Supplementary Materials

Daytime sleep-tracking performance of four commercial wearable devices during unrestricted home sleep

Evan D. Chinoy[1,2]

Joseph A. Cuellar[1,2]

Jason T. Jameson[1,2]

Rachel R. Markwald[1]

[1]Sleep, Tactical Efficiency, and Endurance Laboratory, Warfighter Performance Department, Naval Health Research Center, San Diego, CA, USA

[2]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, USA

Tel: +1 619 767 4494

Email: rachel.r.markwald.civ@health.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 all commercial wearable devices were kept consistent throughout the entire data collection period (May – November 2021) for all participants. The app and firmware versions for the Oura Ring had to be updated partway through the study, but the updates were not reported to affect the sleep-tracking data. The following were the app software versions and device firmware versions, respectively, for each device over the testing period: Fatigue Science ReadiBand: 3.5.2, 2021-05-05; Fitbit Inspire HR: 3.42, 20001.88.11; Oura Ring: 3.9.2 (updated to 3.9.3, 3.10.0, 3.11.0, 3.11.1, 4.0.0), 3.96 (updated to 4.0.2); Polar Vantage V Titan: 5.3.0, 5.1.8.

Supplementary Table

**Supplementary Table S1:** Bland-Altman Summary Agreement (Expanded Results)

Outcome	Device	n	Device Mean (SD)	Reference Mean (SD)	CI Type	CI Level	Propor-tional Bias	Hetero-scedasticity	Norm-ality	Log Transform	Bias	Bias CI	LOA Lower	LOA Lower CI	LOA Upper	LOA Upper CI
TIB	Fatigue Science ReadiBand	80	328.14 (123.68)	343.98 (125.76)	classic	0.95	FALSE	FALSE	FALSE	FALSE	-15.84 (28.81)	(-22.25, -9.43)	-72.31	(-83.42, -61.21)	40.64	(29.53, 51.74)
TIB	Fitbit Inspire HR	79	351.42 (113.77)	353.37 (115.59)	classic	0.95	FALSE	FALSE	FALSE	FALSE	-1.95 (45)	(-12.03, 8.13)	-90.14	(-107.6, -72.69)	86.24	(68.79, 103.7)
TIB	Oura Ring	78	362.0 (108.8)	350.6 (115.6)	classic	0.95	FALSE	TRUE	FALSE	FALSE	11.45 (32.56)	(4.11, 18.79)	bias - 2.46 * 48.53 - 0.07 * ref	c0 = (32.22, 64.84); c1 = (-0.12, -0.03)	bias - 2.46 * 48.53 - 0.07 * ref	c0 = (32.22, 64.84); c1 = (-0.12, -0.03)
TIB	Polar Vantage V Titan	50	394.04 (81.23)	405.86 (73.75)	classic	0.95	TRUE	FALSE	FALSE	FALSE	-51.53 + 0.1 * ref	b0 = (-91.58, -11.47); b1 = (0, 0.2)	bias - 51	bias ci - 51	bias + 51	bias ci + 51
Start Time	Fatigue Science ReadiBand	80	641.81 (184.91)	633.98 (186.65)	classic	0.95	FALSE	FALSE	FALSE	FALSE	7.84 (19.98)	(3.39, 12.28)	-31.32	(-39.02, -23.62)	46.99	(39.29, 54.69)
Start Time	Fitbit Inspire HR	79	625.15 (173.69)	620.04 (173.91)	classic	0.95	FALSE	FALSE	FALSE	FALSE	5.13 (27.61)	(-1.05, 11.32)	-48.98	(-59.7, -38.27)	59.25	(48.54, 69.96)
Start Time	Oura Ring	78	626.9 (171.9)	625.8 (177.1)	classic	0.95	TRUE	FALSE	FALSE	FALSE	19.84 - 0.03 * ref	b0 = (4.33, 35.34); b1 = (-0.05, -0.01)	bias - 36	bias ci - 36	bias + 36	bias ci + 36
Start Time	Polar Vantage V Titan	50	601.76 (152.99)	590.72 (147.34)	classic	0.95	TRUE	FALSE	FALSE	FALSE	-11.43 + 0.04 * ref	b0 = (-30.66, 7.8); b1 = (0.01, 0.07)	bias - 32	bias ci - 32	bias + 32	bias ci + 32
End Time	Fatigue Science ReadiBand	80	969.95 (156.35)	977.95 (151.43)	classic	0.95	TRUE	FALSE	FALSE	FALSE	-39.25 + 0.03 * ref	b0 = (-67.83, -10.67); b1 = (0, 0.06)	bias - 39	bias ci - 39	bias + 39	bias ci + 39
End Time	Fitbit Inspire HR	79	976.56 (152.90)	973.41 (151.68)	classic	0.95	FALSE	FALSE	FALSE	FALSE	3.18 (35.87)	(-4.85, 11.22)	-67.11	(-81.03, -53.2)	73.48	(59.57, 87.39)
End Time	Oura Ring	78	989.0 (152.9)	976.4 (153.4)	classic	0.95	FALSE	FALSE	FALSE	FALSE	12.62 (26.5)	(6.64, 18.59)	-39.33	(-49.68, -28.98)	64.56	(54.21, 74.91)
End Time	Polar Vantage V Titan	50	995.80 (143.95)	996.58 (142.49)	classic	0.95	FALSE	FALSE	FALSE	FALSE	-0.78 (22.05)	(-7.05, 5.49)	-43.99	(-54.85, -33.14)	42.43	(31.58, 53.29)

Notes: Summary agreement statistics between devices and reference for the Bland-Altman plots for TIB, sleep episode start time, and sleep episode end time. The sleep log served as reference for all outcome measures. All daytime sleep episodes with available data for both the device and reference were included in analyses, with the total number of nights (n) indicated in each row. True/false outputs for proportional bias, normality, and heteroscedasticity columns indicate statistical significance at p<0.05. Bias is the mean difference between the device and reference. Regression equations are returned for bias and LOAs when proportional bias or heteroscedasticity are significant, respectively. Values for sleep start and end times are minutes since midnight.

Abbreviations: CI, confidence interval; LOA, limits of agreement; SD, standard deviation; TIB, time in bed.

Supplementary Figures

Supplementary Figure S1: Individual Participant Sleep Schedules
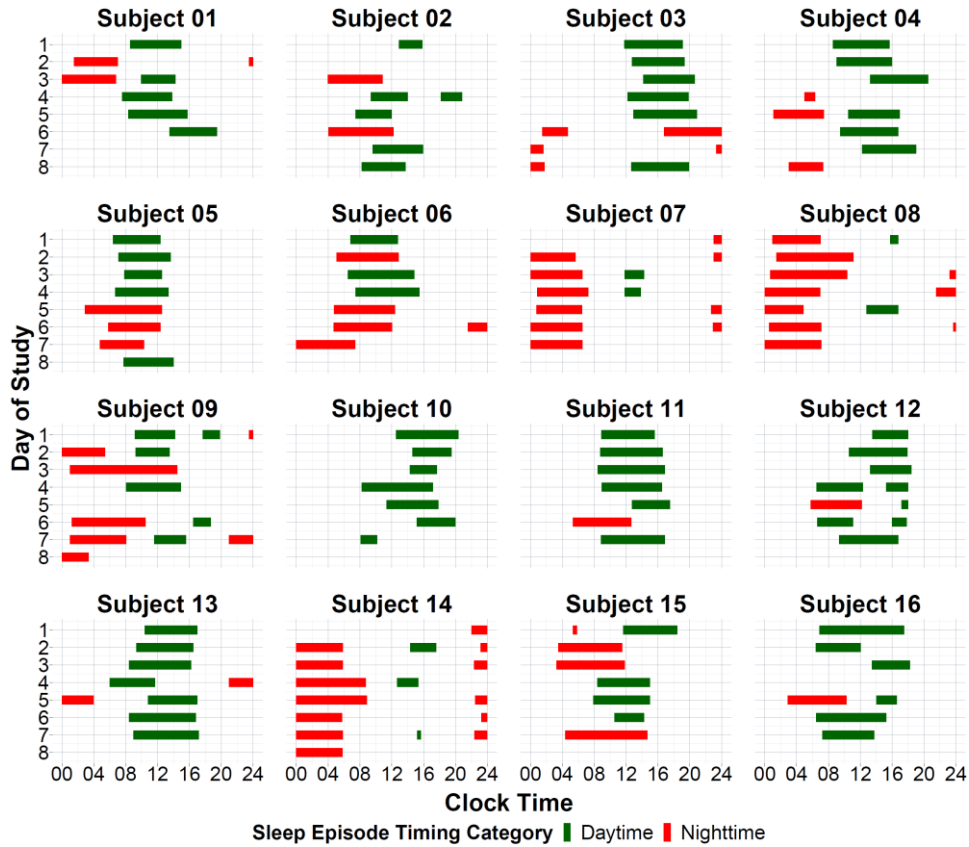
Notes: Plots depict the TIBs and classifications of all sleep episodes for each participant across their study week (7 nights / 8 days), as reported in their reference sleep logs. Sequential days of study run down the y-axis and clock time is shown on the x-axis (note that nighttime sleep episodes with TIB that cross midnight wrap across study day rows). Classification for sleep episodes that met criteria for "daytime sleep" were included in the data analyses and are depicted with green bars. Sleep episodes that did not meet the criteria for daytime sleep were classified as "nighttime sleep," and were not included in the data analyses for this study (and are depicted with red bars).

Abbreviations: TIB, time in bed.

Supplementary Figure S2: Aggregated Sleep Schedules Across All Participants
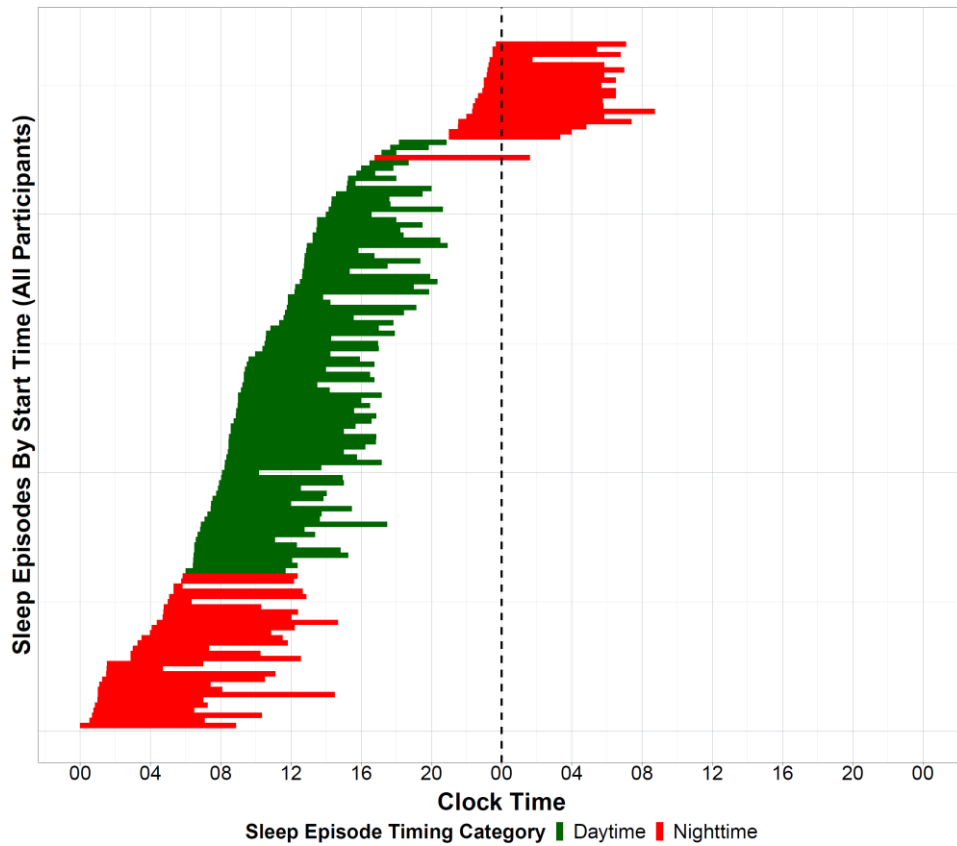
Notes: This figure depicts the TIBs and classifications aggregated across all sleep episodes and participants, as reported in their reference sleep logs. Rows on the y-axis represent each sleep episode, ordered from bottom to top by start of TIB (starting with midnight). Clock time is shown on the x-axis. Like Supplementary Figure S1, classification for sleep episodes that met criteria for "daytime sleep" were included in the data analyses and are depicted with green bars. Sleep episodes that did not meet the criteria for daytime sleep were classified as "nighttime sleep," and were not included in the data analyses for this study (and are depicted with red bars).

**Abbreviations:** TIB, time in bed.

Supplementary Figure S3: Examples of Device TIB Differences Within Participants
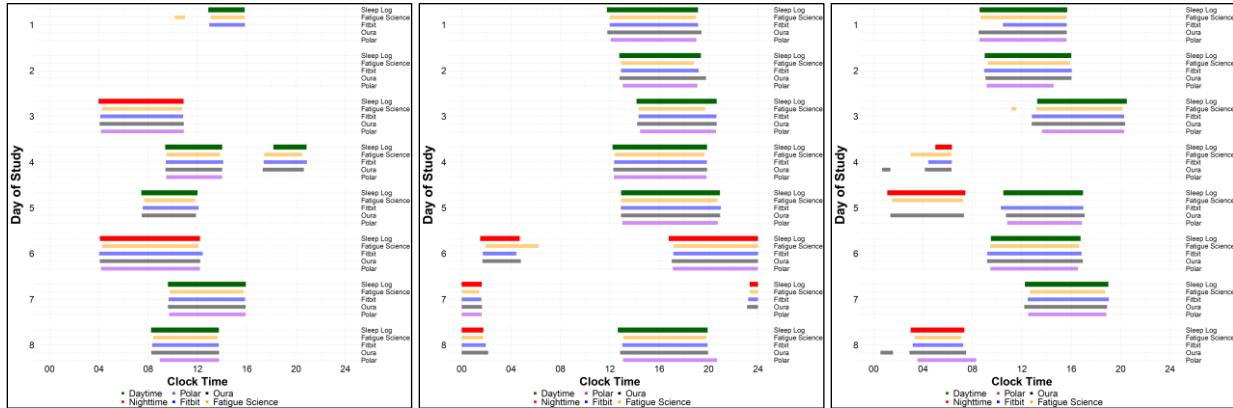
Notes: TIB comparisons between the sleep log and all devices across the study week for three example participants. Like Supplementary Figure S1, plots depict the TIBs and classifications of all sleep episodes for each participant across their study week (7 nights / 8 days), as reported in their reference sleep logs. Sequential days of study run down the y-axis and clock time is shown on the x-axis (note that nighttime sleep episodes with TIB that cross midnight wrap across study day rows). Sleep log classifications of sleep episodes that met criteria for "daytime sleep" were included in the data analyses and are depicted with green bars. Sleep episodes that did not meet the criteria for daytime sleep were classified as "nighttime sleep," and were not included in the data analyses for this study (and are depicted with red bars). Color-coded bars depicting each commercial device's TIB are shown on rows below the sleep log TIB for each day. Devices occasionally missed detecting a logged sleep episode, and therefore have no TIB data depicted for that episode where other devices may have TIB data depicted.

Abbreviations: TIB, time in bed.

Supplementary Figure S4: All Device Daytime Sleep Episode Recordings
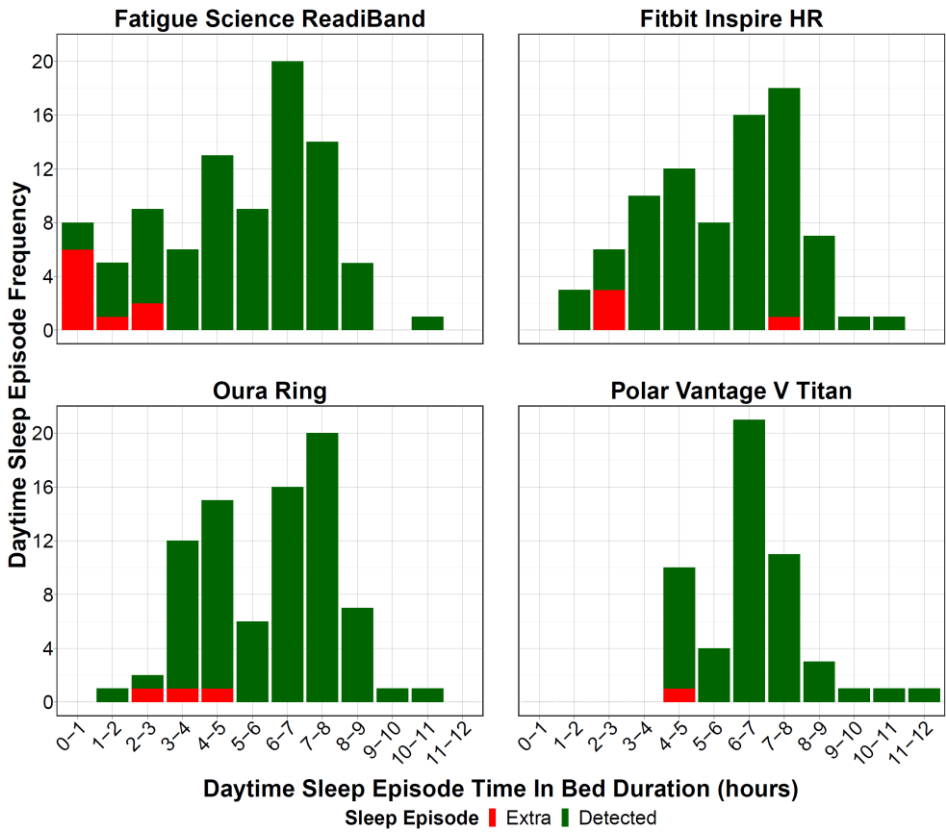
Notes: Plots depict all device-determined daytime sleep episode recordings, organized as stacked histogram bars indicating the number of correctly detected and extra daytime recordings for the devices compared with those reported in the reference sleep log. Green bars represent correctly detected episodes and red bars represent extra false-positive episodes recorded by the device (but were not reported in the sleep log), for each 1-hour TIB bin. TIB refers to the minutes of TIB as recorded by the device for each daytime sleep episode, which may not necessarily equal the TIB of the corresponding daytime sleep episode reported in the sleep log. Only the device recordings that met criteria from the device recording's TIB alone are depicted. Therefore, the total number of correctly detected device recordings depicted in this figure may differ slightly from those depicted in Figure 4 in the main text (which included sleep episodes that met criteria for daytime sleep as determined from the sleep log's TIB).

Abbreviations: TIB, time in bed.