NEUROFIT

Toward Rapid-Acting, AI-Personalized Somatic Care for Warfighter Readiness: NF8 Stress Trajectories in 10,174 Users

Reported somatic stress falls within 7 days of first app use — and stays down.

MHSRS SESSION
Machine Learning & AI in Military Health, Readiness, & Performance
Single-arm real-world outcomes · US cohort, Jan 2023 – Jul 2026
10,174 users · 100,875 daily check-ins
Python primary · independently re-implemented in R

Andrew Hogue, BS[1] · Lisa Kang, MSPH[2]

[1] NEUROFIT, Miami, Florida

d = 0.74
Observed NF8 change at ~7 days (Mild+, primary)

35%
Reduction in reported somatic stress by day 7

9,831
Users in the primary (Mild+) population

d = 0.05
Regression to the mean (negligible)

d = 0.79
Censoring-weighted estimate (higher, not lower)

THE PROBLEM

In military contexts the intervention window is days, not months.

- **Most digital trials are small** (N<300), short-term, and single-site.
- **Standard instruments recall 1–4 weeks,** so they cannot resolve day-level change.
- **Time-to-effect for CBT, PE and EMDR** is 8–16 weeks.
- **No scalable bridge** exists between screening and definitive care.

APPROACH AND DESIGN

A daily somatic read-out, with exercises matched to it.

NEUROFIT classifies autonomic state algorithmically and personalizes somatic regulation exercises. The **NF8** is an 8-domain daily somatic stress scale (0–16) completed in about one minute.

Retrospective, single-arm, observational. A user is US-based if their first check-in is in the US; all of that user's later check-ins are retained.

Pre-designated primary endpoint: observed standardized NF8 change ≈7 calendar days after a user's first check-in, in Mild+ (NF8 ≥ 1/16, N = 9,831). Moderate+ (NF8 ≥ 5/16, N = 7,038) was secondary. One fixed full-population baseline SD serves as the common denominator, so every population and timepoint is directly comparable.

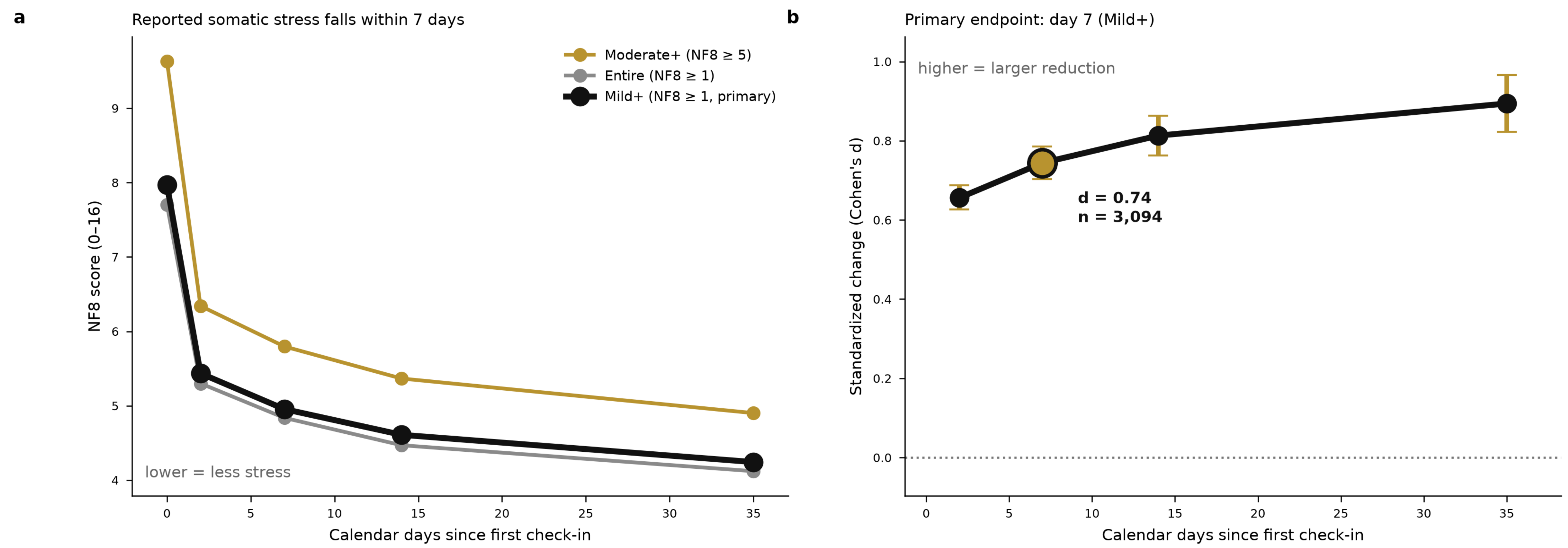

a
Reported somatic stress falls within 7 days
Moderate+ (NF8 ≥ 5)
Entire (NF8 ≥ 1)
Mild+ (NF8 ≥ 1, primary)
NF8 score (0–16)
lower = less stress
Calendar days since first check-in
b
Primary endpoint: day 7 (Mild+)
higher = larger reduction
Standardized change (Cohen's d)
d = 0.74
n = 3,094
Calendar days since first check-in

Figure 1. (a) NF8 marginal means by calendar day; the primary Mild+ series is emphasised. (b) Standardized change with 95% CIs; the gold marker is the pre-designated day-7 endpoint, d = 0.74 [0.70, 0.79] on n = 3,094 retained users.

Day	n (Mild+)	Retention	Reduction	Cohen's d [95% CI]	Mod+ reduction	Mod+ d
2	4,760	48.4%	30.4%	0.66 [0.63, 0.69]	33.5%	0.83
7	**3,094**	**31.5%**	**35.2%**	**0.74 [0.70, 0.79]**	**38.3%**	**0.94**
14	2,226	22.6%	39.0%	0.81 [0.76, 0.86]	43.2%	1.06
35	1,167	11.9%	43.3%	0.89 [0.82, 0.97]	47.3%	1.15

Bold row = pre-designated primary endpoint. The analysis is observed-case, so later rows describe progressively smaller, more engaged subsets. **Responders (≥30% reduction) at day 7: 57.8% of users retained near day 7, or 18.2% of all Mild+ users when every non-retained user counts as a non-responder.**

WHY IT MATTERS OPERATIONALLY

A one-minute daily read-out, on a phone already in the pocket.

- **No added hardware:** the check-in and the optional camera-PPG reading both run on a standard smartphone.
- **Day-level resolution:** detects change inside a training cycle or return-to-duty timeline, where a 1–4 week recall instrument cannot.
- **Self-directed and low-stigma:** no appointment, no clinician in the loop, no record of care-seeking.
- **Scales without staffing:** 10,174 users were observed here with no per-user clinical time.

COULD IT BE AN ARTIFACT?

Four independent attacks on the retention explanation.

- **Regression to the mean:** negligible (d = 0.05), bounded with the observed 24-hour test–retest reliability.
- **Censoring weights:** IPCW *raised* the estimate to d = 0.79.
- **Zero-change assumption:** counting every non-retained user as unchanged still gives d = 0.23.
- **Natural experiment:** users gifted free access were retained 2.6× as often yet changed similarly (d = 0.60 vs 0.76), separating retention from effect size.

IS THE INSTRUMENT SOUND?

Convergent with standard scales; physiology moves as expected.

In independent randomized-administration samples (N = 120–400) the NF8 converged with the **PHQ-9 (r = 0.70)**, GAD-7 (r = 0.55) and DASS-21 Stress (r = 0.52); internal consistency α = 0.81 [0.78, 0.84].

Pre-check-in smartphone-camera PPG tracked reported stress in both cardiac markers — heart rate r = 0.14 [0.07, 0.21], HRV-RMSSD r = -0.10 [-0.17, -0.03] — the expected joint autonomic direction. Breathing rate was null. Correlations are user-balanced; all are small and exploratory.

WHAT THIS DOES AND DOES NOT SHOW

Single-arm observational evidence of outcomes during use — not a controlled efficacy estimate.

Rapid, sustained NF8 reduction occurred within 7 days at population scale (d = 0.74 at day 7, rising to 0.89 by day 35), and the result survived four independent attacks on the retention explanation.

There is no control arm, no randomization and no blinding. Expectancy, natural remission and unmeasured co-intervention remain viable contributors; causal attribution requires a randomized trial. The outcome is a proprietary self-report measure. Retention near day 7 was 32%, typical for consumer wellness apps but a real limit on population-level yield. **No participant was identified as military** — relevance to service members is the hypothesis motivating a planned military validation study, not a finding here.

US-based cohort only. Effect sizes use one fixed full-population baseline SD. Python primary with independent R re-implementation; agreement verified leaf-by-leaf. Companion preprints in preparation (Hogue & Kang, 2026).

MHSRS 2026