



# Enhancing the Athlytx Score Algorithm: Research Insights and Implementation Plan

## Introduction

The Athlytx Score is a holistic readiness metric designed for endurance athletes and active individuals. To elevate its accuracy and credibility, we conducted deep research into sports science and commercial best practices. Key focus areas include recovery metrics (HRV, resting HR, subjective readiness, "Body Battery"-type energy models), training load quantification (TRIMP, ACWR, CTL/ATL fatigue-fitness models, and methods used by WHOOP, Strava, Garmin), sleep analysis (sleep efficiency, deep/REM sleep, latency, Oura's scoring), multi-device data integration, personalization via machine learning, and handling missing data with confidence estimates. Below we summarize evidence-based insights in each area and then present a comprehensive implementation plan for an improved Athlytx Score algorithm.

## 1. Recovery Metrics and Readiness Indicators

**Heart Rate Variability (HRV):** HRV, especially vagal metrics like rMSSD, is a well-established marker of recovery and autonomic balance. Reduced resting HRV is frequently associated with fatigue, overtraining, or illness <sup>1</sup> <sup>2</sup>. However, research emphasizes context – high training load doesn't always mean low HRV, and vice versa <sup>3</sup>. For example, athletes can exhibit stable or even increased HRV during heavy training if they are adapting well, whereas life stressors can suppress HRV despite low training load <sup>3</sup>. Recent studies suggest combining HRV with other signals improves its utility. A 2025 swimmer study found a ratio of HRV to heart period (rMSSD/AVNN) better distinguished overload vs. taper phases than HRV alone <sup>4</sup> <sup>5</sup>, indicating that normalizing HRV by resting heart rate can enhance sensitivity to fatigue states. **Resting Heart Rate (RHR):** Elevated RHR is a traditional red flag for accumulated fatigue or illness, reflecting higher sympathetic drive <sup>2</sup>. It's long been used by coaches as a simple morning readiness check – a sudden jump in RHR can signal insufficient recovery. Modern research highlights the *interplay* of HRV and RHR: because HRV is mathematically influenced by average heart rate, tracking both together yields better insight <sup>6</sup>. In fact, combining HRV and RHR has shown promise for detecting overtraining versus recovery states in high-load athletes <sup>7</sup>. This combined approach can catch subtle fitness/fatigue changes that one metric alone might miss.

**Subjective Readiness:** Self-reported well-being (mood, energy, soreness, sleep quality) is a proven predictor of performance and overtraining risk. Psychological signs like worsened mood, fatigue, or poor sleep often appear *before* objective metrics change <sup>8</sup>. Simple wellness questionnaires can sensitively track these early warnings <sup>8</sup>. A pivotal review concluded that subjective measures of an athlete's status can be more consistent and sensitive than any single objective measure <sup>9</sup> <sup>10</sup>. However, subjective and objective metrics should be seen as *complementary* rather than either-or. As HRV researcher Marco Altini notes, "training load, HRV, and subjective metrics all provide important information that needs to be integrated daily" – there is no single winner; a smart system uses all of them together <sup>11</sup>. Therefore, Athlytx should incorporate an athlete's own perception of readiness (e.g. morning questionnaires or RPE-based recovery scores) alongside physiological data.

**Body Battery Models:** Garmin's "Body Battery" is a notable commercial approach to quantifying recovery. It provides a 0-100 "energy reserve" score by analyzing HRV, stress, and sleep data in real time

<sup>12</sup>. In essence, high HRV and restful sleep charge the battery, while stress (high sympathetic activity) and exertion drain it. This kind of running energy gauge is intuitive for users and grounded in autonomic science. It aligns with research that parasympathetic activity (high HRV) promotes recovery and sympathetic activity correlates with stress <sup>13</sup> <sup>14</sup>. Athlytx can emulate this approach: use nocturnal HRV baselines and daytime HRV/stress fluctuations to model recovery throughout the day. In summary, the evidence urges a *multi-pronged recovery score* that blends HRV (especially relative to personal baseline), resting HR, and subjective readiness. This combination yields a more robust recovery index than any one metric alone <sup>15</sup>. For example, if HRV is suppressed and RHR is elevated (physiological strain) *and* the user reports feeling fatigued, the recovery score should drop significantly. If metrics conflict (e.g. HRV low but user feels great), the system may flag an “uncertain” recovery status, prompting caution. By fusing these inputs, Athlytx’s recovery component will mirror state-of-the-art readiness scores like WHOOP’s Recovery (which integrates HRV and RHR overnight) and Oura’s Readiness (which incorporates HRV, RHR, sleep, and fatigue).

## 2. Training Load Quantification

Properly quantifying training load is vital for balancing fitness gains and fatigue. We examined classic models and current practices used by leading platforms:

- **Banister’s TRIMP:** The concept of *Training Impulse (TRIMP)*, introduced by Banister, remains foundational. TRIMP combines exercise duration and intensity (using heart rate zones weighted exponentially) into a single score <sup>16</sup>. In Banister’s model, each heartbeat contributes to load based on how intense it is relative to one’s maximum, with higher-intensity minutes counted more heavily <sup>17</sup>. TRIMP provides an “internal load” metric (cardiovascular strain) for each session. Athlytx can use a modern TRIMP variant (as Firstbeat and Strava do) to quantify each workout’s load. In fact, Firstbeat (the physiology engine behind Garmin) still uses a modified Banister TRIMP calculation—updating it more frequently via beat-to-beat HR data for accuracy during intervals <sup>18</sup>. The *cumulative* TRIMP over days is crucial for understanding fatigue.
- **Chronic vs. Acute Load (CTL/ATL):** Sports science often characterizes training stress in terms of *chronic* (long-term) and *acute* (short-term) loads. Chronic Training Load (CTL) is essentially a running average of daily load (e.g. over 4–6 weeks), representing the athlete’s fitness base, whereas Acute Training Load (ATL) is the recent load (e.g. past week) representing fatigue <sup>19</sup>. The balance of the two is sometimes called *Training Stress Balance (TSB)* or simply the fitness-fatigue model: when acute load far exceeds chronic load, the athlete is likely in a fatigued, high-risk state; when acute load is much lower than chronic, the athlete may be tapering or detraining. Firstbeat’s system quantifies exactly this: the sum of 7-day TRIMP as acute load and 28-day average TRIMP as chronic load, then analyzes their ratio <sup>19</sup>. This is analogous to the popular **Acute:Chronic Workload Ratio (ACWR)** used to monitor injury risk. ACWR is typically computed as last 7 days’ load divided by last 28 days’ load <sup>20</sup>. Values above 1.5 (meaning acute load 50%+ higher than chronic) have been associated with elevated injury risk, whereas a ratio around 1.0–1.3 is often considered a “sweet spot” for progressive training <sup>21</sup>. This model stems directly from Banister’s fitness-fatigue idea, where performance is a function of positive fitness minus negative fatigue <sup>22</sup>. Athlytx should incorporate CTL/ATL tracking under the hood, and can surface an ACWR-style indicator or “training balance” status. For instance, if a user’s 7-day load is dangerously higher than their 4-week baseline, the app can flag potential overreaching. Conversely, a very low ACWR could indicate a recovery period or insufficient stimulus.
- **Training Load Scores in Practice (Strava, TrainingPeaks, Garmin):** It’s insightful to see how commercial tools present these concepts. **Strava’s Relative Effort (RE)** is a user-friendly

implementation of TRIMP. Strava collaborated with HRV expert Marco Altini to refine TRIMP by weighting intensity even more heavily and normalizing across sports <sup>23</sup> <sup>24</sup>. A short, hard interval run might score as high as a much longer easy run – this reflects the principle that intensity drives strain. Strava also scales the score using its large dataset so that what counts as, say, “100 points” of effort is comparable whether you’re cycling, running, or swimming <sup>24</sup>. In Athlytx, we can adopt a similar zone-weighted approach: e.g. using heart rate or power zones with coefficients (zone 5 minutes count exponentially more than zone 2 minutes) <sup>24</sup>. This ensures the score rewards truly strenuous efforts and isn’t fooled by long low-intensity volume alone. Notably, Strava’s system feeds into a *Fitness & Freshness* chart, which is effectively CTL/ATL as well (Strava’s “Fitness” is an exponentially-weighted load akin to CTL, and “Form” is analogous to TSB). Strava even uses a 3-week rolling baseline to contextualize your current week’s load, coloring it green if within range, red if sharply above recent average, blue if below (taper) <sup>25</sup> <sup>26</sup>. We can present similar visuals in Athlytx to help users manage training stress week-to-week.

**Garmin (Firstbeat) Training Load:** Garmin devices report a 7-day Training Load number and categorize your training status (e.g. Productive, Overreaching, Recovery) based on how that 7-day load compares to your optimal range (which itself is based on longer-term load and fitness). Garmin’s load metric is derived from EPOC (Excess Post-exercise Oxygen Consumption) – essentially measuring how much you “disturbed homeostasis” in each workout <sup>27</sup>. Each session’s Exercise Load is computed from EPOC curves and then summed over days. Like Relative Effort, it primarily reflects internal cardiovascular load. Athlytx can ingest Garmin’s Training Load via their API (since Garmin is a data source) to cross-verify our own calculations. If multiple sources are present (e.g. a user has both Strava and Garmin data), highlighting discrepancies could be useful – but more on data blending below.

**TrainingPeaks TSS and Fitness/Fatigue:** Serious endurance athletes often use TrainingPeaks’ **TSS (Training Stress Score)**, which is analogous to TRIMP but based on *external* load (power or pace relative to threshold). TSS is calibrated so that 100 points ≈ one hour at lactate threshold. While Athlytx may not initially implement power-based TSS, it’s good to note that a TSS-based CTL/ATL model is conceptually similar to a TRIMP-based one (just different inputs). In the future, if we integrate with platforms like Strava or directly with power data, we could incorporate TSS as well.

- **WHOOP Strain:** WHOOP uses a proprietary strain metric on a 0–21 scale to represent total daily cardiovascular load. It’s conceptually similar to TRIMP/Relative Effort but scaled differently. A strain of 20+ corresponds to an exceptionally hard day (marathon or equivalent), whereas a light recovery day might be under 10. WHOOP intentionally aligned the 0–21 scale loosely with Borg’s RPE (Rating of Perceived Exertion) <sup>28</sup>. This makes it easier for users to intuitively understand (e.g. 15+ is a “high strain” day). WHOOP’s innovation was also to include *non-exercise stress*: the device monitors elevated heart rate throughout the day (from lifestyle stress, work, etc.) so that the “Day Strain” captures total load on the body, not just workouts <sup>29</sup>. Additionally, WHOOP 4.0/5.0 introduced a **muscular load** component: using accelerometer and strain sensors to estimate musculoskeletal strain (e.g. weightlifting effort) which pure heart rate might miss <sup>30</sup> <sup>31</sup>. The WHOOP strain algorithm combines cardiovascular and muscular load into the one strain score <sup>32</sup>. For Athlytx, the take-home is to quantify *total stress*. We can incorporate daily active calories or steps as a proxy for non-exercise strain, and include both endurance and strength sessions in the load score. If a user has WHOOP data connected, Athlytx could even import their strain to use directly or for validation. At minimum, our strain component should accumulate across the entire day, not just planned workouts, because as WHOOP highlights, “running errands, parenting, or even mental stress can all contribute to strain” <sup>32</sup>.

In summary, the improved Athlytx **Training Load component** will use a TRIMP-like metric (normalized for intensity) for each activity, sum into daily load, and track acute vs chronic load. It will present the user with insights like "Today's strain: 250 (High)" and "7-day load vs 28-day load: 1.4 ratio (slightly above optimal; ensure recovery)". Backed by research, this approach helps optimize performance while preventing overtraining and injury <sup>33</sup> <sup>34</sup>.

### 3. Sleep Analysis and Readiness

Quality sleep is fundamental to recovery. Our research looked at which sleep metrics most strongly influence readiness and how systems like Oura's quantify sleep quality:

**Sleep Quantity & Efficiency:** Total sleep hours matter – most adults need ~7–9 hours for full recovery <sup>35</sup>. But it's not just quantity; *efficiency* (the percentage of time in bed that one is actually asleep) is a key quality indicator. An efficient sleep (>85% asleep vs time in bed) signifies minimal disturbances <sup>36</sup>. Efficiency below 85% or taking longer than ~20 minutes to fall asleep will lower Oura's Sleep Score <sup>37</sup>. In Athlytx, sleep efficiency can similarly feed into the Sleep Component: frequent wake-ups or long latency means the user's sleep was less restorative than the duration alone suggests. If, say, a user spent 8 hours in bed but only slept 6.5 hours (81% efficiency), we'd flag that as suboptimal sleep quality.

**Sleep Stages – Deep and REM:** Deep sleep (slow-wave sleep) and REM sleep serve different recovery functions. Deep sleep is when the body repairs muscle and tissues, releases growth hormone, and fortifies the immune system <sup>38</sup> <sup>39</sup>. REM sleep is critical for neurocognitive recovery – memory consolidation, learning, and emotional regulation <sup>40</sup>. In healthy adults, REM typically comprises ~20-25% of the night (~1.5+ hours) and deep sleep ~15-20% (~1-1.5 hours) <sup>41</sup> <sup>42</sup>. Oura's guidelines consider ~90 minutes of REM and ~60-90 minutes of deep sleep as optimal targets for a full night <sup>41</sup> <sup>43</sup>. From a performance perspective, evidence suggests deep sleep is especially crucial for physical recovery. One sports medicine article noted an athlete's body "repairs muscles best when ~50% or more of sleep is spent in deep sleep" – emphasizing the value of maximizing deep sleep before competitions <sup>44</sup>. While 50% is higher than most people naturally get, it underscores that *more deep sleep = more physical recovery*. On the other hand, REM sleep loss can impair cognitive sharpness, decision-making, and emotional balance, which are vital on game day. Therefore, Athlytx's Sleep Score should reward not just total sleep but the composition: adequate deep and REM portions. If wearable data shows unusually low deep sleep (e.g. <5% of the night) or very short REM, the Sleep Component would be lower even if total hours were okay, reflecting likely incomplete recovery.

**Latency and Restfulness:** Sleep latency (how quickly one falls asleep) is a revealing metric. Ideally, it takes 15–20 minutes to fall asleep – that indicates a normal level of tiredness at bedtime <sup>45</sup>. Falling asleep almost immediately (<5 minutes) can be a sign of sleep deprivation or extreme exhaustion <sup>46</sup>. Conversely, taking much longer than 20 minutes may indicate insomnia or stress. Either scenario (too fast or too slow) is considered undesirable in Oura's scoring, and Athlytx can adopt a similar heuristic. We will incorporate latency as a minor factor – very low latency will slightly reduce the sleep score (flagging that the user might be overtired even if they got a full night), whereas a moderate latency is ideal. **Restfulness** (or its inverse, restlessness) tracks awakenings and movement. Lots of tossing and turning, or multiple awakenings, fragment one's sleep cycles and reduce restorative value <sup>47</sup>. Our algorithm can use the wearables' motion data or reported sleep disturbances to gauge restfulness. For example, Oura's *Restfulness* contributor looks at how much the user moved or got up during the night <sup>48</sup>. If the user's device logged frequent motion or an elevated resting heart rate late into the night (indicating unrest), Athlytx will interpret that as less effective sleep.

**Oura Sleep Score Model:** Oura's sleep score provides a great template with its **seven contributors** <sup>49</sup>: Total Sleep, Efficiency, Latency, REM, Deep, Timing, and Restfulness. We have covered the main ones except *Timing*. Timing refers to aligning sleep with circadian rhythm – essentially, whether you go to bed and wake up at biologically appropriate times. Oura considers a mid-sleep point roughly between 12am-3am as optimal for most (i.e. not too far off "normal" human diurnal rhythm) <sup>50</sup>. If someone's schedule is shifted (e.g. they sleep 3am-10am), their Timing score might drop, which can reflect suboptimal alignment even if hours are sufficient. For Athlytx, we can incorporate a simplified timing factor or simply let the user's sleep consistency be a factor (sleeping and waking at consistent times each day supports better recovery).

Overall, the *Sleep Component in Athlytx will combine*: nightly sleep duration vs. need, sleep efficiency, deep/REM amounts, and an indicator of sleep disturbances (restfulness), possibly yielding a sub-score out of 100. This will mirror validated scoring systems like Oura's, which users find informative and which correlate with how recovered they feel. Moreover, highlighting specific issues – e.g. "low deep sleep" or "short total sleep" – can educate users on where to improve. Since sleep drives changes in HRV and recovery, this component will heavily influence the overall Athlytx Score. For instance, if a user's sleep score is poor (bad night of sleep), the next day's Athlytx Score will likely be capped or reduced, aligning with how WHOOP's recovery score often dips after insufficient or low-quality sleep.

## 4. Blending Data from Multiple Wearables

Many athletes use multiple devices or platforms (Garmin watch, WHOOP strap, Oura ring, etc.), each with its own metrics. A major challenge is *data integration*: ensuring Athlytx can ingest and reconcile overlapping or even conflicting data to produce a unified score. There is no universal standard for this, but we can follow best practices from data fusion research and practical experience:

**Establish a Hierarchy of Data Sources:** One recommended strategy is to identify a "primary" source for each type of metric and use others as secondary <sup>51</sup> <sup>52</sup>. For example, if a user has an Oura ring and a Garmin watch both providing HRV, we might choose one as the authoritative HRV source (perhaps Oura for nocturnal HRV which it specializes in) and use the other only if primary data is missing. The logic is to avoid double-counting or confusion from small discrepancies. In a data-fusion framework proposed by Levine (2017), you first *define the primary measurement* for the problem, then integrate secondary streams hierarchically based on their accuracy and relevance <sup>51</sup> <sup>52</sup>. Applying that: Athlytx could treat *training load* from Garmin/Strava as primary for workouts (since those capture exercise data well), *sleep and HRV* from Oura or WHOOP as primary for recovery (since those devices excel at overnight tracking), etc., if a user has all connected. Secondary sources would fill any gaps or provide cross-checks.

**Normalization and Scale Harmonization:** Different platforms have different scales (e.g. WHOOP strain 0-21, Garmin load might be 0-1000 for 7-day, Oura readiness 0-100). Athlytx will translate external scores into our internal 0-100 scale for consistency. For instance, a WHOOP Recovery score of 80% could be taken as is, while a Garmin Body Battery of 50/100 might equate to a moderate recovery level internally. We will rely on the actual underlying metrics whenever possible instead of the proprietary scores (e.g., use HRV value in ms rather than WHOOP's % score) to recompute Athlytx Score with our own formula. However, *leveraging validated metrics* is smart: if a user's WHOOP Recovery score is very low on a given day, our system should incorporate that warning (since it's derived from HRV and sleep too). Essentially, Athlytx will act as a meta-aggregator: blending inputs from multiple wearables into one coherent picture, rather than forcing the user to interpret three separate readiness scores.

**Avoiding Data Conflicts:** If data streams conflict (e.g., Garmin says the user slept 7h, Oura says 6h30, WHOOP says 6h45), the differences are usually minor. We can either average them or prefer the device most specialized for that metric. In this case, Oura/WHOOP might be more accurate for sleep than a watch, so we'd go with Oura's value. We can indicate in the `DailyMetric.dataSources` field which devices contributed each component <sup>53</sup>. For transparency, if there's a large discrepancy (say one device logged a nap and another didn't), we can flag it to the user or in our confidence score (discussed below).

**Data Fusion and AI:** As datasets grow, we can employ machine learning to improve multi-sensor integration. For example, using a weighted ensemble or regression model that learns how to predict performance or recovery outcomes from the combination of all available metrics. Over time, the system could personalize how much weight it gives to one device or metric based on which ones correlate best with the user's actual performance or self-reported feelings. This aligns with the idea of "system retraining via AI" to individualize the integrated algorithm <sup>54</sup> <sup>55</sup>. In practice, initially we will implement straightforward rules (priority orders, simple weighting), but later we could introduce a machine learning layer that adjusts weights of HRV vs. sleep vs. strain for each user based on their historical data patterns.

**Real-World Example:** If an athlete wears a Garmin watch (for workouts and daily HR), a WHOOP (for 24/7 strain and recovery), and an Oura (for sleep and HRV), Athlytx might do the following: Pull workouts from Garmin/Strava for detailed training load (pace, HR, etc.), use WHOOP's continuous HR data to not miss any effort and get WHOOP's Recovery (morning HRV) data, and use Oura's nightly HRV and sleep staging data for sleep quality. Our algorithm would then combine Garmin's training load, Oura's sleep quality, and WHOOP's HRV into our proprietary score. Each provides a piece of the puzzle – by integrating them, we present one cohesive readiness score rather than three separate ones. This multi-source approach is a selling point: Athlytx becomes the one-stop dashboard aggregating all your wearables. To implement this cleanly, our **backend** will need modular data importers and a unifying logic (see Implementation Plan). The result for users will be confidence that Athlytx isn't ignoring any data – it uses the "best of each device" to inform the score.

## 5. Personalization and Machine Learning

Every individual is different. A given HRV of 30 ms might be normal for one person and alarming for another; 6 hours of sleep could leave one athlete refreshed but another groggy. Thus, an optimal readiness algorithm must adapt to personal baselines and trends.

**Individual Baselines:** The first step in personalization is establishing each user's normal range for key metrics. For instance, tracking a user's average HRV and resting HR over a few weeks to set their baseline. Many apps do this: WHOOP, Oura, and others compare your current values to your own 30-day baseline. Athlytx will similarly calibrate the score to the individual. If an athlete usually has rMSSD around 50 ms and today it's 30 ms, that's a significant drop for them (even if 30 might be "average" in the population) – our score should reflect that as a warning. Conversely, someone with a normally low HRV shouldn't be penalized every day; it's the deviation from their norm that matters most. Research supports this approach: studies have found that *relative changes* in HRV (vs. personal baseline) are more useful for guiding training than absolute values <sup>3</sup>. Also, subjective scales work best personalized; e.g. if an athlete rates their sleep quality 2/5 when normally they rate 4/5, that's a red flag. Athlytx will track rolling averages for sleep hours, HRV, RHR, etc., and use those to contextualize daily readings.

**Machine Learning for Readiness:** We can leverage ML both for *tailoring the score* and for *predictive analytics*. Tailoring the score might involve a model that learns how each component (sleep, HRV, load,

etc.) correlates with the user's own performance or reported fatigue. For example, some athletes might be more sensitive to high training load, while others handle volume well but get affected by poor sleep. A machine learning model (even a simple weighted linear regression or a more complex classifier) could, over time, adjust the weighting of components in the Athlytx Score for each user. If our dataset grows (many users, longitudinal data), we could employ techniques like clustering or personalized regression to categorize users and adjust scoring algorithms accordingly. The literature indicates that combining multiple metric inputs improves training outcomes <sup>56</sup>, and it underscores using "advanced analytics" to personalize training guidance <sup>57</sup> <sup>58</sup>. In one study on cyclists, the group that trained with guidance from combined HRV+RHR+subjective data improved performance more than the HRV-only group <sup>59</sup>. This suggests an algorithm that dynamically responds to an individual's data (and perhaps flags when an extra rest day is needed) can tangibly benefit performance.

**Tailored Readiness Scores:** We envision the Athlytx Score eventually evolving beyond a one-size-fits-all formula. Initially, we'll implement a general algorithm based on the evidence, but machine learning can refine it. For instance, we might build a simple predictive model that takes a week's worth of metrics and tries to predict the user's next-day subjective readiness or performance (if they log how they feel or their training results). The model might learn that for a given user, sleep quality is the strongest predictor, whereas for another user, acute training load is the key driver. Then the algorithm could upweight or downweight components in the readiness score formula accordingly. This user-specific feedback loop is akin to having a coach who learns an athlete's unique responses. It's important, however, to maintain transparency – users should understand their score and not feel it's a "black box." So any ML adjustments should be subtle and always explainable in terms of the components (e.g. "you seem to handle high strain well, so the model is less concerned with your strain score and more with your sleep").

**Predictive Readiness and Trends:** Beyond daily scoring, personalization enables *predictive analytics*. For example, if we detect patterns (like a persistent drop in HRV and rise in RHR over several days), we could predict an impending performance drop or illness. Researchers modeling stress-recovery status have noted that integrating HR and HRV trends can detect abnormal recovery states <sup>60</sup>. Our system could provide a forecast like "At this trajectory, your readiness might be low by the weekend – consider adding recovery." Also, using historical data, we might forecast how ready a user will be for an upcoming event (e.g. "If you maintain your current training and recovery, your projected Athlytx Score on race day is 90, indicating high readiness"). These predictive features will rely on ML models trained on sequences of data. However, implementing this will come after we have accumulated enough user data and verified the accuracy of such predictions.

In summary, personalization in Athlytx will mean **continual calibration**. On the simplest level, it's baseline comparisons and individualized thresholds. On a more advanced level, it's ML models that tune the algorithm per user and even predict future readiness. This ensures the Athlytx Score is *relevant to you*, not just a generic population metric.

## 6. Confidence Scoring and Fallback Logic for Missing Data

No data system is perfect – wearables can fail to record, users forget to wear devices, or certain metrics might be missing on a given day. Instead of outputting a potentially misleading score, Athlytx will implement confidence scoring and smart fallback strategies:

**Confidence Index:** Each daily Athlytx Score will carry a confidence level (e.g. ○○○ out of 5, or a percentage) that reflects how complete and reliable the input data was. For instance, if a user didn't wear their HRV device last night, our recovery component is missing a key input. The score might still be

computed using other data (like their subjective rating and yesterday's strain), but we would tag it as "low confidence" because an important piece is absent. This practice is analogous to having error bars on a measurement. It communicates to the user and coaches that *we have less certainty in the score today*. WHOOP does something similar informally – if your overnight HRV reading is flagged as poor quality, they caution interpreting that day's recovery score. We will formalize it into our UI: perhaps with a color or icon indicating confidence.

**Missing Data Fallbacks:** When data is missing, Athlytx will use hierarchical fallback logic: - If last night's sleep data is missing (device not worn), we could default to using the previous night's data or the user's average sleep as a proxy, but only if needed. We'll clearly indicate that sleep was not recorded. Alternatively, the score might rely more heavily on training load and subjective measures that day. The algorithm might say, "Sleep data unavailable – using 7-day sleep average for today's score (confidence reduced)." - If HRV is missing but we have resting HR and perhaps WHOOP recovery, we might use resting HR and any subjective recovery score as a stand-in for the recovery component, again with reduced confidence. - If training load is missing (e.g. user didn't log an activity and isn't wearing any device that tracks steps/heart rate), we could assume it was a rest day (zero strain) unless other evidence suggests otherwise. - In general, the system will favor *using whatever data is available* rather than giving up entirely. Partial data is still valuable, but we must mark the limitations.

**Medical Best Practices:** In healthcare, dealing with missing data often involves techniques like *imputation* or carrying forward last values, but always with caution. We will adopt a conservative approach: prefer to under-estimate readiness if critical data is missing, to avoid false confidence. For example, if we lack both HRV and sleep for a day, the Athlytx Score might be capped at a moderate level with a low confidence flag, even if training load was low (because we truly don't know if the user recovered). This errs on the side of safety, akin to how a doctor would be cautious making a diagnosis with incomplete test results.

**Quality of Data:** Confidence scoring also accounts for data quality issues. If a heart rate trace is noisy or an HRV reading had too many artifacts, we downgrade confidence. The user interface can display messages like "⚠️ HRV reading was inconclusive last night" or "⚠️ Optical HR signal lost during workout – strain estimate uncertain." Studies in wearable tech have noted that the accuracy of input data heavily influences the reliability of the metrics <sup>61</sup>. For example, a faulty HR max setting or a glitchy optical sensor can skew training load calculations <sup>62</sup>. By monitoring such anomalies, we can adjust the confidence. If a user's heart rate data looked erratic (perhaps they wore the watch loosely), the strain score might still calculate but we'd mark it as lower reliability.

**User Transparency:** We will include in the UI a breakdown of which sources contributed each day (this is already captured in the `DailyMetric.dataSources` JSON <sup>53</sup>). Additionally, the Athlytx Score card could have an info tooltip that says "Today's score is based on: Sleep – from Oura (7h, good quality), HRV – missing data, RHR – 60 bpm (above baseline), Training load – 300 (high), Subjective – 4/5. **Confidence: Low (missing HRV)**." This level of transparency builds trust and helps users understand when to trust the score fully versus when to be a bit cautious. Over time, users will appreciate that Athlytx doesn't try to sweep missing data under the rug – we acknowledge it and adapt.

For development, implementing this means writing robust checks in the score calculation module: e.g. if `hrvAvg` is null for the day, set a flag and adjust calculations to exclude HRV (or substitute baseline). We might assign weights dynamically based on available data. We'll also create a *confidence score* (perhaps a simple percentage of how much of the "ideal" data is present, weighted by importance of each metric). If sleep, HRV, training load, and subjective are all present, confidence = 100%. If one of four is missing, maybe confidence = 75%, etc., adjusted if the missing one is particularly important (HRV and sleep are quite significant). This confidence value can then drive UI indicators.

---

Having synthesized the research and best practices above, we now propose an **enhanced Athlytx Score algorithm** and a roadmap to implement it, covering backend logic, UI improvements, trend tracking, predictive features, and validation.

## Implementation Plan for Athlytx Score 2.0

### A. Algorithm and Formula Improvements (Evidence-Based)

Based on the research insights, we will formulate the Athlytx Score as a weighted composite of at least three components: **Recovery**, **Strain**, and **Sleep**, with personalization and context-aware modifiers. In the database, these correspond to `recoveryComponent`, `strainComponent`, `sleepComponent` for each day <sup>63</sup>, which will now be computed with updated formulas:

- **Recovery Component (0-100):** Calculated each morning from HRV, resting HR, and subjective readiness (and potentially any morning orthostatic test or questionnaire data). For example, we can derive a recovery score where 100 is “fully recovered”. A possible formula (to be refined) could be:

$$\text{Recovery} = 100 - a \cdot (\text{RHR}_{\text{dev}}) - b \cdot (\text{HRV}_{\text{dev}}) - c \cdot (\text{subjectiveFatigue})$$

where  $\text{RHR}_{\text{dev}}$  is the deviation of resting HR from baseline,  $\text{HRV}_{\text{dev}}$  is the negative deviation of HRV from baseline (i.e. how much lower than your 7-day average), and `subjectiveFatigue` is a scaled value from a morning questionnaire. The coefficients  $a$ ,  $b$ ,  $c$  will be tuned (possibly via regression on pilot data). The inclusion of both HRV and RHR aligns with research that a combined index is more indicative <sup>7</sup>. We could also integrate a *HRV/RHR ratio* as the swimmers' study suggested <sup>4</sup>. For instance, use rMSSD divided by average R-R interval (or equivalently, rMSSD \* RHR) as one input to capture parasympathetic vs sympathetic balance <sup>64</sup>. If available, we incorporate recovery scores from devices: e.g. if WHOOP recovery% or Oura Readiness is provided via API, we can map that onto our 0-100 scale and compare with our internally computed value, perhaps averaging them for robustness.

- **Strain (Training Load) Component (0-100):** Computed on an ongoing basis each day and finalized at end of day. We will implement a **Banister TRIMP-based calculation**: using either heart rate zones or training impulse formulas. Each workout contributes points; non-exercise activity can contribute small points. We'll likely adopt Strava's approach where the highest-intensity efforts get significantly higher weighting <sup>17</sup>. The raw daily strain might be, say, 0-400 points (which could correspond to minutes in various zones). To convert to 0-100, we'll define what constitutes “100” strain (perhaps a very hard training day for that user, like their max in last 30 days). Alternatively, we might not strictly cap at 100 – some systems don't (TrainingPeaks TSS can exceed 100). But for a consistent Athlytx Score, we might normalize strain as a percentage of the user's weekly average. For instance, if the user's typical daily load is 50 and today they did 100, that might be considered a 100 (an extremely high day). This ensures personalization (a pro athlete with huge volume is judged against their own norm). Also, we incorporate acute vs chronic context: If the user has chronically high fitness (high CTL), a given strain might impact them less. We could reflect that by adjusting today's strain score down if their 7-day load isn't high relative to 28-day (meaning they're accustomed). However, to keep it simple, the strain component will initially just reflect today's internal load score relative to personal scale. It will also include non-training stress: e.g. a high number of steps or high all-day average heart rate (which might indicate a physically active job or high stress day) will bump up strain.

- **Sleep Component (0-100):** Derived after each night's sleep. We'll implement a scoring function that takes: total sleep vs. target (based on user's set goal or 8h default), sleep efficiency, deep sleep %, REM %, and sleep latency. Each factor can contribute some points. For example, we could assign 0-40 points for sleep duration (0 if <4h, 40 if  $\geq 8h$ ), 0-20 for sleep efficiency (0 if  $<70\%$ , 20 if  $\geq 90\%$ ), 0-20 for deep+REM (0 if very low deep/REM, 20 if hit targets), and 0-20 for latency and restfulness combined. These numbers can be adjusted with testing. The sum becomes the sleep quality score. Oura's Sleep Score logic can guide our thresholds 65 41 38. We will also allow the user's *sleep debt* or past nights to influence it: e.g. if they've been undersleeping all week, even one decent night might not fully restore them, so we might factor a multi-day sleep bank. This can feed into Recovery component too (Oura Readiness does this by considering prior sleep debt).
- **Aggregate Athlytx Score:** Finally, the Athlytx Score itself (0-100) will be a weighted blend of Recovery, Strain, and Sleep components. A starting proposition is to weight Recovery (acute recovery state) and Sleep (last night's effect) more heavily in determining *today's readiness*, while Strain (training load) influences readiness inversely (more strain usually lowers next-day readiness). For example, formula:

$$\text{Athlytx Score} = w_r \cdot \text{Recovery} + w_s \cdot \text{Sleep} - w_t \cdot \text{StrainAdjustment}$$

where StrainAdjustment could be something like a scaled value of yesterday's strain or cumulative fatigue. We might set  $w_r = 0.4$ ,  $w_s = 0.3$ ,  $w_t = 0.3$  initially. If the user took a rest day (low strain), then strainAdjustment is small and doesn't penalize the score. If they had an unusually hard training day yesterday, we subtract some points unless their Recovery score (HRV etc.) miraculously remained high. Essentially, the model should capture that *high strain needs high recovery to compensate*. If both strain was high and recovery metrics are poor, the score will be low (the athlete is likely in need of rest). On the other hand, if strain was high but recovery metrics are still good (indicating the athlete handled it well), the score might only be moderately reduced. We will test and iterate on these weightings with pilot users.

- **Personalization & Baselines:** The algorithm will incorporate personal baselines at multiple points. We'll maintain running averages for HRV, RHR, and sleep for each user, updating them continuously. The formula will use deviations from these rather than absolute values, as discussed. It will also adjust for the user's typical training load. If someone's average daily strain is 150 (a high-volume athlete), a day of 150 should perhaps count as "moderate" strain for them, whereas for a novice with average 50, a 150 day is huge. So normalization by individual history is key, as noted above with Strava's normalization across users 24. We'll implement functions that map raw metrics to scores using the individual's context. For example, a percentile approach: convert today's HRV to a percentile relative to the user's last 60 days of HRV values – if it's in the bottom 10%, that's a likely poor recovery (score it low), if in top 10%, score high.
- **Confidence Calculation:** Alongside the score, our algorithm module will compute a confidence score as discussed. This likely won't affect the numeric Athlytx Score (we won't, say, reduce the score itself, just flag it). But one approach could be to incorporate it if needed: e.g. if data is missing, we could regress the score toward 50 (neutral) slightly to avoid extreme scores on partial data. However, it might be better to keep the score formula pure and handle confidence in presentation.

We will document the exact formula in the code (and possibly a whitepaper) for transparency. The modular design means the code will likely have separate functions: `computeRecoveryScore(userData)`, `computeSleepScore(userData)`,

`computeStrainScore(userData)`, which then feed into `computeAthlytxScore(...)`. This modular approach allows swapping in data source-specific calculations (like if only WHOOP is connected vs only Garmin, we adjust accordingly).

## B. Backend Updates (Modular Architecture & Integrations)

To support the new algorithm and multi-source data, the backend needs enhancements in data management and processing:

- **Modular Data Ingestion:** We will refactor the data sync service (`backend/services/syncService.js` or a new service) <sup>66</sup> to pull in data from all connected providers on a daily schedule. Each provider (Garmin, Oura, WHOOP, Strava) has different endpoints and data structures. We have existing OAuth integration points <sup>67</sup> <sup>68</sup>. Next, we need to fetch and store daily metrics from each: for example, Garmin's daily summary (`GET /api/garmin/v2/dailies`) provides steps, calories, Body Battery, etc.; Oura's `sleep` and `readiness` endpoints provide nightly details <sup>69</sup>; WHOOP's `recovery` and `sleep` endpoints provide HRV, sleep, etc. <sup>70</sup>. We will expand the `DailyMetric` model population. The `DailyMetric` table already has fields for `hrvAvg`, `restingHr`, `sleepHours`, `sleepQuality`, `trainingLoad`, etc. <sup>71</sup>. We'll map data from each source into these fields:
  - Fill `hrvAvg` with overnight HRV from whichever device (WHOOP Recovery's RMSSD, or Oura's average HRV).
  - Fill `restingHr` with the lowest overnight resting heart rate (both Oura and WHOOP provide this, Garmin dailies might as well).
  - Fill `sleepHours` from Oura/WHOOP or Garmin sleep duration.
  - Fill `sleepQuality` – perhaps a composite or the sleep score from Oura if available.
  - Fill `trainingLoad` – from Garmin's 7d load or Strava's Relative Effort (we might calculate our own if needed).
- The `dataSources` JSON can record details like  
`"HRV": "Oura", "Sleep": "WHOOP", "Load": "Garmin"`.
- **Calculation Engine:** We will implement the Athlytx Score calculation in the backend (likely as part of the sync job or a separate cron task that runs after data ingestion). The calculation will use the modular functions described. This could be done in JavaScript within the Node backend, or even in SQL if simple, but likely JS for flexibility. We will ensure that if not all data is present, the functions handle nulls gracefully (applying the fallback logic described).
- **Garmin Integration:** Special attention to Garmin – Garmin has its own metrics like Body Battery and stress score. In v2, we have endpoints for Garmin dailies <sup>72</sup>. We should parse Body Battery highs/lows, daily stress, etc. Body Battery can indirectly inform our recovery score (if Body Battery went up to 90 in the morning, user is well-recovered). We may incorporate those values in future versions. In this phase, integrating Garmin mainly means using their sleep and activity summaries to supplement data if Oura/WHOOP are not present.
- **Database Adjustments:** The current schema seems sufficient (fields exist for what we need). We might add fields for additional nuance (for example, an `hrvBaseline` or `ctl` / `atl` value for each day if we want to store those for trend calculations). But these can also be computed on the fly. We will likely add a *confidence* field in `DailyMetric` or a separate table to store the confidence score for that day's Athlytx Score. Alternatively, we can compute confidence on the fly each time we present the data.
- **Ensuring Data Consistency:** We will implement checks to avoid duplicate data entries or overlaps. If a user has multiple connections (e.g., both Oura and Garmin providing sleep for the same night), our sync service should merge them logically instead of double-counting. Possibly, we give precedence or average values. This logic can be configured per user (maybe let user

select primary device for certain metrics in settings eventually). Initially, we can hardcode priorities.

- **APIs for Coaches/Athletes:** Since coaches will view athletes' scores, our new data should be exposed via the dashboard API <sup>73</sup>. We'll update `/api/athlete/dashboard` to include the Athlytx Score, sub-scores, and trend data for the requested period <sup>73</sup>. This means this endpoint will run the query to fetch DailyMetrics for last X days and return those values. We will ensure that the calculations either are stored in DB or computed on request. Storing daily Athlytx Score in the DB (it looks like `DailyMetric.athlytxScore` exists <sup>71</sup>) is useful so we don't recalc every time unless inputs changed.
- **Real-time considerations:** Most of these metrics update once per day (morning after sleep). We might set up the sync so that early each morning, once devices have uploaded their nightly data, we compute the new score. Possibly integrate a push or webhook if available (some APIs allow push of new data). But a daily cron job is simplest.
- **Modularity for New Devices:** Our design should allow adding future integrations (e.g., Apple Health or Fitbit) by writing a new importer that maps data into our model. By separating device-specific logic from the core algorithm, we ensure extensibility.

## C. User Interface Enhancements

The UI will be revamped to surface the new Athlytx Score in a clear, informative way, with transparency and context. Proposed UI changes:

- **Athlytx Score Card:** On the athlete dashboard, one of the four key metric cards is Athlytx Score <sup>74</sup>. We will redesign this card to display:
  - The **Athlytx Score** as a prominent number (0–100) with a colored ring or gauge (e.g., green for high, yellow moderate, red low).
  - A subtitle or icon indicating the confidence (e.g., a "?" icon if low confidence, hover text "Missing HRV data").
  - Possibly small sub-indicators or a sparkline showing trend (e.g., an up/down arrow if score is higher or lower than yesterday).
- **Component Breakdown:** When the user taps/clicks the Athlytx Score card, we'll show a breakdown of components:
  - Recovery Score (with maybe a short label like "Recovery: 85/100 – HRV above baseline"),
  - Strain Score ("Strain: 70/100 – Yesterday's training was high"),
  - Sleep Score ("Sleep: 90/100 – 8h good quality"),
  - Any other factors (perhaps an illness flag or something if detected). This addresses transparency – users see what's driving the score. Each of those could have tooltips explaining "Recovery is calculated from HRV and resting HR <sup>2</sup>. Your HRV was 5% below baseline (slight fatigue)." etc. This level of detail is inspired by Oura's approach (they show contributors for Readiness like HRV balance, sleep balance, previous day activity, etc., with notes if something is off).
- **Trends and History:** We will implement charts for daily and long-term trends. For example, a line chart of Athlytx Score over the past 30 days (perhaps overlayed with a line for 7-day average to smooth it). This helps users and coaches spot patterns, like steadily declining scores signaling accumulating fatigue, or how a deload week boosts scores. The **dashboard** might include a "Training Load vs Readiness" chart: since we already plan a CTL/ATL chart <sup>75</sup>, we could overlay Athlytx Score or color the CTL/ATL chart background by readiness (high readiness days vs low).
- **Notifications and Alerts:** UI can incorporate gentle guidance based on the score. For example, if Athlytx Score is very low (e.g. <40), show a suggestion like "Your readiness is low – consider

prioritizing recovery today." Or if consistently low for a week, maybe "Scores have been low all week; ensure you're recovering adequately or reduce load." These are not strict prescriptions but educational nudges. Conversely, high score could prompt "You're in great shape today – a hard workout is manageable."

- **Confidence and Data Transparency:** As discussed, indicate missing data. Perhaps an icon next to the score or in the breakdown where something is missing (e.g., if no HRV, the Recovery line might say "Recovery: -- (HRV data not available)"). We want to avoid user confusion if, say, they forgot their device and see a weird score – the UI should make it clear why.
- **Coach View:** On the coach dashboard (which can see multiple athletes), it might show each athlete's latest Athlytx Score and maybe a colored dot (like green/yellow/red) for quick status. This allows a coach to scan who is at risk of overtraining or who is primed for hard training. We will add Athlytx Score to the coach's athlete list and detail pages.
- **Customization:** Eventually, we might let users customize how their score is calculated (like toggling "include subjective" or adjusting their sleep need). In this iteration, we'll focus on making the UI accommodating enough to display all new info without overwhelming the user. The design tokens and glassmorphism style mentioned for v2.0 will be used to keep it visually appealing <sup>76</sup> <sup>77</sup>, but with added data complexity we need to maintain clarity.

## D. Daily and Long-Term Trends

As part of the UI and backend, tracking trends is important: - We will store historical Athlytx Scores and component values in the database (DailyMetric table). This allows queries for trends over any period. - In the UI, we implement charts: likely using Chart.js as planned <sup>78</sup>. A line chart for Athlytx Score over time, and perhaps separate lines for Recovery, Strain, Sleep components (the user could toggle them). Another useful visualization is an *area chart or bar chart by day* showing contributions – e.g., stacked bars where one color is strain, one is sleep, etc., illustrating how the components varied. - We will integrate the trend with CTL/ATL chart. For example, overlay the readiness score on the Training Load chart <sup>79</sup> <sup>75</sup>. This can show relationships (e.g., every time acute load spiked far above chronic, readiness dipped accordingly a couple days later). - **Long-term analytics:** Provide a "trend summary" that says things like "Over the last 60 days, your average Athlytx Score was 78. Your best week was Oct 5-11 (avg 85) when you had consistently good sleep. Your lowest scores happened following the marathon race on Nov 2." Such insights can be generated once we have enough data, possibly using the AI module to find patterns. This is more of a phase 2 feature, but we mention it as a goal to help users learn from the data.

## E. Predictive Readiness Features

With the foundation in place, we can experiment with predictive modeling: - **Next-Day Prediction:** Using the last week or two of data, predict tomorrow's score (or range). This could simply be regression to the mean (most likely tomorrow will be around your recent average) adjusted by known factors (e.g., if you just had two very hard days and your HRV is trending down, tomorrow will likely be low). We might incorporate this as "Projected Tomorrow: 65 (if you train as planned and sleep normally)". This helps planning training – for example, an athlete might postpone a hard interval session if they see their projected readiness is poor. - **Event Readiness:** If a user inputs an upcoming race date, we could simulate what their score might be given certain taper strategies. This requires modeling how training load reductions and extra rest translate to score increases. Based on fitness-fatigue dynamics <sup>22</sup>, if you taper (lower acute load) while maintaining some chronic fitness, your readiness should peak. We can use the CTL/ATL model to estimate that. For instance, if race is in 10 days, and we know current ATL and CTL, we could suggest "to maximize readiness, aim for an ACWR of ~0.8 in the final week" and forecast the score. - **Injury Risk Alerts:** Although not explicitly asked, coupling ACWR and readiness allows soft predictions of injury risk (when load is high and readiness continuously low). Athlytx could alert "Injury

risk elevated – your acute load is high while recovery scores are low. Consider additional rest." This feature would use thresholds from research (like ACWR >1.5 with poor recovery metrics is a red flag). - **Machine Learning Models:** We might deploy a simple model using logistic regression or decision trees on our internal data (once we have enough) to predict whether a user will have a "good" or "bad" readiness day, or to classify days as "optimal, warning, or critical". This predictive classification could then feed into recommendations.

We will likely pilot predictive features carefully, as erroneous predictions could hurt credibility. They will be presented as probabilities or suggestions, not guarantees. For example, "It's likely (~70% chance) you'll feel recovered by tomorrow if you get at least 8 hours sleep tonight."

## F. Pilot Testing and Validation

Before full rollout, we will run a pilot program to test the new score algorithm: - **Internal Testing:** Use our own data (developers or a small group of test users with wearables) to compute Athlytx Scores and see if they intuitively match perceived readiness. We'll tweak the weights/formulas as needed. We can also simulate scenarios from research papers (e.g., feed in data corresponding to an overtraining period vs a taper and see if the score behaves as expected – rising during taper, dropping during overload <sup>64</sup> ). - **Comparison with Established Metrics:** We will compare Athlytx Score against WHOOP Recovery and Oura Readiness on days where we have those. A high correlation would validate we're capturing similar phenomena. If there are discrepancies, we investigate why – maybe our score is considering something they don't (or vice versa). We can adjust or at least know how to explain differences. - **User Feedback:** Select a few beta users (possibly coaches and athletes known to us) to use the updated platform. Gather subjective feedback: do they feel the score reflects their actual readiness? Are the insights helpful? This qualitative feedback is crucial to fine-tune how we present information. For example, maybe users want a simpler summary or find one of the sub-metrics confusing – we can address that. - **Scientific Validation:** While a formal study is beyond initial scope, we can do a mini validation: e.g., measure correlation between Athlytx Score and performance metrics like running pace in a workout, or how score trends correlate with when the athlete says they felt "in the zone" vs "burnt out." If possible, we could partner with a coach to track a small team: have them follow Athlytx Scores for a training cycle and note any improvements in managing training (this could later become a case study or even publication if results are strong). - **Iterate Quickly:** Given our 2-3 week build timeline for MVP in earlier plan <sup>80</sup>, we should allocate some days in Phase 7 (Premium Features) or Phase 8 (Testing) specifically for refining the scoring algorithm based on pilot results. The beauty of having it modular is we can adjust weights or logic without overhauling the whole system.

- **Validation Metrics:** We can set some target goals, like: Athlytx Score < 50 should correspond with user reporting high fatigue X% of the time, Score > 80 should correspond with user feeling great Y% of time. Also, if we have any objective markers (like did the user complete their planned workout or had to abort it due to fatigue), we could see if low scores preceded those events. We might use such analysis to claim "Athlytx Score predicts low-energy days with Z% accuracy."

Finally, after pilot and iteration, we'll proceed to full deployment. Throughout, we will maintain references to the scientific rationale for marketing/communication purposes – e.g., we can publish a blog or whitepaper citing key studies (like HRV-guided training benefits <sup>81</sup> <sup>82</sup>, or the importance of sleep for athletic performance <sup>44</sup>) to lend credibility to Athlytx's methodology.

## G. Roadmap and Future Considerations

After launching the improved Athlytx Score, the roadmap could include: - Integration of additional metrics like *blood biomarkers* or *nutrition* if available (for example, if a user logs soreness or hydration,

that can adjust readiness). - Incorporating *mental health/stress questionnaires* more formally, given how important psychological state is <sup>8</sup>. - Continuous learning: deploying the ML personalization so the system gets smarter the more you use it. - Possibly seeking *third-party validation*: e.g., working with a sports science lab to test Athlytx Score against lab measures or against other products. - Adding an *injury likelihood score* or recommending *daily optimal training zones* (e.g., "today is a good day for high-intensity" vs "keep it easy today"), essentially turning readiness into actionable training guidance - many users find that valuable (WHOOP does something like "suggested strain target" based on recovery).

By implementing the above plan, Athlytx will significantly enhance its core value proposition: an evidence-based, personalized readiness score that helps athletes train smarter and recover better. With transparent components and integration of the best practices from Garmin, WHOOP, Oura and the latest research, Athlytx Score 2.0 will rival or exceed the insights of leading platforms, all while presenting the information in a user-friendly, coach-friendly manner.

## References:

- Buchheit, M. (2014). "HRV: Ready or Not?" – discusses HRV use in training decisions <sup>11</sup>.
- Altini, M. (2020). *On Heart Rate Variability, Training Load and Subjective Metrics* – emphasizes integrating objective and subjective data <sup>11</sup>.
- Firstbeat Technologies. *White Paper: Understanding Athlete Training Load* – defines TRIMP-based load, acute vs chronic load <sup>19</sup> <sup>16</sup>.
- Strava Engineering (2018). *The Science of Relative Effort* – explains how Relative Effort builds on TRIMP with intensity weighting <sup>23</sup>.
- the5runner (2025). *Strava Relative Effort Guide* – notes Strava's use of zone weighting and normalization across sports <sup>23</sup> and how RE relates to Garmin's EPOC-based load <sup>27</sup>.
- WHOOP (2023). *How Does WHOOP Strain Work?* – describes the 0–21 strain scale and inclusion of muscular load <sup>28</sup> <sup>30</sup>.
- Oura Support. *Sleep Score Contributors* – outlines the factors in Oura's Sleep Score (sleep stages, efficiency, latency, etc.) <sup>36</sup> <sup>38</sup>.
- Stephenson, M. (2024). *How Does Sleep Affect Athletic Performance?* – highlights importance of deep sleep for muscle recovery (50% of sleep in deep) <sup>44</sup>.
- Scientific Reports (2025). *HRV, RHR, and Well-being in Cyclists* – showed combining HRV, resting HR, and subjective wellness improved training outcomes and stresses that neither metric alone suffices <sup>56</sup> <sup>15</sup>.
- Bulte et al. (2025). *HRV Readiness in Swimmers* – found HRV/RHR ratio more responsive to overload vs taper <sup>4</sup> <sup>5</sup>.
- Gabbett, T. (2016). *ACWR and Injury Risk* – introduced acute:chronic workload ratio to manage fatigue and injury (the "sweet spot" ~0.8–1.3, danger >1.5) <sup>83</sup> <sup>21</sup>.
- Various sources from Garmin, WHOOP documentation for algorithm insights <sup>12</sup> <sup>84</sup>.

---

<sup>1</sup> <sup>2</sup> <sup>6</sup> <sup>7</sup> <sup>8</sup> <sup>13</sup> <sup>14</sup> <sup>15</sup> <sup>56</sup> <sup>57</sup> <sup>58</sup> <sup>59</sup> <sup>81</sup> <sup>82</sup> Individual training prescribed by heart rate variability, heart rate and well-being scores in experienced cyclists | Scientific Reports  
[https://www.nature.com/articles/s41598-025-13540-z?error=cookies\\_not\\_supported&code=76dff965-0aae-4171-bd96-111083a7bd0f](https://www.nature.com/articles/s41598-025-13540-z?error=cookies_not_supported&code=76dff965-0aae-4171-bd96-111083a7bd0f)

<sup>3</sup> <sup>11</sup> On Heart Rate Variability (HRV), training load and subjective metrics  
<https://www.hrv4training.com/blog2/on-heart-rate-variability-hrv-training-load-and-subjective-metrics>

- 4 5 64 Use of Heart-Rate Variability to Examine Readiness to Perform in Response to Overload and Taper in Swimmers - PubMed  
<https://pubmed.ncbi.nlm.nih.gov/40360152/>
- 9 Monitoring the athlete training response: subjective self-reported ...  
<https://bjsm.bmjjournals.org/content/50/5/281>
- 10 Monitoring the athlete training response: subjective self ... - PubMed  
<https://pubmed.ncbi.nlm.nih.gov/26423706/>
- 12 Body Battery Frequently Asked Questions | Garmin Customer Support  
<https://support.garmin.com/en-US/?faq=VOFJAsiXut9K19k1qEn5W5>
- 16 18 19 content.firstbeat.com  
[https://content.firstbeat.com/hubfs/Sports/Sports%20Guides/Sports%20Guides%20\(English\)/ENG-Sports-Guide-Training%20Load.pdf](https://content.firstbeat.com/hubfs/Sports/Sports%20Guides/Sports%20Guides%20(English)/ENG-Sports-Guide-Training%20Load.pdf)
- 17 23 24 25 26 27 61 62 84 Strava Relative Effort – Everything You Need to Know (2025 Update)  
<https://the5runner.com/2025/11/17/strava-relative-effort-guide-tss-2025/>
- 20 22 33 34 Acute:Chronic Workload Ratio  
<https://www.scienceforsport.com/acutechronic-workload-ratio/>
- 21 The acute:chronic workload ratio – science or religion?  
<https://www.sportsinjurybulletin.com/improve/the-acutechronic-workload-ratio--science-or-religion>
- 28 29 30 31 32 How Does WHOOP Strain Work? | WHOOP  
<https://www.whoop.com/au/en/thelocker/how-does-whoop-strain-work-101/>
- 35 36 37 38 39 40 41 42 43 45 46 47 48 49 50 65 Sleep Contributors – Oura Help  
<https://support.ouraring.com/hc/en-us/articles/360057792293-Sleep-Contributors>
- 44 How Does Sleep Affect Athletic Performance? | Mass General Brigham  
<https://www.massgeneralbrigham.org/en/about/newsroom/articles/sleep-and-athletic-performance>
- 51 52 54 55 A Five-Step Strategy to Combine Data Sources from Multiple Wearable Sensors  
<https://www.scirp.org/journal/paperinformation?paperid=73908>
- 53 63 66 67 68 69 70 71 72 README.md  
<https://github.com/zwiersd/athlytx-backend/blob/40206c1d389faea72dce715c8673dd893ef4b23d/README.md>
- 60 Modeling Stress-Recovery Status Through Heart Rate Changes ...  
<https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2020.576308/full>
- 73 74 75 76 77 78 79 80 ATHLYTX-ELITE-BUILD-PLAN.md  
<https://github.com/zwiersd/athlytx-backend/blob/40206c1d389faea72dce715c8673dd893ef4b23d/Documentation/ATHLYTX-ELITE-BUILD-PLAN.md>
- 83 Editorial: Acute: Chronic Workload Ratio: Is There Scientific Evidence?  
<https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2021.669687/full>