ORIGINAL RESEARCH

Stone et al

# Evaluations of Commercial Sleep Technologies for Objective Monitoring During Routine Sleeping Conditions

**Supplemental Materials**

**Table S1.** Outlier Prevalence by Participant

|  |  |  |  |
| --- | --- | --- | --- |
| **Subject** | **Outlier Count** | **Total Observations** | **Total Nights Slept** |
| 1 | 13 | 315 | 21 |
| 2 | 19 | 295 | 18 |
| 3 | 4 | 280 | 18 |
| 4 | 5 | 219 | 11 |
| 5 | 6 | 166 | 10 |

**Notes.** Table S1 includes the breakdown of the number of outliers identified with respect to the total observations and nights slept. This breakdown is based on the individual data sets derived from each of the participants.

**Table S2.** Outlier Prevalence by Device

|  |  |  |  |
| --- | --- | --- | --- |
| **Device** | **Outlier Count** | **Total Observations** | **Percent Outlier** |
| Beddit | 2 | 120 | 1.67 |
| Fatigue Science | 6 | 210 | 2.86 |
| Fitbit | 8 | 169 | 4.73 |
| Garmin | 7 | 121 | 5.79 |
| Oura | 17 | 373 | 4.56 |
| Polar | 0 | 57 | 0.00 |
| Sleep++ | 0 | 16 | 0.00 |
| SleepWatch | 0 | 13 | 0.00 |
| WHOOP | 7 | 196 | 3.57 |

**Notes.** Table S2 details the proportion of outliers that were derived from each device, relative to the total observations for the device.

**Table S3.** Individual Subject Differences for TST

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Subject** | **n** | **MAPE** | **Min. (%)** | **MdAPE** | **Max. (%)** | **IQR** |
| 1 | 58 | 6.33 | 0.19 | 5.12 | 26.06 | 5.08 |
| 2 | 70 | 9.77 | 0.18 | 5.41 | 50.07 | 13.47 |
| 3 | 78 | 19.27 | 0.15 | 15.57 | 85.69 | 14.07 |
| 4 | 48 | 8.70 | 0.22 | 8.58 | 29.05 | 7.48 |
| 5 | 36 | 8.96 | 0.00 | 4.54 | 44.35 | 11.64 |

**Notes.** Table S3 includes summary statistics for absolute percent error when measuring total sleep time, with respect to each individual participant.

**Abbreviations:** APE, absolute percent error; IQR, interquartile range; MAPE, mean absolute percent error; Max, maximum; MdAPE, median absolute percent error; Min, minimum; n, number of trials; TST, total sleep time

**Table S4.** Individual Subject Differences for TWT

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Subject** | **n** | **MAPE** | **Min. (%)** | **MdAPE** | **Max. (%)** | **IQR** |
| 1 | 61 | 39.44 | 1.37 | 36.95 | 96.30 | 33.31 |
| 2 | 62 | 58.76 | 1.01 | 56.73 | 135.29 | 41.55 |
| 3 | 66 | 54.78 | 0.00 | 62.28 | 127.14 | 53.49 |
| 4 | 49 | 44.06 | 2.13 | 42.55 | 125.35 | 49.33 |
| 5 | 32 | 46.97 | 0.00 | 38.80 | 140.58 | 65.87 |

**Notes.** Table S4 includes summary statistics for absolute percent error when measuring total wake time, with respect to each individual participant.

**Abbreviations:** APE, absolute percent error; IQR, interquartile range; MAPE, mean absolute percent error; Max, maximum; MdAPE, median absolute percent error; Min, minimum; n, number of trials; TWT, total wake time

**Table S5.** Individual Subject Differences for SE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Subject** | **n** | **MAPE** | **Min. (%)** | **MdAPE** | **Max. (%)** | **IQR** |
| 1 | 59 | 6.04 | 0.01 | 5.54 | 13.31 | 5.38 |
| 2 | 61 | 8.50 | 0.16 | 5.31 | 35.17 | 10.87 |
| 3 | 66 | 9.83 | 0.13 | 10.09 | 23.05 | 9.06 |
| 4 | 49 | 6.22 | 0.04 | 5.37 | 14.66 | 8.30 |
| 5 | 33 | 5.65 | 0.66 | 4.84 | 19.21 | 6.15 |

**Notes.** Table S5 includes summary statistics for absolute percent error when measuring total wake time, with respect to each individual participant.

**Abbreviations:** APE, absolute percent error; IQR, interquartile range; MAPE, mean absolute percent error; Max, maximum; MdAPE, median absolute percent error; Min, minimum; n, number of trials; SE, sleep efficiency

**Table S6.** Steel-Dwass Summary Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Subject** | **Comparison Subject** | **Steel-Dwass**  **Z-Score** | ***p*-Value** |
| TST | 2 | 1 | 0.67 | 0.962 |
| 3 | 1 | 6.76 | < 0.001\* |
| 3 | 2 | 4.94 | < 0.001\* |
| 4 | 1 | 2.16 | 0.195 |
| 4 | 2 | 0.63 | 0.971 |
| 4 | 3 | -4.93 | < 0.001\* |
| 5 | 1 | 0.01 | 1 |
| 5 | 2 | -0.70 | 0.956 |
| 5 | 3 | -4.54 | < 0.001\* |
| 5 | 4 | -0.80 | 0.931 |
| TWT | 2 | 1 | 3.68 | 0.002\* |
| 3 | 1 | 2.84 | 0.036\* |
| 3 | 2 | -0.68 | 0.960 |
| 4 | 1 | 0.66 | 0.966 |
| 4 | 3 | -1.93 | 0.299 |
| 4 | 2 | -2.67 | 0.058 |
| 5 | 1 | 0.25 | 0.999 |
| 5 | 4 | 0.00 | 1 |
| 5 | 3 | -1.17 | 0.766 |
| 5 | 2 | -1.66 | 0.462 |
| SE | 2 | 1 | 0.51 | 0.986 |
| 3 | 1 | 3.49 | 0.004\* |
| 3 | 2 | 1.95 | 0.292 |
| 4 | 1 | -0.02 | 1 |
| 4 | 2 | -0.94 | 0.8796 |
| 4 | 3 | -3.21 | 0.012\* |
| 5 | 1 | -0.91 | 0.8925 |
| 5 | 2 | -1.26 | 0.7161 |
| 5 | 3 | -3.37 | 0.007\* |
| 5 | 4 | -0.55 | 0.9822 |

**Notes:** Table S6 includes post-hoc comparisons via the Steel-Dwass method following the Kruskal Wallis Test. Significant results are indicated with a “\*”, and in these cases there is enough evidence to conclude that a single participant significantly differed from the comparison subject for that respective sleep metric.

**Abbreviations:** SE, sleep efficiency; TST, total sleep time; TWT, total wake time

**Table S7.** Bland Altman Proportional Bias Sleep Staging Summary Statistics: All Devices

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Device** | **R2** | **Adjusted p-Value** |
| Light | Fitbit | 0.210 | 0.653 |
| Garmin | 0.002 | 1 |
| Oura | 0.082 | 0.855 |
| WHOOP | 0.109 | 1 |
| Deep | Fitbit | 0 | 1 |
| Garmin | 0.423 | 0.049\* |
| Oura | 0.186 | 0.023\* |
| WHOOP | 0.013 | 1 |
| REM | Fitbit | 0 | 1 |
| Garmin | 0.002 | 1 |
| Oura | 0.075 | 1 |
| WHOOP | 0.038 | 1 |

**Notes:** Table S7 includes Bland Altman summary statistics for proportional biases, which assesses the degree of proportionality with respect to the bias between reports of the Sleep Profiler and a given device, for all devices across the three sleep staging metrics. Significant results are indicated with a “\*”, and in these cases there is enough evidence to conclude that the device is not “proportionally unbiased” thus demonstrating a tendency to overestimate or underestimate at higher or lower magnitudes for a given sleep staging metric. Data presented for each individual metric are sorted alphabetically by device for each of the three sleep staging metrics displayed.

**Abbreviations:** REM, rapid eye movement



**Figure S1.** Boxplots by Participant for Sleep Summary Metrics

**Notes:** Figure S1 depicts a box plot representation of absolute percent error (APE) for total sleep time (TST; A), total wake time (TWT; B), and sleep efficiency (SE; C) for each participant.

**Abbreviations:** TST, total sleep time; TWT, total wake time; SE, sleep efficiency; APE, absolute percent error; MdAPE, median absolute percent error



**Figure S2A-D.** Light Time Bland Altman Plots: All Devices.

**Notes:** Figure S2 depicts a four panel Bland Altman plot comparison for all devices that estimated the amount of time spent, in hours, in light sleep (Figure S2A, WHOOP; Figure S2B, Fitbit; Figure S2C, Oura; Figure S2D, Garmin).

**Abbreviations:** N/A



**Figure S3A-D.** Deep Time Bland Altman Plots: All Devices.

**Notes:** Figure S3 depicts a four panel Bland Altman plot comparison for all devices that estimated the amount of time spent, in hours, in light sleep (Figure S3A, WHOOP; Figure S3B, Fitbit; Figure S3C, Oura; Figure S3D, Garmin).

**Abbreviations:** N/A



**Figure S4A-D.** REM Time Bland Altman Plots: All Devices.

**Notes:** Figure S4 depicts a four panel Bland Altman plot comparison for all devices that estimated the amount of time spent, in hours, in REM sleep (Figure S4A, WHOOP; Figure S4B, Fitbit; Figure S4C, Oura; Figure S4D, Garmin).

**Abbreviations:** REM, rapid eye movement