Predictive Detection of Cognitive Degradation in Operational Environments Using Wearable Sensors

Leslie Saxon, Jill Boberg, Matthew Bartels, Trevor Barrett
Center for Body Computing · USC Institute for Creative Technologies
MHSRS-26-5107 · ADVANCED PRODUCT DEVELOPMENT

USC Institute for Creative Technologies

USC Center for Body Computing

0.65 vs 0.58
Detection accuracy (AUC) in the **high-demand unit** vs the **opposing force**; directional across both rotations

57% / 82%
Of the training unit below the vendor's BAC > 0.08% effectiveness tier on day 3 in the box (vs 14–17% OPFOR)

53%
Of objectively degraded days **went unreported**: the soldier reported no impairment

465
Service members enrolled across four operational deployments; 336 wearable-instrumented

The Gap

The Department has no scalable method to detect cognitive degradation in the field. Fitness-for-duty decisions rely on self-report; as this poster shows, more than half of impaired soldiers report no impairment.

The Capability

Continuous wearable sleep and activity sensing, a vendor fatigue model, continuous glucose, and a 3-minute mobile cognitive battery, fielded end-to-end under field conditions for a full rotation.

The resulting detection model is deliberately **low-cost**: three sleep and time-awake channels a wrist device already produces, plus **one tap** of self-reported sleepiness, each referenced to **the soldier's own baseline**.

Design

Two full National Training Center rotations, Jan–Apr 2024. 93 soldiers, 1,389 person-days, 1,459 PVT sessions, 937 instrumented nights.

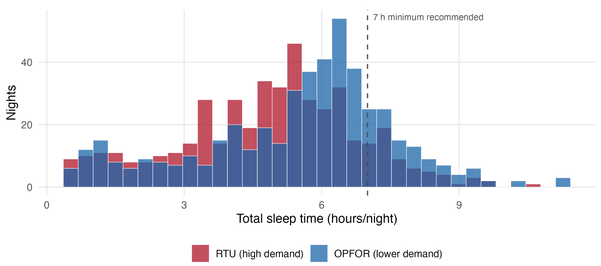
Two cohorts share the same maneuver area (the NTC "box"), weather, and dates, but carry very different operational load:

- **RTU**: the rotational training unit under evaluation (battalion/brigade staff, predominantly O4+/E8+).
- **OPFOR**: the resident opposing force.

That contrast approximates a controlled comparison in field data and anchors the analyses that follow.

	RTU	OPFOR
Sleep (h/night)	4.84	5.48
PVT mean RT (ms)	511	460
Steps/day	8,916	7,067

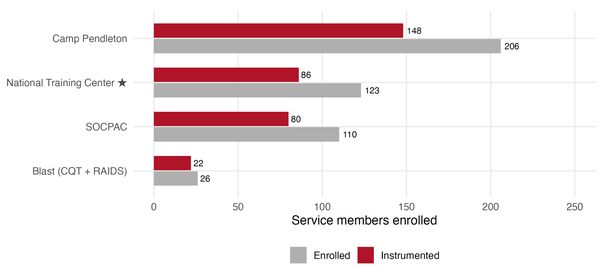
The Exposure



Nightly total sleep across the two cohorts; the dashed line marks the 7 h recommendation. Sleep is the interval union of overlapping wearable stage segments.

Across 937 instrumented nights, soldiers averaged **5.2 h** (median 5.5), and the training unit **4.84 h**. Sleep is computed as an **interval union** of overlapping wearable stage segments; summing them, as is common, inflates total sleep by **49%**.

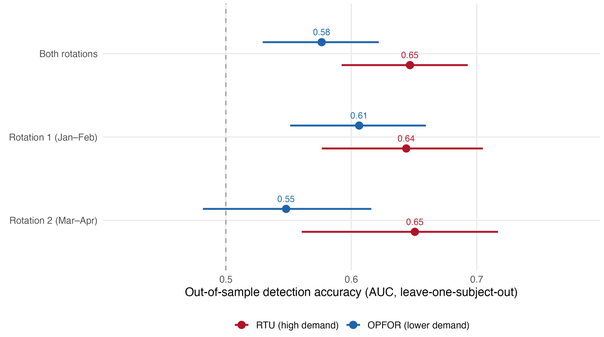
Deployment Scale



Enrollment and wearable instrumentation across the four operational deployments. ★ The analytic sample for this poster is **NTC alone** (93 soldiers); the other three establish that the platform fields and are **not pooled** into any result.

1 · Detection Concentrates Under Load

Detection accuracy is not uniform: it is strongest in the unit under the greatest operational load.



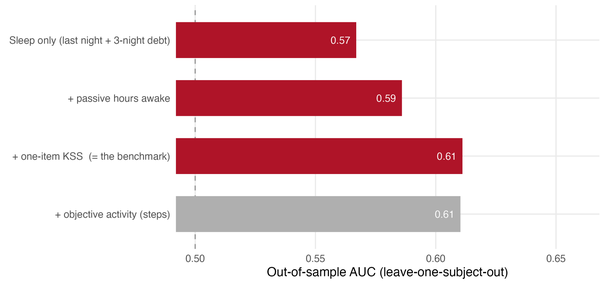
Leave-one-subject-out validation: the model never sees the soldier it scores. Outcome: a day in the slowest third of *that soldier's own* PVT distribution. Bars are 95% **subject-level** bootstrap intervals.

RTU 0.65 vs **OPFOR 0.58**, P(RTU > OPFOR) = 0.98. The RTU leads in **both rotations independently**: two different units, two months apart.

As a screening signal: in the RTU the model flags **39% of a soldier's degraded days** at an 80% true-negative rate, providing useful triage rather than a diagnosis.

Interpretation. The per-rotation cells are small and their confidence intervals overlap. This is a **replicated, directional** result (the capability concentrates under load), *not* a statistically significant cohort difference. It is a hypothesis to be powered prospectively.

2 · Low-Cost Channels Carry the Signal



Out-of-sample AUC as channels are added.

Sleep alone is a ceiling. No sleep representation we or the vendor can build exceeds AUC ≈ 0.59: not raw sleep, not engineered features, not the proprietary model. Passive hours-awake adds little; **one self-report item adds more than any objective channel we captured**. Objective activity adds nothing (0.61 → 0.61).

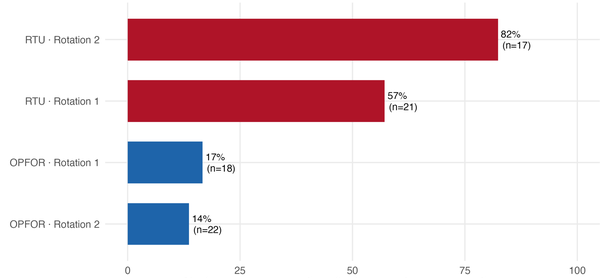
For advanced product development, the central result is that the capability is low-cost *because the expensive channels added no measurable predictive value.*

How Detection Is Scored

ELEMENT	METHOD
Outcome	A day in the slowest third of *that soldier's own* PVT distribution: a **personal**, not a population, threshold.
Inputs	3-night sleep debt · last night's sleep · passive hours awake · one-item KSS, each centered on their own baseline.
Validation	**Leave-one-subject-out.** The model is refit without a soldier before it scores them, so no soldier contributes to their own prediction.
Uncertainty	Bootstrap over **soldiers**, not person-days: days within a soldier are not independent, and resampling rows would yield intervals that are far too tight.

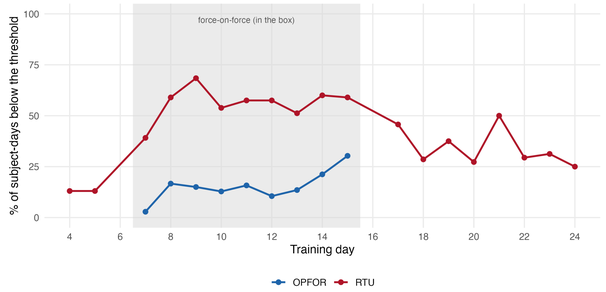
3 · Day 3 in the Box

On the third day of force-on-force operations, 57% (Rotation 1) and 82% (Rotation 2) of the training unit fell below the cognitive-effectiveness threshold the SAFTE/READI model equates to a blood alcohol content > 0.08%, versus 14–17% of the opposing force in the same rotations.



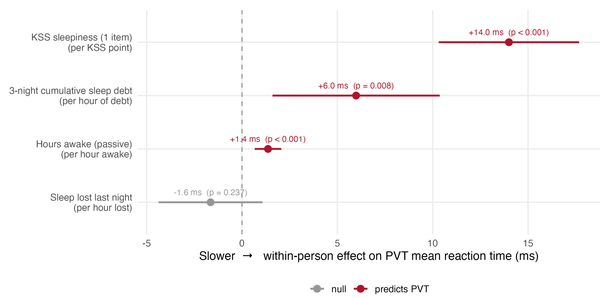
Percent of each unit below the effectiveness threshold on the third day in the box, reported by rotation.

What this measures. A **vendor model output computed from sleep** (Fatigue Science SAFTE/READI modelled effectiveness < 70), *not* a measured reaction time. Reported **by rotation, never pooled**: Rotation 2's bands were not issued until training day 8, and pooling would let a band-issue artifact masquerade as a training effect.



A **dose-response**, not a threshold artifact: the training unit climbs from ~13% in reception/staging to 60–68% in the box and recedes in recovery. The opposing force never leaves the 10–30% band. Whole rotation: **44.5%** of RTU subject-days below threshold vs **14.6%** OPFOR.

4 · Slowing Follows Accumulated Debt

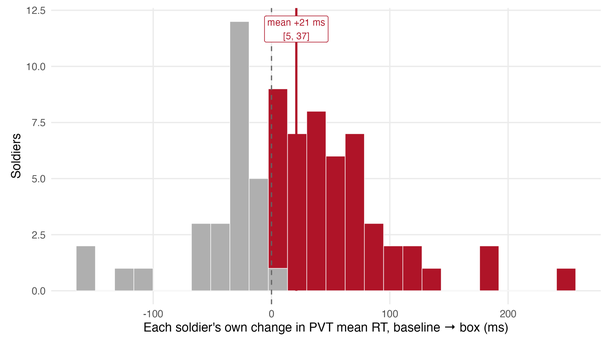


Every predictor is oriented so **more = worse**. Units differ by row; magnitudes are not comparable across rows.

Soldiers slept **5.2 h/night** (median 5.5, IQR 3.9–6.5). A single poor night predicts **nothing** (p = 0.24). **Three nights of accumulated debt do** (+6.0 ms of slowing per hour of debt, p = 0.008), reproduced on the vendor's independent sleep measure (+7.3 ms, p = 0.007).

Degradation here is a **cumulative** process, which is why continuous monitoring is more informative than a single retrospective self-report of sleep.

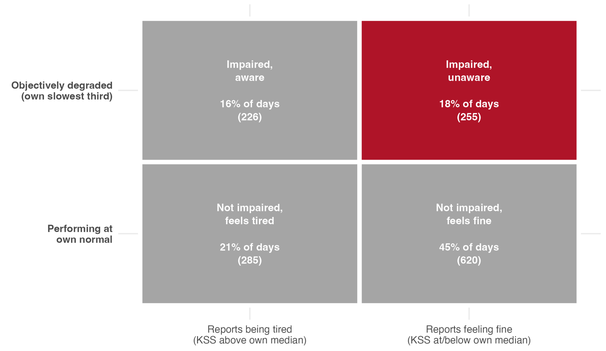
5 · Objective Slowing



Each soldier is their own control; red = slowed, grey = sped up on entering the box. PVT **mean** reaction time.

63% of soldiers slowed on entering the box: **+21 ms** [5, 37], n = 75, p = 0.011, with subjective sleepiness rising in parallel (+0.66 KSS points, p < 0.0001). The spread matters as much as the mean; soldiers differ substantially in the magnitude of the effect.

6 · Impairment Goes Unreported



Objective impairment (own slowest third of PVT) versus subjective report (own KSS median), each against the soldier's own baseline. Tiles show the share of person-days in each quadrant.

53% of objectively degraded person-days went unreported: more than half of impaired soldiers report no impairment when polled.

Impairment and self-report are **both** scored against their own baseline, so this is not a between-person effect.

This is concurrent unawareness, not lead time. The cross-lag runs the other way: yesterday's sleepiness predicts today's slowing (p = 0.03), while yesterday's slowing does *not* predict today's sleepiness (p = 0.87). We do not claim to detect degradation before the soldier experiences it; we detect it **while it is going unreported**, which self-report cannot surface.

What This Capability Enables

- **Role 0–1 monitoring** of degradation risk with equipment already on the wrist, plus one tap of self-report.
- A flag referenced to **the individual**. A fixed population threshold is largely a trait detector: $R^2 = 0.81$ of it is explained by baseline reaction time alone.
- Visibility into the **half of impaired soldiers who report no impairment**.

Limits.
- No HRV was captured in this cohort; this poster makes **no autonomic claim**.
- Detection is **same-day**: there is no day-to-day momentum to forecast on.
- No injury, safety, or AAR outcome exists here, so the endpoint is **degradation risk**, not mission failure.
- Accuracy at AUC ≈ 0.6 is a **screening** signal, not a diagnosis; the ceiling is set by a 3-minute PVT.

Acknowledgements. Supported by CDMRP W81XWH-22-1-0956. We thank the soldiers of the 11th ACR and the rotational units at Fort Irwin.