

## Supplementary Materials

# **Performance of four commercial wearable sleep-tracking devices tested under unrestricted conditions at home in healthy young adults**

Evan D. Chinoy<sup>1,2</sup>

Joseph A. Cuellar<sup>1,2</sup>

Jason T. Jameson<sup>1,2</sup>

Rachel R. Markwald<sup>1</sup>

<sup>1</sup>Sleep, Tactical Efficiency, and Endurance Laboratory, Warfighter Performance Department, Naval Health Research Center, San Diego, CA, USA; <sup>2</sup>Leidos, Inc., San Diego, CA, USA

Correspondence: Rachel R. Markwald; Sleep, Tactical Efficiency, and Endurance Laboratory, Warfighter Performance Department, Naval Health Research Center; 140 Sylvester Road, San Diego, CA 92106  
Tel: +1 619 767 4494  
Email: [rachel.r.markwald.civ@mail.mil](mailto:rachel.r.markwald.civ@mail.mil)

## Device Firmware and App Software Versions

Device firmware and software versions during testing were kept as consistent as possible between participants. App software versions for the Dreem headband and all commercial wearable devices were kept consistent throughout the entire data collection period (July – November 2020) for all participants. However, device firmware versions for only the Polar Vantage V Titan and Oura Ring devices had to be updated over the testing period, but we confirmed with representatives from both companies (i.e., Polar and Oura) that these firmware updates between participants did not affect the sleep-tracking data or algorithms of those two devices. The following were the app software versions and device firmware versions, respectively, for each device over the entire testing period: Dreem: 2.13.3, 4.0.92; Fatigue Science Readiband: 3.2.4, 2019-12-13; Fitbit Inspire HR: 3.24.1, 20001.63.18; Oura Ring: 3.1.0 (also updated to 3.2.0, 3.3.0, and 3.4.0), 2.26.1 (also updated to 2.36.1); Polar Vantage V Titan: 4.4.8, 5.1.4 (also updated to 5.1.8).

## Actiwatch Sensitivity Threshold Comparisons

Sleep-wake summary and EBE results for the Actiwatch analyzed with the low and high sensitivity thresholds are presented in Supplementary Tables S4-S6 and depicted with Bland-Altman plots in Supplementary Figure S1.

In general, the different sensitivity threshold settings exhibited mixed results for the sleep-wake detection capability of the Actiwatch compared with the recommended medium setting and the commercial devices. The low and high settings had opposite effects on the EBE results, with the medium setting results remaining in between low and high. The low setting decreased sensitivity and increased specificity, indicating that detection of Dreem-scored wake improved and detection of Dreem-scored sleep worsened. Sensitivity and specificity with the low setting were to varying extents lower and higher, respectively, than the commercial devices. Despite specificity increasing, NPV decreased with the low setting – indicating that while more Actiwatch epochs were scored as wake and more often corresponded to the Dreem's wake epochs, a lower rate of the Actiwatch's low setting wake epochs were actually wake epochs according to Dreem. The high setting had the opposite effects on EBE results as described with

the low setting, indicating similar (but opposite) performance tradeoffs with the other Actiwatch settings and the commercial devices.

The low and high settings also altered the sleep-wake summary outcomes in ways that corresponded to the EBE results – the low setting generally increased the amount of wake and decreased the amount of sleep across sleep episodes, and the high setting did the opposite. The mean levels of TST, SE, and WASO with both the low and high settings were more biased versus the Dreem than the medium setting. Overall, the recommended medium Actiwatch setting most closely estimated mean bias for sleep-wake summary outcomes compared with the Dreem. These results suggests that, depending on the population and setting, the different actigraphy analysis settings may be warranted in some studies, but that the decision may result in performance tradeoffs.

## Supplementary Tables

Supplementary Tables S1-S6 are included in a separate Excel file and contain:

**Table S1:** EBE Contingency Tables

**Table S2:** Sleep-Wake Summary Agreement (Expanded Results)

**Table S3:** Sleep Stage Summary Agreement (Expanded Results)

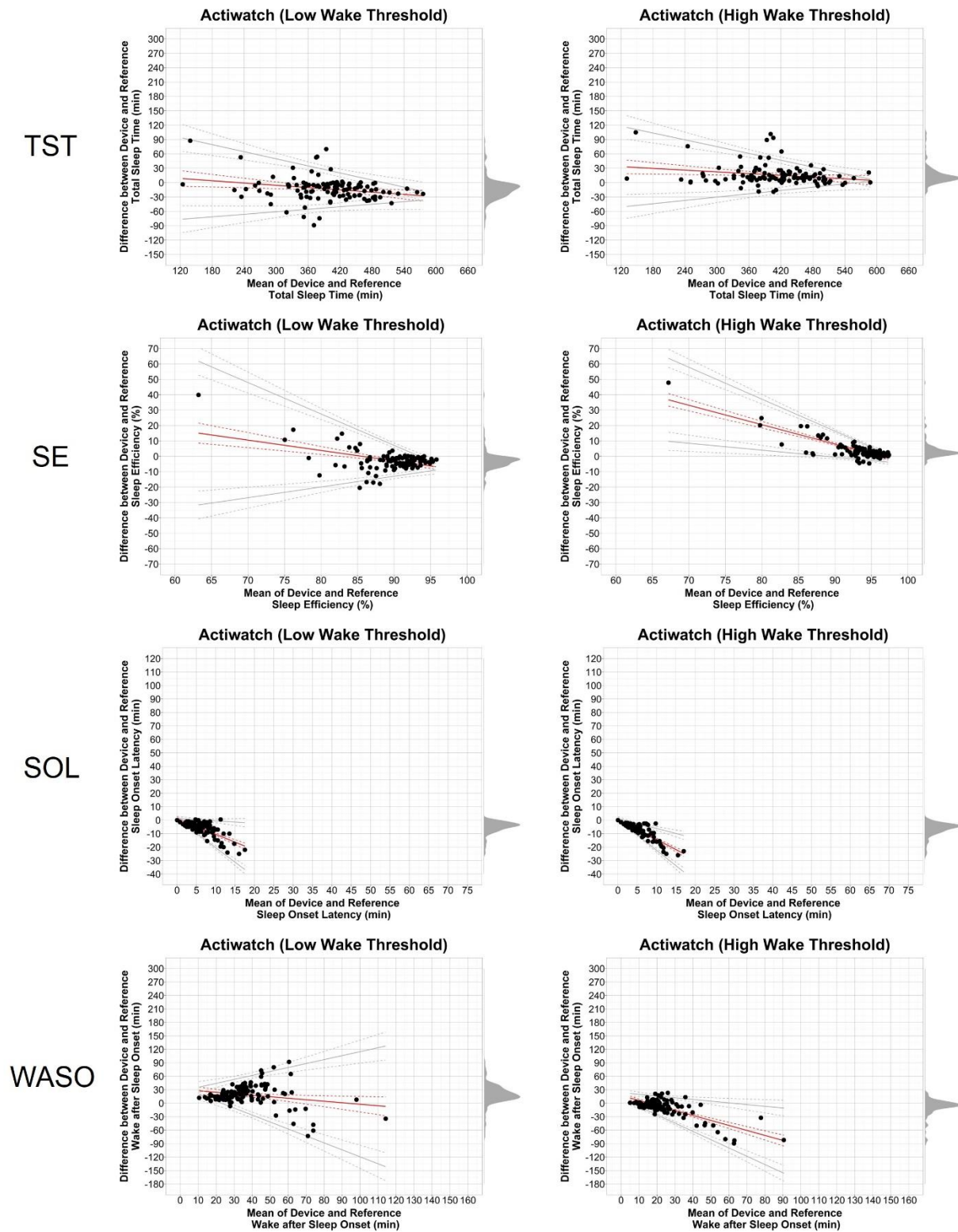
**Table S4:** Actiwatch Sleep-Wake Summary Agreement with Low and High Threshold Settings

**Table S5:** Actiwatch EBE Sleep-Wake Contingency Tables with Low and High Threshold Settings

**Table S6:** Actiwatch EBE Sleep-Wake Agreement with Low and High Threshold Settings

# Supplementary Figure

Supplementary Figure S1: Bland-Altman Plots: Actiwatch Low and High Sensitivity Threshold Settings



**Notes:** Plots depict the mean bias (solid red line) and upper and lower limits of agreement (solid gray lines) for deviation in TST, SE, SOL, and WASO for the Actiwatch with low and high sensitivity threshold settings compared with the reference Dreem. Black circles are individual nights. Dashed lines represent the 95% confidence intervals around the bias and limits of agreement lines. Gray shaded regions on the right y-axis are density plots showing the distribution of individual night biases. Zero on the y-axis represents no difference, with positive and negative y-axis values indicating an overestimation or underestimation, respectively, compared with the reference. Diagonal mean bias lines indicate significant proportional bias. Non-parallel limits of agreement lines indicate significant heteroscedasticity.

**Abbreviations:** TST, total sleep time; SE, sleep efficiency; SOL, sleep onset latency; WASO, wake after sleep onset.