THROUGH THE RINGER: STRESS TESTING OURA RING 4 SLEEP-TRACKING PERFORMANCE IN MULTIPLE OPERATIONAL U.S. NAVY CONTEXTS

Remington Mallett[1,2], Andrew G. Kubala[1,2], Trevor B. Viboch[1,2], Emily M. Moslener[1,2], Joseph A. Cuellar[1,2], Krisztina Marosi[1,2], Sarah A. Chabal[1,2], Rachel R. Markwald[1], Evan D. Chinoy[1]

[1]Warfighter Performance Department, Naval Health Research Center, San Diego, CA; [2]Leidos, Inc., San Diego, CA

NAVY MEDICINE RESEARCH & DEVELOPMENT

BACKGROUND

- Consumer wearable health technology continues to rapidly advance and is now being adopted in operational contexts by individual service members and as part of comprehensive programs to monitor readiness.
  - Capabilities that aggregate sleep data at the group level may inform processes and decisions to mitigate fatigue.
  - Sleep data collected at the individual level may be used to help make manning decisions.
- Nearly all wearable validation studies have been conducted in general populations and controlled settings, limiting their applicability in operational contexts that likely include irregular sleep patterns, disruptive sleep environments, etc.
- Oura Ring is a widely used device in current operational efforts tracking service member sleep.

- <u>**Objective**</u>**: The current study evaluated the sleep-tracking performance of a recent Oura Ring model against portable electroencephalography (EEG) in multiple U.S. Navy operational contexts.**

METHODS

Reference device – Somfit

Somfit portable EEG (*Compumedics Ltd.*)
A "silver-standard" EEG device

Photo courtesy of Naval Health Research Center

- A single-use forehead adhesive-gel electrode patch
- 78 Hz sampling frequency
- Measures neural activity with EEG electrodes
- 3 electrode locations measured: Fp1, Fp2, Fpz
- 4 differential channels: EEG (Fp1-Fp2), EMG (Fp1-Fp2), EOG (Fp1/2-Fpz)
- Utilizes previously validated automated sleep scoring algorithm

Test device – Oura

Oura Ring 4 (*Oura Health Oy*)
A commercial multi-sensor wearable device

U.S. Navy photo by Ramin Khalili from DVIDS/Released

- PPG sensors: red/infrared LEDs measure blood oxygen
- PPG sensors: green/infrared LEDs measure heart rate
- Accelerometer and temperature sensors
- Smart sensing (accounts for skin tone, BMI, age, etc.)
- Sleep staging was performed with Oura's current algorithm (v2.0)
- Data extracted from Oura Teams commercial interface

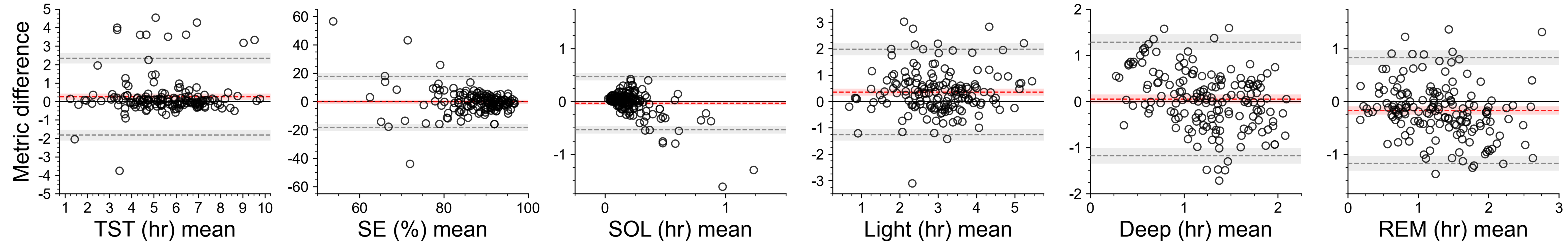
Samples

- Participants were active duty U.S. Navy Sailors from multiple operational studies and wore both devices during either:
  - Shipboard sleep from one command (n = 49, up to 9 sleep episodes per participant)
  - Shore-based sleep from two commands working in port on rotating shifts (n = 9, up to 15 sleep episodes per participant)

RESULTS

Discrepancy analysis – Comparison of sleep episode summary metrics

Biases from the Oura Ring 4 show means and spreads typical of previous wearable evaluation studies ($N_{participants}$ = 47; $N_{episodes}$ = 60)



Metric difference = Oura scores minus Somfit scores; Positive values indicate Oura overestimation, negative values indicate Oura underestimation

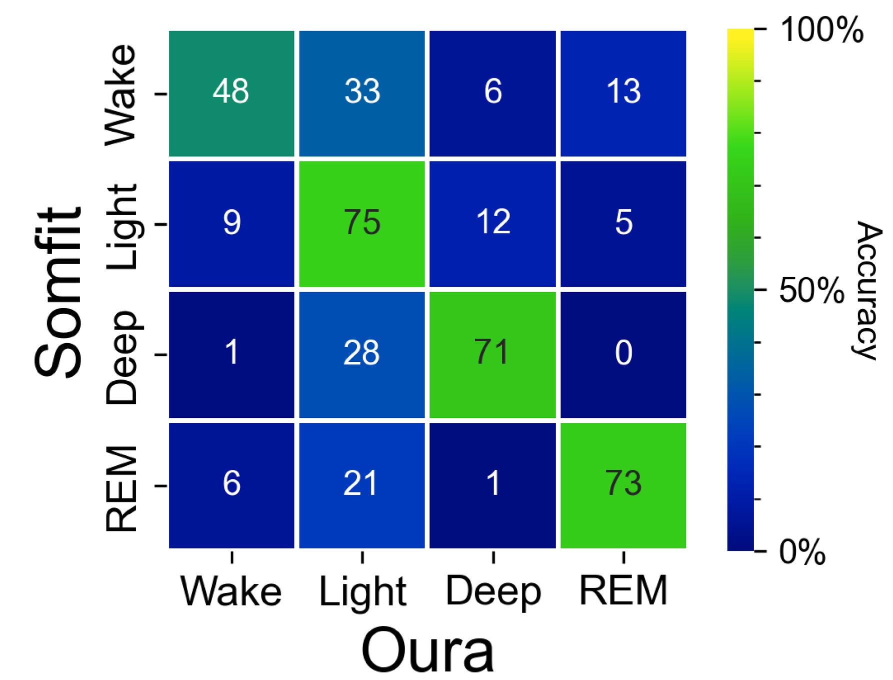
TST = total sleep time; SE = sleep efficiency; SOL = sleep onset latency; Light = light sleep; Deep = deep sleep; REM = rapid eye movement sleep

Red dashed line = mean bias; gray dashed lines = limits of agreement; shaded regions = 95% confidence intervals

Epoch-by-epoch analysis – Comparison of sleep stages per 30 seconds

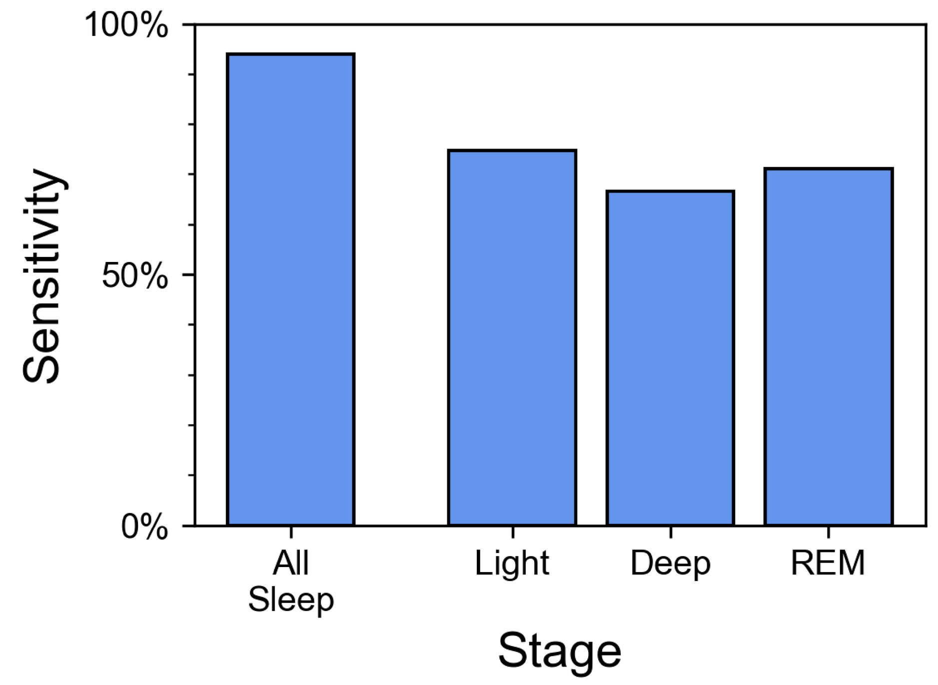
Staging from the Oura Ring 4 shows agreement representative of previous wearable evaluation studies ($N_{participants}$ = 60; $N_{episodes}$ = 301).

Classification agreement

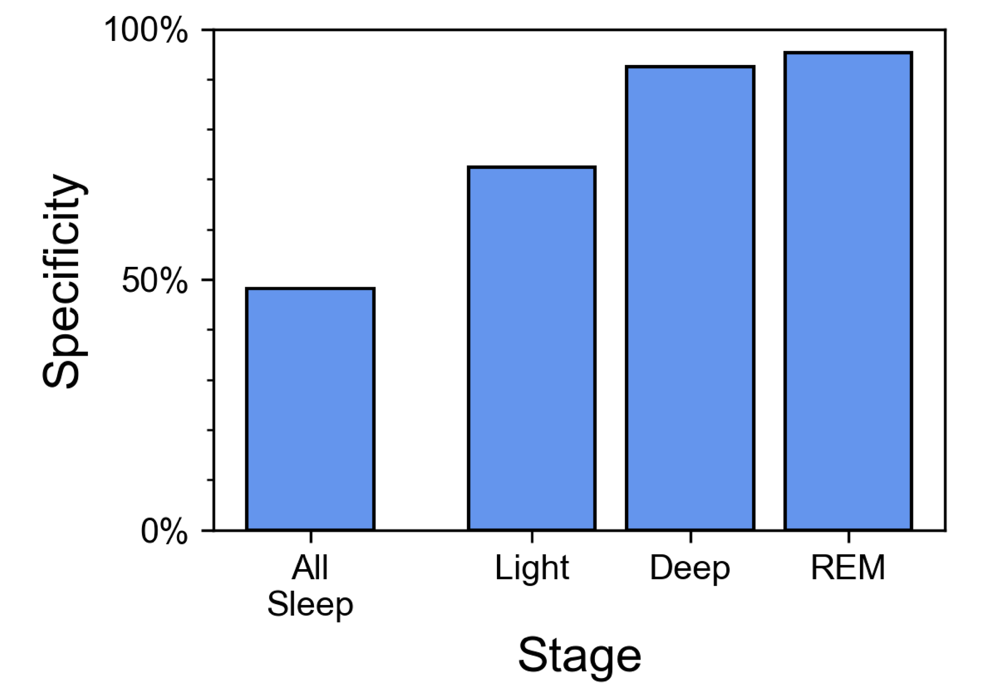


Values represent the percentage of each Oura state classification for a given Somfit classification

Agreement metrics



Sensitivity = True positive rate; % correctly identified epochs where a given sleep stage was present



Specificity = True negative rate; % correctly identified epochs where a given sleep stage was absent

CONCLUSIONS

- **Overall, the Oura Ring 4 performed as well in military operational contexts as prior Oura models did in laboratory and general population studies.**
- Sleep summary outcomes showed relatively small mean biases, which are typically deemed acceptable for wearables.
- The epoch-level sensitivity and specificity trade-offs follow similar patterns to those in prior studies and offer valuable insights into what conclusions can be drawn from wearables in large datasets (e.g., more sensitive to sleep epochs than wake epochs, mixed performance for sleep stage detection).
- Though the current study addressed shipboard sleep and shiftwork, future studies in military operational contexts to evaluate wearables' native sleep algorithms will be critical to understanding their strengths and limitations in fatigue risk management efforts.
- The Oura Ring 4 (release date OCT 2024) was recently surpassed by the Oura Ring 5 (JUN 2026). If used in similar contexts, the Oura Ring 5 would benefit from a similar evaluation.

This work was supported by the Office of Naval Research (Code 34), Naval Medical Forces Pacific, and Naval Advanced Medical Development under work unit nos. N1701 and 62513. The views expressed in this work are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. The study protocol was approved by the Naval Health Research Center Institutional Review Board in compliance with all applicable federal regulations governing the protection of human subjects. Research data were derived from approved Naval Health Research Center Institutional Review Board protocol numbers NHRC.2017.0008 and NHRC.2025.0007.