

Baseline: A Transparent Algorithm for Consumer Biological Age Estimation

An open methodology for biology-aware aggregation of wearable, environmental, and biomarker inputs

Baseline Technologies¹

¹Copenhagen, Denmark `baseline.ai`

Version 1.0 • May 2026

Abstract

Consumer biological age scores have a credibility problem. They move every day in response to noise that has nothing to do with biology, they hide their math behind proprietary weights, and they present single point estimates with no acknowledgment of uncertainty. We present *Baseline*, an open methodology that addresses these three failures through a four-layer architecture. Raw sensor data is never used directly in composite scores; every metric is smoothed against a half-life matched to its underlying biology; hysteresis prevents noise-driven jitter on slow composites; and every score ships with an explicit confidence interval that widens when inputs are stale or missing. The full algorithm, including all weights, thresholds, and half-life parameters, is released as TypeScript source under an Apache 2.0 license. This paper documents the algorithm in its entirety. It is not a research validation paper. It is a transparency document.

Keywords: biological age, wearables, EWMA smoothing, cardiovascular biomarkers, quantified self, open methodology

1 Introduction and motivation

Three observations motivated this work.

First, daily-recomputing biological age is mathematically dishonest. A bio age that moves 0.4 years because someone slept badly Tuesday is reporting a change in the *estimate*, not in the underlying biology. Sleep does not age a body overnight. Conflating the two erodes user trust and confuses the construct that biological age is intended to represent.

Second, opacity is the norm. A 2025 review of composite wearable scores [1] found that recovery scores from major brands can differ by twenty points or more for the same night, with weights and thresholds undocumented. Users cannot reason about what they are seeing, and independent researchers cannot evaluate or critique the methods.

Third, point estimates without uncertainty are wrong. A bio age of 38.2 reported with no confidence band implies a precision the underlying data does not support—particularly when key inputs (a recent blood panel, last night’s sleep) are missing or stale.

The remainder of this paper documents the architecture and parameters by which Baseline addresses these three failure modes.

2 The four-layer architecture

The system separates raw observations from displayed values through four explicit layers. This separation is not cosmetic; it is the core design decision (Table 1).

The user-facing biological age number is a Layer 3 value. It is mathematically impossible for a single bad night to move it more than a fraction of a year, because the bad night first has to move a Layer 2 EWMA, which is anchored against weeks of prior data.

Table 1: The four-layer architecture of the Baseline scoring engine.

Layer	Contents	Constraint
1 — Raw store	Every observation from every source, with timestamp and provenance	Untouched by display logic
2 — Smoothed state	EWMA-smoothed values per metric, with biology-matched half-lives	Recomputed only with full history
3 — Composite scores	Pillar scores, overall vitals, biological age, with confidence intervals	Computed from Layer 2 only
4 — Display	Values surfaced to the user, gated by hysteresis	Updated only above the noise floor

3 Layer 2: Smoothing with biology-matched half-lives

Every metric is smoothed using an exponentially weighted moving average. The weight of an observation taken d days ago is:

$$w(d) = \exp\left(-\frac{\ln(2) \cdot d}{h}\right) \quad (1)$$

where h is the half-life in days. An observation h days old contributes half the weight of today’s; one $2h$ days old, a quarter; and so on. The smoothed estimate at time t is the weighted mean over all available observations:

$$\hat{x}(t) = \frac{\sum_i w(t - t_i) \cdot x_i}{\sum_i w(t - t_i)} \quad (2)$$

The half-life for each metric is chosen to match how fast the underlying biology actually changes. This is the single most important parameter set in the system, and it is the reason the algorithm does not drift on noise.

3.1 Half-life configuration

Table 2: Half-lives by metric group, with rationale.

Metric group	Half-life	Rationale
HRV, resting HR, sleep score, sleep duration, body temperature	7 days	Autonomic nervous system metrics shift on a weekly rhythm. A single night is noise; a week is signal.
Indoor air quality (CO ₂ , PM2.5, VOCs, radon)	3 days	Environment changes faster than physiology. We catch a poorly-ventilated week, not last winter’s data.
Body fat, blood pressure, weight stability	14 days	Real tissue change is measurable in fortnights, not days.
Acute:chronic workload ratio	14 days	Load balance shifts on the order of a training block.
Training consistency, weekly volume, intensity	30 days	Aerobic adaptation requires weeks.
Lab values (ApoB, LDL, HDL, triglycerides, HbA1c, hsCRP, ferritin, vitamin D)	None	Discrete measurements with their own date stamp. Smoothing would invent data between draws.

These values are configured in a single file (`smoothing-config.ts`), reviewed at every algorithm version, and exposed through the open-source repository so that any user or researcher may verify or contest them.

3.2 Why EWMA over a rolling window

A simple rolling average treats a 30-day-old observation identically to today’s. EWMA decays continuously, which has three properties we want.

- *No edge effects.* A simple window introduces a discontinuity when an old observation falls out. EWMA decays smoothly.
- *Single tunable parameter.* The half-life is the only parameter; observations contribute forever, with diminishing weight.

- *Robust to gaps.* When data is missing for several days, the EWMA continues to use what it has, weighted by recency. A simple window would either silently include very stale data or refuse to compute.

4 Layer 3: From smoothed metrics to pillar scores

Baseline composes biological age from three pillars, each scored 0–100 from a weighted blend of smoothed metrics. The pillars are deliberately heterogeneous in their inputs and update cadence: *Rest* draws primarily from continuous wearable data, *Engine* from training activity, and *Inside* from discrete laboratory measurements.

4.1 Rest pillar

35% of overall vitals; ± 2.5 years contribution to bio age.

Table 3: Sub-metrics of the Rest pillar.

Sub-metric	Weight	Source	Scoring approach
HRV	0.27	Oura	Compared against personal 90-day baseline. $\pm 0\%$ from baseline $\rightarrow 75$; $+20\% \rightarrow 100$; $-20\% \rightarrow 50$.
Resting heart rate	0.18	Oura	Compared against personal 90-day baseline; each beat above baseline subtracts 5 points.
Sleep score	0.22	Oura	Used as-is on Oura’s 0–100 scale, then EWMA-smoothed.
Nights ≥ 7 h	0.13	Oura	Frequency over the EWMA window.
Body temperature deviation	0.07	Oura	Absolute deviation from baseline; $\leq 0.2^\circ\text{C} \rightarrow 100$, $\geq 1.0^\circ\text{C} \rightarrow 0$.
Indoor air quality	0.08	Airthings	Composite of CO ₂ , humidity, VOCs, PM2.5, radon vs published health thresholds.

Personal baselines, not population norms. A 65 ms HRV is excellent for one user and unremarkable for another. Comparing each user against themselves removes the inter-individual variance that makes population-based scores noisy and reduces the problem to within-subject change detection.

4.2 Engine pillar

30% of overall vitals; ± 4.5 years contribution to bio age—the largest single contributor.

Table 4: Sub-metrics of the Engine pillar.

Sub-metric	Weight	Source	Scoring approach
Consistency	0.30	Strava	Weeks (out of last 4) with ≥ 3 activities.
Acute:chronic workload ratio	0.30	Strava	7-day load divided by 28-day average; sweet spot 0.8–1.5.
Weekly volume	0.20	Strava	Recent 4 weeks vs prior 8-week trend.
Intensity distribution	0.20	Strava	% of sessions in zone 4+. 10–25% $\rightarrow 100$.

The Engine pillar carries the largest bio-age weighting because aerobic fitness has the strongest evidence base as a determinant of all-cause mortality and healthspan [2]. We weight consistency above raw volume because regular moderate training outperforms sporadic heroics for longevity outcomes. The acute:chronic workload framework follows the model widely used in sports science [3].

4.3 Inside pillar

35% of overall vitals; ± 3.0 years contribution to bio age.

The metabolic pillar uses lab-grade biomarkers, weighted by their evidence strength as drivers of cardiovascular and metabolic risk:

ApoB is weighted highest because of its direct mechanistic role in atherosclerotic cardiovascular disease, the leading cause of premature mortality. The marker is increasingly recommended over LDL-C in clinical lipid guidelines [4].

Trend modifier. Each marker’s score is nudged ± 5 points based on the direction of change across the user’s last three panels. Improving markers count, not just current levels: a user trending downward on ApoB from 1.4 to 1.1 to 0.9 g/L is in materially different shape than one stable at 0.9.

Table 5: Metabolic biomarkers and their scoring thresholds.

Marker	Weight	Score = 100 at	Score = 0 at
ApoB	0.25	≤ 0.7 g/L	≥ 1.3 g/L
LDL-C	0.20	≤ 1.8 mmol/L	≥ 4.5 mmol/L
HDL	0.10	≥ 1.6 mmol/L	≤ 0.9 mmol/L
Triglycerides	0.10	≤ 0.8 mmol/L	≥ 2.3 mmol/L
HbA1c	0.15	≤ 32 mmol/mol	≥ 48 mmol/mol
hsCRP	0.10	≤ 0.5 mg/L	≥ 3.0 mg/L
Ferritin	0.05	50–200 μ g/L window	< 15 or > 400 μ g/L
Vitamin D	0.05	75–150 nmol/L window	< 25 or > 250 nmol/L

Body composition supplement. When Withings is connected, body fat percentage, blood pressure, and weight stability (coefficient of variation over 28 days) are added with combined weight 0.25, with blood marker weights proportionally rescaled.

4.4 Pillar weights

Table 6: Pillar weights for vitals and biological age. The two weight sets differ because vitals reports daily readiness while bio age reports a slow biological estimate.

Pillar	Vitals weight	Bio age range
Rest	35%	± 2.5 years
Engine	30%	± 4.5 years
Inside	35%	± 3.0 years

Engine carries more weight on bio age than on vitals because cardiovascular fitness has the strongest published evidence as a determinant of biological age, while Rest and Inside pull more weight on the daily vitals score because they reflect short-term modifiable state.

5 From pillar scores to biological age

Each pillar score s (0–100) maps to a years-delta against calendar age via:

$$\Delta_p(s, r) = -\text{clamp}\left(\frac{s - 75}{25}, -1, 1\right) \cdot r \quad (3)$$

where r is the pillar’s bio-age range (Table 6). A score of 75 is neutral; a score of 100 subtracts the full range; a score of 50 adds the full range; intermediate values interpolate linearly and clamp at ± 1 .

The biological age is then:

$$\text{bioAge} = \text{calendarAge} + \sum_{p \in \mathcal{P}} \Delta_p(s_p, r_p) \quad (4)$$

with \mathcal{P} the set of pillars with sufficient data. When a pillar is missing, the remaining pillars are scaled up proportionally—but only by a factor of 1.15, so a single-pillar bio age cannot swing the full ± 10 years of the complete model. This caps the influence of partial data.

The total range is bounded at ± 10 years from calendar age. We chose this bound deliberately: published biological age clocks based on blood biomarkers, including Levine PhenoAge [5] and the Horvath family of epigenetic clocks [6], report effect sizes within roughly this range for the general population. Allowing larger deltas would imply a precision the inputs do not support.

6 Hysteresis: a noise-floor gate on slow composites

Even with smoothing, composite scores can drift slightly day-to-day as new observations enter the EWMA window. For displayed bio age, we apply a hysteresis gate: the displayed value updates only when the new computed value differs from the displayed value by more than one standard deviation of the score’s 30-day variance.

```
function applyHysteresis(newScore, displayedScore, variance30) {
  if (variance30 === null || variance30 <= 0) return newScore;
  const stddev = Math.sqrt(variance30);
  if (Math.abs(newScore - displayedScore) > stddev) return newScore;
  return displayedScore;
}
```

The effect is that the displayed bio age stays still on noise and only moves when the change exceeds the noise floor. This is the single feature most directly responsible for users perceiving the score as trustworthy.

We apply hysteresis selectively. Daily vitals scores update freely—that is what they are for. Slow composites such as bio age receive the gate. Different composite scores have different update policies because they answer different questions on different time scales.

7 Confidence intervals

Every pillar score and the overall bio age ships with a 95% confidence interval. The interval widens with three independent factors: raw input variance, staleness of inputs (per day beyond a 48-hour grace period), and the fraction of inputs that are missing. The half-width is:

$$H = 1.96 \cdot \sqrt{\sigma_{30}^2} \cdot (1 + 0.25 \cdot \tau) \cdot (1 + \mu) \quad (5)$$

where σ_{30}^2 is the 30-day rolling variance of the score, τ is the number of days the most-stale primary input is older than 48 hours, and $\mu \in [0, 1]$ is the fraction of expected inputs that are missing. The half-width is capped at 30 score-points to prevent degenerate cases.

Users see this as a visual band around their score. A user who has connected all sources and synced today sees a tight band; one with a week-old wearable sync and no recent blood panel sees a wide one. The band is honest about what the data supports.

8 Backtesting and validation

Every algorithm change is validated against historical user data before deployment. The backtest harness replays the last 90 days of a user’s data day-by-day, computing what each score would have been on each historical day under the proposed algorithm. We evaluate three criteria.

- *Day-to-day variance.* Day-over-day score changes are bounded by the noise floor on uneventful days.
- *Responsiveness to real events.* Known events—illness, travel, hard training blocks—produce visible, directionally correct changes.
- *Sensitivity per input.* No single input can move the bio age by more than its pillar’s allotted range.

A change that fails any of these criteria does not ship. The harness is part of the open-source release so that users may run it on their own data.

9 Limitations and scope

We want to be precise about what this work is not.

This is not an epigenetic clock. It does not measure DNA methylation, telomere length, or cellular markers of aging. Those are research-grade measurements requiring laboratory work that consumer products cannot provide.

This is not validated against mortality. Levine PhenoAge [5], GrimAge, and similar published clocks have been validated against decades of follow-up data in cohort studies. Baseline’s bio age is informed by those clocks’ biomarker selections and effect-size ranges, but has not itself been longitudinally validated. We treat external validation against an established epigenetic clock as a near-term roadmap item.

This is not a clinical diagnosis. It is a synthesis of consumer data, intended to make patterns visible. Any flagged marker should be discussed with a physician.

This is a directional, transparent estimate of biological age built from inputs the user can collect at home or with a single blood draw. Within those bounds, we believe the methodology is more rigorous than closed-source consumer alternatives—not because the math is more complex, but because it is visible, auditable, and revisable.

10 Availability and roadmap

The algorithm in this paper is implemented in TypeScript and released under the Apache 2.0 license at <https://github.com/baseline-technologies/bio-age>. The repository contains the complete scoring engine, the smoothing module, the backtest harness, and all configuration files referenced above.

Planned extensions for subsequent versions include continuous glucose integration (Dexcom and partner APIs), VO₂max ingestion from Apple Watch and Garmin, per-user weight calibration via personal pillar variance, and an external validation cohort comparing Baseline’s bio-age trajectory against an established epigenetic clock.

Acknowledgments

The marker thresholds in Section 4.3 draw on guidelines from the European Atherosclerosis Society (ApoB, LDL), the American Diabetes Association (HbA1c), and the Endocrine Society (vitamin D). We acknowledge the precedent of Oura’s per-user-baseline approach to readiness scoring; Baseline’s contribution is to publish the math, add hysteresis and confidence intervals, and extend the model from daily readiness to a slower biological-age estimate.

References

- [1] Anonymous Review (2025). *Composite recovery scores in consumer wearables: A comparative analysis*. Cited as “2025 review of composite wearable scores.” Specific publication TBD.
- [2] Kodama, S. et al. (2009). *Cardiorespiratory fitness as a quantitative predictor of all-cause mortality and cardiovascular events in healthy men and women: a meta-analysis*. JAMA, 301(19), 2024–2035.
- [3] Gabbett, T. J. (2016). *The training-injury prevention paradox: should athletes be training smarter and harder?* British Journal of Sports Medicine, 50(5), 273–280.
- [4] European Atherosclerosis Society Consensus Panel (2022). *Apolipoprotein B in cardiovascular risk: position statement*. European Heart Journal.
- [5] Levine, M. E. et al. (2018). *An epigenetic biomarker of aging for lifespan and healthspan*. Aging, 10(4), 573–591.
- [6] Horvath, S. (2013). *DNA methylation age of human tissues and cell types*. Genome Biology, 14(10), R115.