerobic and resistance exercises. *Journal of Strength and Conditioning Research*, 26(8), 2293–2307. doi:10.1519/JSC.0b013e31823a3e2d
4. **Sabag, A., Najafi, A., Michael, S., Esgin, T., Halaki, M., & Hackett, D.** (2018). The compatibility of concurrent high intensity interval training and resistance training for muscular strength and hypertrophy. A systematic review and meta-analysis. *Journal of Sports Sciences*, 36(21), 2472–2483. doi:10.1080/02640414.2018.1464636
5. **Coffey, V. G., & Hawley, J. A.** (2017). Concurrent exercise training. Do opposites distract? *The Journal of Physiology*, 595(9), 2883–2896. doi:10.1113/JP272270

Intensity distribution and its measurement

1. **Casado, A., González-Mohíno, F., González-Ravé, J. M., & Foster, C.** (2022). Training periodization, methods, intensity distribution, and volume in highly trained and elite distance runners. A systematic review. *International Journal of Sports Physiology and Performance*, 17(6), 820–833. doi:10.1123/ijspp.2021-0435
2. **Rosenblat, M. A., Watt, J. A., Arnold, J. I., Treff, G., Sandbakk, Ø. B., Esteve-Lanao, J., et al.** (2025). Which training intensity distribution intervention will produce the greatest improvements in maximal oxygen uptake and time-trial performance in endurance athletes? A systematic review and network meta-analysis of individual participant data. *Sports Medicine*, 55(3), 655–673. doi:10.1007/s40279-024-02149-3
3. **Sylta, Ø., Tønnessen, E., & Seiler, S.** (2014). From heart-rate data to training quantification. A comparison of 3 methods of training-intensity analysis. *International Journal of Sports Physiology and Performance*, 9(1), 100–107. doi:10.1123/IJSPP.2013-0298

Readiness, wearables and measurement error

1. **Ibrahim, A. H., Beaumont, C. T., & Strohacker, K.** (2024). Exploring regular exercisers' experiences with readiness and recovery scores produced by wearable devices. A descriptive qualitative study. *Applied Psychophysiology and Biofeedback*, 49(3), 395–405. doi:10.1007/s10484-024-09645-2
2. **Chevance, G., Golaszewski, N. M., Tipton, E., Hekler, E. B., Buman, M., Welk, G. J., Patrick, K., & Godino, J. G.** (2022). Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures. Systematic review and meta-analysis. *JMIR mHealth and uHealth*, 10(4), e35626. doi:10.2196/35626
3. **Fuller, D., Colwell, E., Low, J., Orychock, K., Tobin, M. A., Simango, B., et al.** (2020). Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate. Systematic review. *JMIR mHealth and uHealth*, 8(9), e18694. doi:10.2196/18694
4. **Chinoy, E. D., Cuellar, J. A., Huwa, K. E., Jameson, J. T., Watson, C. H., Bessman, S. C., et al.** (2021). Performance of seven consumer sleep-tracking devices compared with polysomnography. *Sleep*, 44(5), zsaa291. doi:10.1093/sleep/zsaa291
5. **Molina-Garcia, P., Notbohm, H. L., Schumann, M., Argent, R., Hetherington-Rauth, M., Stang, J., et al.** (2022). Validity of estimating the maximal oxygen consumption by consumer wearables. A systematic review with meta-analysis and expert statement of the INTERLIVE network. *Sports Medicine*, 52(7), 1577–1597. doi:10.1007/s40279-021-01639-y
6. **Manresa-Rocamora, A., Sarabia, J. M., Javaloyes, A., Flatt, A. A., & Moya-Ramón, M.** (2021). Heart rate variability-guided training for enhancing cardiac-vagal modulation, aerobic fitness, and endurance performance. A methodological systematic review with meta-analysis. *International Journal of Environmental Research and Public Health*, 18(19), 10299. doi:10.3390/ijerph181910299
7. **Vesterinen, V., Nummela, A., Heikura, I., Laine, T., Hynynen, E., Botella, J., & Häkkinen, K.** (2016). Individual endurance training prescription with heart rate variability. *Medicine & Science in Sports & Exercise*, 48(7), 1347–1354. doi:10.1249/MSS.0000000000000910
8. **Javaloyes, A., Sarabia, J. M., Lamberts, R. P., & Moya-Ramon, M.** (2019). Training prescription guided by heart-rate variability in cycling. *International Journal of Sports Physiology and Performance*, 14(1), 23–32. doi:10.1123/ijspp.2018-0122
9. **Javaloyes, A., Sarabia, J. M., Lamberts, R. P., Plews, D., & Moya-Ramon, M.** (2020). Training prescription guided by heart rate variability versus block periodization in well-trained cyclists. *Journal of Strength and Conditioning Research*, 34(6), 1511–1518. doi:10.1519/JSC.0000000000003337
10. **Nuuttila, O.-P., Nummela, A., Korhonen, E., Häkkinen, K., & Kyröläinen, H.** (2022). Individualized endurance training based on recovery and training status in recreational runners. *Medicine & Science in Sports & Exercise*, 54(10), 1690–1701. doi:10.1249/MSS.0000000000002968
11. **Haddad, M., Stylianides, G., Djaoui, L., Dellal, A., & Chamari, K.** (2017). Session-RPE method for training load monitoring. Validity, ecological usefulness, and influencing factors. *Frontiers in Neuroscience*, 11, 612. doi:10.3389/fnins.2017.00612
12. **Wallace, L. K., Slattery, K. M., & Coutts, A. J.** (2009). The ecological validity and application of the session-RPE method for quantifying training loads in swimming. *Journal of Strength and Conditioning Research*, 23(1), 33–38. doi:10.1519/JSC.0b013e3181874512

Derived numbers, technique and equipment

1. **Cerezuela-Espejo, V., Hernández-Belmonte, A., Courel-Ibáñez, J., Conesa-Ros, E., Mora-Rodríguez, R., & Pallarés, J. G.** (2021). Are we ready to measure running power? Repeatability and concurrent validity of five commercial technologies. *European Journal of Sport Science*, 21(3), 341–350. doi:10.1080/17461391.2020.1748117
2. **Austin, C., Hokanson, J., McGinnis, P., & Patrick, S.** (2018). The relationship between running power and running economy in well-trained distance runners. *Sports*, 6(4), 142. Equipment supplied by the manufacturer, declared. doi:10.3390/sports6040142
3. **Borszcz, F. K., Tramontin, A. F., & Costa, V. P.** (2019). Is the functional threshold power interchangeable with the maximal lactate steady state in trained cyclists? *International Journal of Sports Physiology and Performance*, 14(8), 1029–1035. doi:10.1123/ijspp.2018-0572
4. **Wong, S., Burnley, M., Mauger, A., Fenghua, S., & Hopker, J.** (2022). Functional threshold power is not a valid marker of the maximal metabolic steady state. *Journal of Sports Sciences*, 40(23), 2578–2584. doi:10.1080/02640414.2023.2176045
5. **Mooney, R., Corley, G., Godfrey, A., Quinlan, L., & ÓLaighin, G.** (2016). Inertial sensor technology for elite swimming performance analysis. A systematic review. *Sensors*, 16(1), 18. doi:10.3390/s16010018
6. **de Magalhaes, F. A., Vannozzi, G., Gatta, G., & Fantozzi, S.** (2015). Wearable inertial sensors in swimming motion analysis. A systematic review. *Journal of Sports Sciences*, 33(7), 732–745. doi:10.1080/02640414.2014.962574
7. **Gay, A., Ruiz-Navarro, J. J., Cuenca-Fernández, F., López-Belmonte, Ó., Abraldes, J. A., Fernandes, R. J., & Arellano, R.** (2022). The impact of wetsuit use on swimming performance, physiology and biomechanics. A systematic review. *Physiologia*, 2(4), 198–230. doi:10.3390/physiologia2040016

Strength, durability and injury

1. **Denadai, B. S., de Aguiar, R. A., de Lima, L. C., Greco, C. C., & Caputo, F.** (2017). Explosive training and heavy weight training are effective for improving running economy in endurance athletes. A systematic review and meta-analysis. *Sports Medicine*, 47(3), 545–554. doi:10.1007/s40279-016-0604-z
2. **Berryman, N., Mujika, I., Arvisais, D., Roubeix, M., Binet, C., & Bosquet, L.** (2018). Strength training for middle- and long-distance performance. A meta-analysis. *International Journal of Sports Physiology and Performance*, 13(1), 57–63. Erratum 13(3), 398. doi:10.1123/ijspp.2017-0032
3. **Llanos-Lagos, C., Ramirez-Campillo, R., Moran, J., & Sáez de Villarreal, E.** (2024). The effect of strength training methods on middle-distance and long-distance runners' athletic performance. A systematic review with meta-analysis. *Sports Medicine*, 54(7), 1801–1833. doi:10.1007/s40279-024-02018-z
4. **Andersen, C. A., Clarsen, B., Johansen, T. V., & Engebretsen, L.** (2013). High prevalence of overuse injury among iron-distance triathletes. *British Journal of Sports Medicine*, 47(13), 857–861. doi:10.1136/bjsports-2013-092397
5. **Crunkhorn, M. L., Toohey, L. A., Charlton, P., Drew, M., Watson, K., & Etxebarria, N.** (2024). Injury incidence and prevalence in elite short-course triathletes. A 4-year prospective study. *British Journal of Sports Medicine*, 58(9), 470–476. doi:10.1136/bjsports-2023-107327
6. **Rhind, J. H., Dass, D., Barnett, A., & Carmont, M.** (2022). A systematic review of long-distance triathlon musculoskeletal injuries. *Journal of Human Kinetics*, 81, 123–134. doi:10.2478/hukin-2022-0011
7. **Vleck, V., Millet, G. P., & Alves, F. B.** (2014). The impact of triathlon training and racing on athletes' general health. *Sports Medicine*, 44(12), 1659–1692. doi:10.1007/s40279-014-0244-0

Environment, nutrition and clinical boundaries

1. **Racinais, S., Hosokawa, Y., Akama, T., Bermon, S., Bigard, X., Casa, D. J., et al.** (2023). IOC consensus statement on recommendations and regulations for sport events in the heat. *British Journal of Sports Medicine*, 57(1), 8–25. doi:10.1136/bjsports-2022-105942
2. **Jeukendrup, A. E.** (2014). A step towards personalized sports nutrition. Carbohydrate intake during exercise. *Sports Medicine*, 44(Suppl 1), S25–S33. doi:10.1007/s40279-014-0148-z
3. **Burke, L. M., Hawley, J. A., Wong, S. H. S., & Jeukendrup, A. E.** (2011). Carbohydrates for training and competition. *Journal of Sports Sciences*, 29(Suppl 1), S17–S27. doi:10.1080/02640414.2011.585473
4. **Podlogar, T., & Wallis, G. A.** (2022). New horizons in carbohydrate research and application for endurance athletes. *Sports Medicine*, 52(Suppl 1), 5–23. doi:10.1007/s40279-022-01757-1
5. **Tyler, C. J., Reeve, T., Hodges, G. J., & Cheung, S. S.** (2016). The effects of heat adaptation on physiology, perception and exercise performance in the heat. A meta-analysis. *Sports Medicine*, 46(11), 1699–1724. Erratum 46(11), 1771. Cited for direction and duration; the abstract reports magnitude as verbal descriptors only. doi:10.1007/s40279-016-0538-5
6. **Mountjoy, M., Ackerman, K. E., Bailey, D. M., Burke, L. M., Constantini, N., Hackney, A. C., et al.** (2023). 2023 International Olympic Committee's consensus statement on relative energy deficiency in sport. *British Journal of Sports Medicine*, 57(17), 1073–1097. doi:10.1136/bjsports-2023-106994
7. **Stellingwerff, T., Mountjoy, M., McCluskey, W. T. P., Ackerman, K. E., Verhagen, E., & Heikura, I. A.** (2023). Review of the scientific rationale, development and validation of the International Olympic Committee Relative Energy Deficiency in Sport Clinical Assessment Tool, version 2, by a subgroup of the IOC consensus on REDs. *British Journal of Sports Medicine*, 57(17), 1109–1121. doi:10.1136/bjsports-2023-106914
8. **Burke, L. M., Lundy, B., Fahrenholtz, I. L., & Melin, A. K.** (2018). Pitfalls of conducting and interpreting estimates of energy availability in free-living athletes. *International Journal of Sport Nutrition and Exercise Metabolism*, 28(4), 350–363. doi:10.1123/ijsnem.2018-0142

Pacing, race execution and artificial intelligence

1. **Abbiss, C. R., & Laursen, P. B.** (2008). Describing and understanding pacing strategies during athletic competition. *Sports Medicine*, 38(3), 239–252. doi:10.2165/00007256-200838030-00004
2. **Wu, S. S. X., Peiffer, J. J., Brisswalter, J., Nosaka, K., & Abbiss, C. R.** (2014). Factors influencing pacing in triathlon. *Open Access Journal of Sports Medicine*, 5, 223–234. doi:10.2147/OAJSM.S44392
3. **Etxebarria, N., Anson, J. M., Pyne, D. B., & Ferguson, R. A.** (2013). Cycling attributes that enhance running performance after the cycle section in triathlon. *International Journal of Sports Physiology and Performance*, 8(5), 502–509. doi:10.1123/ijspp.8.5.502
4. **Vleck, V. E., Bürgi, A., & Bentley, D. J.** (2006). The consequences of swim, cycle, and run performance on overall result in elite Olympic distance triathlon. *International Journal of Sports Medicine*, 27(1), 43–48. doi:10.1055/s-2005-837502
5. **Dergaa, I., Ben Saad, H., El Omri, A., Glenn, J. M., Clark, C. C. T., Washif, J. A., et al.** (2024). Using artificial intelligence for exercise prescription in personalised health promotion. A critical evaluation of OpenAI's GPT-4 model. *Biology of Sport*, 41(2), 221–241. doi:10.5114/biolsport.2024.133661
6. **Washif, J. A., Pagaduan, J., James, C., Dergaa, I., & Beaven, C. M.** (2024). Artificial intelligence in sport. Exploring the potential of using ChatGPT in resistance training prescription. *Biology of Sport*, 41(2), 209–220. doi:10.5114/biolsport.2024.132987
7. **Havers, T., Jelonnek, C., Masur, L., Isenmann, E., Sperlich, B., Geisler, S., & Düking, P.** (2025). A professional assessment of training plans for muscle hypertrophy and maximal strength developed by generative artificial intelligence. *Biology of Sport*, 42(4), 353–366. Published online 26 August 2025. doi:10.5114/biolsport.2026.152350

Governance, participation and value

1. **World Triathlon** (2025). *World Triathlon Competition Rules.* Approved by the Executive Board in January 2025; document dated 18 January 2025. Rules 1.7, 2.1, 2.2, 5 and 6.4 referenced in the text.
2. **Raysmith, B. P., & Drew, M. K.** (2016). Performance success or failure is influenced by weeks lost to injury and illness in elite Australian track and field athletes. A 5-year prospective study. *Journal of Science and Medicine in Sport*, 19(10), 778–783. doi:10.1016/j.jsams.2015.12.515
3. **Sport England** (2025). *Active Lives Adult Survey November 2023-24 Report*, and the accompanying activity data tables. Loughborough, Sport England. Released 24 April 2025. Triathlon figures read from the published tables; England only.

WHERE THIS GOES NEXT

A connected athlete pilot

The argument in this paper is testable, and the only honest way to test it is across a squad of real athletes over a full season, with an independent evaluation partner and a measure agreed in advance. The measure this paper proposes is completed planned weeks, because it is countable, auditable and the one most closely linked to hitting a performance goal in the published prospective work.

Athleet.AI works with coaches, squads, federations and technology partners to run connected athlete pilots. The reconciliation layer, the error reporting and the decision record come first. Prediction is not part of the pilot, and Chapter 13 explains why.

What a pilot involves

- One season, one squad, three disciplines and the gym reconciled into a single record.
- Written reconciliation rules, published to the athletes rather than held in the product.
- Error reported alongside every derived number the athlete sees.
- Every recommendation recorded with the decision that followed it, and the override rate measured.
- A published result whichever way it goes, including the honest count of what the system got wrong.