

# Actigraphy-Based Sleep Detection: Validation with Polysomnography and Comparison of Performance for Nighttime and Daytime Sleep During Simulated Shift Work

Chenlu Gao<sup>1-3</sup>, Peng Li<sup>1-3</sup>, Christopher J Morris<sup>1,2</sup>, Xi Zheng<sup>1,3</sup>, Ma Cherrysse Ulsa<sup>1,3</sup>, Lei Gao<sup>1-4</sup>, Frank AJL Scheer<sup>1-3</sup>, Kun Hu<sup>1-3</sup>

<sup>1</sup>Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA; <sup>2</sup>Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA; <sup>3</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA; <sup>4</sup>Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Correspondence: Kun Hu, Brigham and Women's Hospital, 221 Longwood Avenue, 040 BLI, Boston, MA, 02115, USA, Tel +1 617-525 8694, Fax +1 617-732 7337, Email [khu1@bwh.harvard.edu](mailto:khu1@bwh.harvard.edu)

**Purpose:** Actigraphy-based sleep detection algorithms were mostly validated using nighttime sleep, and their performance in detecting daytime sleep is unclear. We evaluated and compared the performance of Actiware and the Cole-Kripke algorithm (C-K) – two commonly used actigraphy-based algorithms – in detecting daytime and nighttime sleep.

**Participants and Methods:** Twenty-five healthy young adults were monitored by polysomnography and actigraphy during two in-lab protocols with scheduled nighttime and/or daytime sleep (within-subject design). Mixed-effect models were conducted to compare the sensitivity, specificity, and F1 score (a less-biased measure of accuracy) of Actiware (with low/medium/high threshold setting, separately) and C-K in detecting sleep epochs from actigraphy recordings during nighttime/daytime. *t*-tests and intraclass correlation coefficients were used to assess the agreement between actigraphy-based algorithms and polysomnography in scoring total sleep time (TST).

**Results:** Sensitivity was similar between nighttime (Actiware: 0.93–0.99 across threshold settings; C-K: 0.61) and daytime sleep (Actiware: 0.93–0.99; C-K: 0.66) for both the C-K and Actiware (daytime/nighttime×algorithm interaction:  $p > 0.1$ ). Specificity for daytime sleep was lower (Actiware: 0.35–0.54; C-K: 0.91) than that for nighttime sleep (Actiware: 0.37–0.62; C-K: 0.93;  $p = 0.001$ ). Specificity was also higher for C-K than Actiware ( $p < 0.001$ ), with no daytime/nighttime×algorithm interaction ( $p > 0.1$ ). C-K had lower F1 (nighttime = 0.74; daytime = 0.77) than Actiware (nighttime = 0.95–0.98; daytime = 0.90–0.91) for both nighttime and daytime sleep (all  $p < 0.05$ ). The daytime-nighttime difference in F1 was opposite for Actiware (daytime: 0.90–0.91; nighttime: 0.95–0.98) and C-K (daytime: 0.77; nighttime: 0.74; interaction  $p = 0.003$ ). Bias in TST was lowest in Actiware (with medium-threshold) for nighttime sleep (underestimation of 5.99 min/8h) and in Actiware (with low-threshold) for daytime sleep (overestimation of 17.75 min/8h).

**Conclusion:** Daytime/nighttime sleep affected specificity and F1 but not sensitivity of actigraphy-based sleep scoring. Overall, Actiware performed better than the C-K algorithm. Actiware with medium-threshold was the least biased in estimating nighttime TST, and Actiware with low-threshold was the least biased in estimating daytime TST.

**Keywords:** Actiware, Cole-Kripke algorithm, sleep scoring, shift worker, circadian rhythms

## Plain Language Summary

Actigraphy is widely used in both research and clinical settings to estimate sleep. Two commonly used algorithms for actigraphy-based sleep scoring are the built-in algorithm of the Actiware software (Philips Respironics) and the Cole-Kripke algorithm. The Cole-Kripke algorithm implements re-scoring rules to reduce the overestimation of sleep time. Existing validations of both algorithms predominantly focused on nighttime sleep, whereas it is unknown how accurate they are in detecting daytime sleep. Here, we reported

that the performance of the two algorithms in identifying sleep epochs remained similar for daytime and nighttime sleep, and both algorithms had increased errors in identifying wake epochs during daytime. These findings may provide useful guidance for objective evaluation of daytime sleep using ambulatory actigraphy and justify the necessity for improving actigraphy-based algorithms in scoring daytime sleep.

## Introduction

While polysomnography (PSG) is the gold standard for measuring sleep, it requires costly equipment, electrode placement and equipment calibrations by knowledgeable sleep technicians, and time-consuming manual scoring.<sup>1</sup> A widely used alternative sleep assessment method is actigraphy. Unlike PSG-derived sleep scores that are based on brain activity (EEG), eye movements (EOG), muscle activity and skeletal muscle activation (EMG), actigraphy measures body (wrist) movement and actigraphy-based algorithms use movement activity data to estimate sleep.<sup>2,3</sup> Compared to PSG, actigraphy-based sleep assessment has several advantages, including greater ecological validity, less influence on sleep,<sup>4</sup> and the utility for continuous monitoring across days.<sup>2</sup> Thus, actigraphy may offer a convenient and naturalistic solution of long-term sleep monitoring in the home environment. In 2007, the American Academy of Sleep Medicine published guidelines on using actigraphy to diagnose sleep disorders.<sup>5</sup> Later, a systematic review and meta-analysis concluded that actigraphy is generally a valid and useful tool for diagnosing and assessing treatment outcomes for some sleep disorders, such as delayed sleep-wake phase disorder.<sup>6</sup>

However, because actigraphy distinguishes wake from sleep based on movements in the forms of activity counts and/or raw acceleration, it may erroneously categorize quiet rest as sleep, especially as occurred to insomnia patients who may stay awake while lying still in bed<sup>7</sup> and patients who are bedbound due to illnesses.<sup>8</sup> Indeed, the widely used Actiware (Philips Respironics) algorithm shows high sensitivity (ie, correctly scoring sleep epochs) but low specificity (ie, correctly scoring wake epochs). For example, specificity of Actiware for overnight sleep was 0.48 among a group of 38 healthy young adults,<sup>9</sup> 0.28–0.48 in sleep-disordered patients across different thresholds,<sup>10</sup> and 0.67 in 33 middle-aged adults with lower back pain;<sup>11</sup> specificity of Actiware for daytime naps ranged from 0.36 to 0.64 in healthy young adults,<sup>12</sup> sensitivity of Actiware was above 0.87 and up to 0.96 in all these studies. Due to low specificity, Actiware as well as other actigraphy-based algorithms may generate ambiguous inference about sleep/wake states, resulting in overestimation of sleep.<sup>2</sup>

To address the issue of overestimation of sleep, re-scoring rules have been introduced in the Cole-Kripke algorithm to score actigraphy data.<sup>13,14</sup> For instance, based on one physiological consideration that people may stop moving for a few minutes before falling asleep (ie, the onset of sleep), the first few epochs (eg, 3 minutes) after a certain amount of wakefulness (eg, 10 minutes) scored as sleep are re-scored to wake. A study of 40 young adults showed that applying the re-scoring rules in the Cole-Kripke algorithm improved the specificity from 0.441 to 0.486 while reducing the sensitivity slightly from 0.975 to 0.970.<sup>15</sup> Though the Cole-Kripke algorithm is a commonly used algorithm for scoring sleep based on actigraphy (eg, Gu et al 2020;<sup>16</sup> Widome et al 2020;<sup>17</sup> and epidemiological studies such as the Jackson Heart Sleep Study),<sup>18</sup> especially for data from accelerometer devices other than the Actiwatch (Bend, OR), its performance as comparison to Actiware is still not clear. Certain studies showed that the Cole-Kripke algorithm can achieve a specificity up to 0.65<sup>13</sup> and can reduce the overestimation of total sleep time from 1.89% to 0.81%,<sup>14</sup> while other studies showed the specificity of the Cole-Kripke algorithm is still low (eg, 0.34 among healthy young adults).<sup>19</sup>

Previous validation studies of actigraphy-based sleep scoring have predominantly used data collected during nighttime sleep.<sup>6</sup> Sleep during daytime, such as recovery sleep during the morning hours after night shift and napping in the afternoon before night shift, is common in today's society and an important component of the 24-hour sleep-wake cycle. Many studies have provided convincing evidence for the physiological importance of daytime sleep, showing that, for instance, mid-afternoon recovery sleep improves vigilance.<sup>20</sup> The performance of actigraphy-based algorithms for detecting daytime sleep has rarely been tested.<sup>21</sup> It is unclear whether these algorithms perform similarly or differently for detecting daytime sleep as for nighttime sleep. Addressing the question is crucial for the design of large-scale field sleep studies, in which actigraphy may be the only feasible tool for sleep assessment.

A related question is whether it is appropriate to perform sleep scoring on 24-h actigraphy data when scheduled sleep episodes are unknown (eg, sleep/wake diary is not available or not reliable). Though sleep diary/log is often used in

conjunction with actigraphy,<sup>22</sup> completing sleep diaries adds extra burden to participants and compliance rates vary across studies. For example, a two-week study by Maich et al found 99.8% completion rate of sleep diary.<sup>23</sup> However, in Thurman's study of 30 healthy adults who were asked to wear actiwatches and keep sleep logs for up to 16 consecutive weeks,<sup>24</sup> the overall completion rate of sleep logs was 73% and only 60% of participants completed  $\geq 75\%$  of their 16-week sleep logs, with lower diary completion rates associated with lower agreement between sleep log and actigraphy. On the other hand, considering the health relevance of sleep, long-term sleep monitoring should become more and more common in health care as well as in clinical research, especially for older adults who may not be able to complete sleep diaries (eg, people with Alzheimer's disease). Therefore, testing whether Actiware and the Cole-Kripke algorithms can perform well with unknown sleep episodes is also of importance for sleep study design using actigraphy.

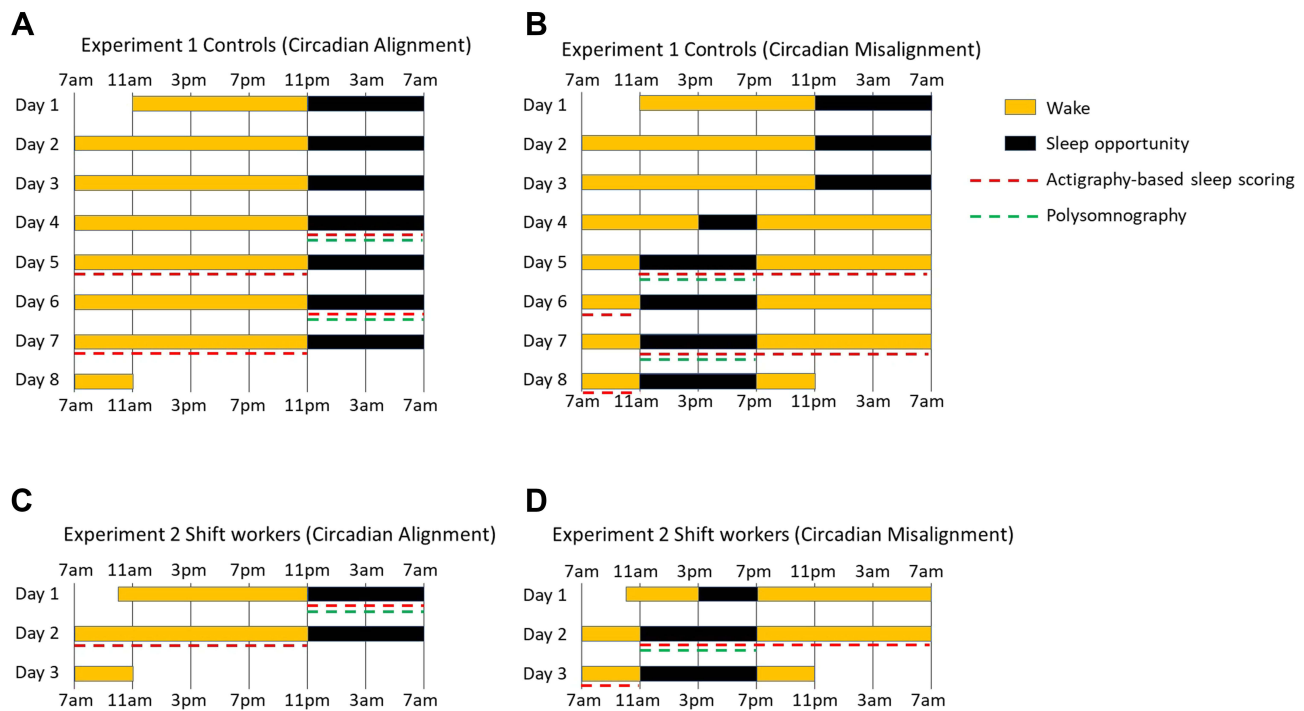
To fill the gaps in the literature, we examined the performance of Actiware and the Cole-Kripke algorithms using the data collected from the same participants during daytime and nighttime sleep. Objective 1 of this study was to test whether actigraphy-based sleep assessment detected sleep epochs during daytime and nighttime sleep episodes equally accurately against polysomnography. Objective 2 of this study was to compare the performance of the two algorithms using the same dataset. If their performance in detecting sleep (ie, sensitivity) differed, we would further explore whether the differences were due to differences in detecting light sleep stages, because the Cole-Kripke's re-scoring rules apply to the first few sleep epochs following wakefulness. Objective 3 of this study was to test the importance of sleep/wake diary by determining the differences in the performance of actigraphy-based sleep assessment when 24-hour actigraphy data were analyzed compared to when only scheduled sleep episodes were analyzed.

## Methods

### Participants and Protocols

We used data from two experiments each with two multi-day in-laboratory protocols. Other aspects of these protocols, designed to test independent hypotheses, have been published previously.<sup>25–38</sup> Experiment 1 included 14 non-shift worker adults (mean age = 28.16y, SD = 9.20y, age range: 20–50y; eight women) who did not perform any shift work in the past three years, were not involved in more than six months of cumulative lifetime shift work, and did not cross more than one time zone in the past three months. Experiment 2 included 11 shift workers (mean age = 34.46y, SD = 7.64y, age range: 24–48y; six women) who had on average 4.52y (SD = 7.74y, range: 1.11–25.09y) of consecutive shift work experience and on average 5.30y (SD = 7.66y, range: 1.11–25.09y) of lifetime cumulative shift work experience. In both experiments, participants were non-smokers and not taking drugs or medications (except for oral contraceptives). Each participant completed a circadian alignment protocol (Experiment 1: [Figure 1A](#); Experiment 2: [Figure 1C](#)) and a circadian misalignment protocol (Experiment 1: [Figure 1B](#); Experiment 2: [Figure 1D](#)). The two protocols were in randomized order and separated by 2–8 weeks (Experiment 1: Mean = 4, SD = 2; Experiment 2: Mean = 4, SD = 1). During the circadian alignment protocol of both experiments, participants had a daily scheduled sleep episode between 11pm–7am. A simulated shift work schedule was used to introduce circadian misalignment in both experiments. Specifically, in Experiment 1 (non-shift workers), following nighttime sleep episodes between 11pm–7am on Days 1–3, participants were scheduled to sleep between 3pm–7pm on Day 4 and 11am–7pm on Days 5–8; in Experiment 2 (chronic shift workers), participants had scheduled sleep episodes between 3pm–7pm on Day 1 and between 11am–7pm on Days 2–3.

Experiment 1 consisted of two 8-day (with 7 nights) laboratory protocols. Sleep was recorded via PSG during the scheduled nighttime sleep episodes (~11pm–7am) on Day 4 and Day 6 of the circadian alignment protocol and during the scheduled daytime sleep episodes (~11am–7pm) on Day 5 and Day 7 of the circadian misalignment protocol ([Figure 1A](#) and [Figure 1](#)). Experiment 2 consisted of two 3-day (with 2 nights) laboratory protocols. Participants were instructed to sleep from 11pm to 7am on the night before the in-laboratory protocols to reduce prior sleep debt. During the study, PSG was recorded during the scheduled nighttime sleep episode (~11pm–7am) on Day 1 of the circadian alignment protocol and during the scheduled daytime sleep episode (~11am–7pm) on Day 2 of the circadian misalignment protocol ([Figure 1C](#) and [Figure 1](#)). For both experiments, each scheduled sleep opportunity episode was 8 hours to allow sufficient sleep and increase generalizability to the general population with normal sleep duration. Wrist actigraphy (Actiwatch Spectrum, Bend, OR) was continuously worn during the whole study period.



**Figure 1** Experiment protocols. **(A and B)** Two 8-day protocols in Experiment 1 were designed for non-shift workers, including **(A)** a circadian alignment protocol with nighttime sleep and **(B)** a misalignment protocol with nighttime sleep from Day 1 to Day 3, an afternoon nap on Day 4, and daytime sleep from Day 5 to Day 8. **(C and D)** Two 3-day protocols in Experiment 2 were designed for chronic shift workers, including **(C)** a circadian alignment protocol with nighttime sleep and **(D)** a misalignment protocol with an afternoon nap on Day 1 followed by nighttime sleep on Days 2–3.

More details on the study methods have been previously reported.<sup>37,38</sup> Both experiments complied with the Declaration of Helsinki, were approved by the Partners Human Research Committee, and were conducted in the Center for Clinical Investigation at Brigham and Women’s Hospital (Boston, Massachusetts). All participants provided written informed consent.

## Data Collection and Analysis

### Polysomnography

During each PSG assessment, data were collected continuously using Vitaport-3 (Temec Instruments, Kerkrade, B.V., The Netherlands) with a sampling rate of 256 Hz, including left and right electrooculography (EOG), bipolar submental electromyography (EMG), bipolar electrocardiography (ECG), and electroencephalography (EEG; F3, F4, C3, C4, O1 and O2 channels, referenced to contralateral mastoids). Research personnel inspected and scored the PSG recordings in 30-second epochs using the Vitascore software (Temec Instruments) based on the American Academy of Sleep Medicine guidelines.<sup>39</sup> Participants were monitored and kept awake outside the scheduled sleep opportunities, based on study protocols.

### Actigraphy

The Actiwatch monitored raw acceleration with a sample frequency of 32 Hz and integrated the data to a proprietary “count” value for every epoch (1 minute). A built-in capacitive sensor determined whether the device was on-wrist or off-wrist. On-wrist epochs of the activity count data were loaded into the following two software packages for sleep/wake scoring. (1) Actiware software (version 5.52; Respironics, Inc. Murrysville, PA)<sup>40</sup> describes the scoring approach as follows:

Whether a particular epoch is scored as wake is determined by comparing [the weighted value of] activity counts [in 5 consecutive epochs centered at] the [current] epoch in question, to a threshold value which can be set by the researcher ... If the number of counts exceeds the threshold the epoch is scored as wake. If it falls below, or is equal to, the threshold the epoch is scored as sleep.

There are three levels for the threshold, 20, 40, and 80 activity counts, that correspond to low-, medium-, and high-sensitivity settings, respectively. (2) The ezActi software was developed by the Medical Biodynamics Program at Brigham and Women's Hospital to provide a graphical user interface in the MATLAB platform for activity data analysis, and the Cole-Kripke algorithm<sup>13,14</sup> was implemented in this software to determine sleep or wake epochs. This software can be accessed at: <https://github.com/pliphd/Actigraphy>.

### Temporal Alignment of PSG and Actigraphy Measurements

Because of the difference in length and start time of epochs between PSG and actigraphy, the beginning of each 1-minute actigraphy epoch did not align perfectly with the beginning of a 30-second PSG epoch. To resolve this, we identified three 30-second PSG epochs that overlapped with each 1-minute actigraphy epoch and considered a 1-minute actigraphy epoch as a true sleep epoch when the three 30-second PSG epochs were all sleep epochs. We further categorized the true sleep epochs into PSG-light sleep (ie, N1 and N2), PSG-slow-wave sleep (SWS; ie, N3), or PSG-rapid eye movement (REM) sleep if their corresponding PSG epochs were all light sleep, SWS, or REM sleep, respectively. Likewise, a 1-min actigraphy epoch was considered a true wake epoch if the corresponding three PSG 30-second epochs were wake. Otherwise, we categorized an epoch as a transitional epoch (1.93% of all epochs), which was excluded from analyses.

### Epoch-by-Epoch Analysis

We compared the epoch-by-epoch scoring from Actiware and the Cole-Kripke algorithm to sleep/wake status obtained from PSG and study protocols (scheduled wakeful episodes). For each participant during each study protocol, we computed sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score (ie, a better measure of accuracy in the case of imbalanced data, which is less biased than the traditional accuracy), and Cohen's kappa using data during the scheduled sleep episodes with PSG. Then, we repeated the calculation of these measures for the 24-h time windows (the 24 hours following the beginning of a sleep opportunity period) in those days with PSG (Figure 1). Sensitivity refers to the proportion of true sleep epochs that was correctly scored as sleep by Actiware/Cole-Kripke algorithm; specificity refers to the proportion of true wake epochs that was correctly scored as wake by Actiware/Cole-Kripke algorithm.<sup>41</sup> PPV is the probability that an epoch scored as sleep by Actiware/Cole-Kripke algorithm was truly a sleep epoch and NPV is the probability that an epoch scored as wake by Actiware/Cole-Kripke algorithm was truly a wake epoch. F1 score is a measure of accuracy that is less affected by the imbalanced samples in a data set.<sup>42</sup> It is the harmonic mean of sensitivity and PPV, specifically:

$$F1 = \frac{2}{\frac{1}{PPV} + \frac{1}{Sensitivity}}$$

Cohen's kappa measures agreement between PSG and Actiware/Cole-Kripke algorithm, with values  $\leq 0.2$  indicating no or slight agreement, 0.2–0.4 indicating fair, 0.4–0.6 indicating moderate, 0.6–0.8 indicating substantial, and 0.8–1.0 indicating perfect agreement.<sup>43</sup> For statistical analyses, we focused on the commonly used metrics (ie, sensitivity, specificity, and F1).

### Summary Sleep Parameters

In addition to epoch-by-epoch analyses, we compared total sleep time (TST) and wake after sleep onset (WASO) measured by the Cole-Kripke algorithm, Actiware, and PSG, because TST and WASO are widely used in both research and clinical settings.

### Statistical Analysis

We first conducted independent-samples *t*-tests (for continuous variables) and chi-square (for dichotomous variables) to compare participants and their sleep characteristics in the two experiments. Because Experiment 1 and Experiment 2 participants were demographically similar and they had similar sleep characteristics, we combined the samples of two experiments for the following analyses. To compare nighttime and daytime sleep characteristics of the same individuals, we conducted mixed-effect models with protocol (2 levels: nighttime/daytime sleep) as the fixed-effect factor and individual subject as the random-effect factor. Outcomes were PSG-measured TST, light sleep, SWS, REM sleep, and

activity counts per epoch during each sleep stage. Mixed-effect models were further used to investigate whether activity counts per epoch were associated with sensitivity.

To evaluate the performance of actigraphy-based sleep, we conducted two models for epoch-by-epoch analyses. Model 1: For objective 1 (nighttime vs daytime sleep) and objective 2 (Cole-Kripke vs Actiware), we conducted a mixed-effect model with protocol (2 levels: nighttime/daytime sleep), Algorithm (4 levels: Actiware low/medium/high threshold or the Cole-Kripke algorithm), and Protocol  $\times$  Algorithm interaction as fixed-effect factors. Individual subject was included in the model as a random-effect factor. Sensitivity, specificity, and F1 scores for *scheduled sleep episodes* and sensitivity during light sleep, SWS, and REM sleep were the outcomes. Post-hoc analyses (pair-wise Tukey LSD test) were conducted for significant main and interaction effects.

Model 2: For objective 3 (scheduled sleep vs 24-hour windows), we added Analysis Window (2 levels: scheduled sleep or 24-h time windows), Analysis Window  $\times$  Protocol interaction, Analysis Window  $\times$  Algorithm interaction, and the three-way interaction of Analysis Window  $\times$  Protocol  $\times$  Algorithm as fixed-effect factors to model 1. Post-hoc analyses (pair-wise Tukey LSD test) were conducted for significant main and interaction effects. Specificity and F1 scores were the outcomes. Sensitivity was not tested as an outcome, because it is not influenced by analysis window.

For analyses on TST and WASO, we computed the mean difference between the Cole-Kripke/Actiware and PSG as a measure of bias and the standard deviation of such difference as a measure of imprecision. We used paired-samples *t*-test to examine whether TST/WASO measured by the Cole-Kripke or Actiware was different from PSG-measured TST/WASO. Complimentarily, we computed the absolute difference between TST/WASO measured by PSG and the Cole-Kripke algorithm or Actiware and used one-sample *t*-test to examine whether such difference was different from 0. To test the agreement in TST between the algorithms and PSG, we computed intraclass correlation coefficients (ICC); we used two-way mixed-effect models to compute ICCs and estimated the reliability of a single measure to examine the consistency between measures.<sup>44</sup> ICC values <0.50, 0.50–0.75, 0.75–0.90, and >0.90 indicate poor, moderate, good, and excellent agreement, respectively. We computed ICCs in SPSS 28 and performed all other analyses in JMP Pro 16; all statistical tests were two-tailed, and statistical significance was accepted at  $\alpha = 0.05$ .

## Results

### Participant and Sleep Characteristics

Participants from Experiment 1 and Experiment 2 had similar age ( $t=1.87$ ,  $p=0.074$ ) and gender distribution ( $\chi^2=0.02$ ,  $p=0.897$ ), as well as PSG-measured TST, light sleep, SWS, and REM sleep minutes per night during nighttime ( $ps>0.05$ ) and daytime sleep ( $ps>0.05$ ; Table 1). Next, we combined participants from the two Experiments and found that TST ( $F=22.85$ ,  $p<0.001$ ), light sleep ( $F=10.47$ ,  $p=0.004$ ), and REM sleep ( $F=14.24$ ,  $p=0.001$ ) were shorter during daytime than nighttime sleep. SWS minutes were similar between nighttime and daytime sleep ( $F=1.78$ ,  $p=0.198$ ).

### Epoch-by-Epoch Analysis

[Supplemental Table 1](#) and [Figure 2](#) present the epoch-by-epoch analysis results for scheduled sleep episodes. [Supplemental Table 2](#) presents the number of epochs scored as sleep or wake.

Model 1 showed a significant difference in sensitivity across algorithms ( $F=248.46$ ,  $p<0.001$ ), no significant difference in sensitivity between nighttime and daytime sleep ( $F=2.41$ ,  $p=0.122$ ), and no significant interaction effect on sensitivity between protocol and algorithms ( $F=1.28$ ,  $p=0.285$ ). Post-hoc analyses showed that Cole-Kripke yielded lower sensitivity than Actiware with different thresholds ( $ps<0.05$ ), and Actiware (low) yielded lower sensitivity than Actiware (high) ( $p<0.05$ ). The difference in sensitivity across algorithms was also observed for each sleep stage (Light:  $F=234.88$ ,  $p<0.001$ ; SWS:  $F=37.58$ ,  $p<0.001$ ; REM:  $F=149.83$ ,  $p<0.001$ ; [Supplemental Table 3](#); [Figure 3](#)). Cole-Kripke was the least sensitive in detecting sleep epochs during any sleep stage ( $ps<0.05$ ). During both light sleep and SWS, sensitivity was similar for nighttime and daytime sleep (light sleep:  $F=0.02$ ,  $p=0.887$ ; SWS:  $F=0.82$ ,  $p=0.367$ ), which was consistent across all algorithms (Protocol  $\times$  Algorithm interactions for light sleep:  $F=0.003$ ,  $p>0.999$ ; SWS:  $F=1.34$ ,  $p=0.265$ ). In contrast, within REM sleep, the algorithms were overall more sensitive for daytime than nighttime sleep ( $F=6.10$ ,  $p=0.015$ ), and this difference in sensitivity was similar across algorithms (Protocol  $\times$  Algorithm interaction:

**Table I** Participant Characteristics and Data Analyzed

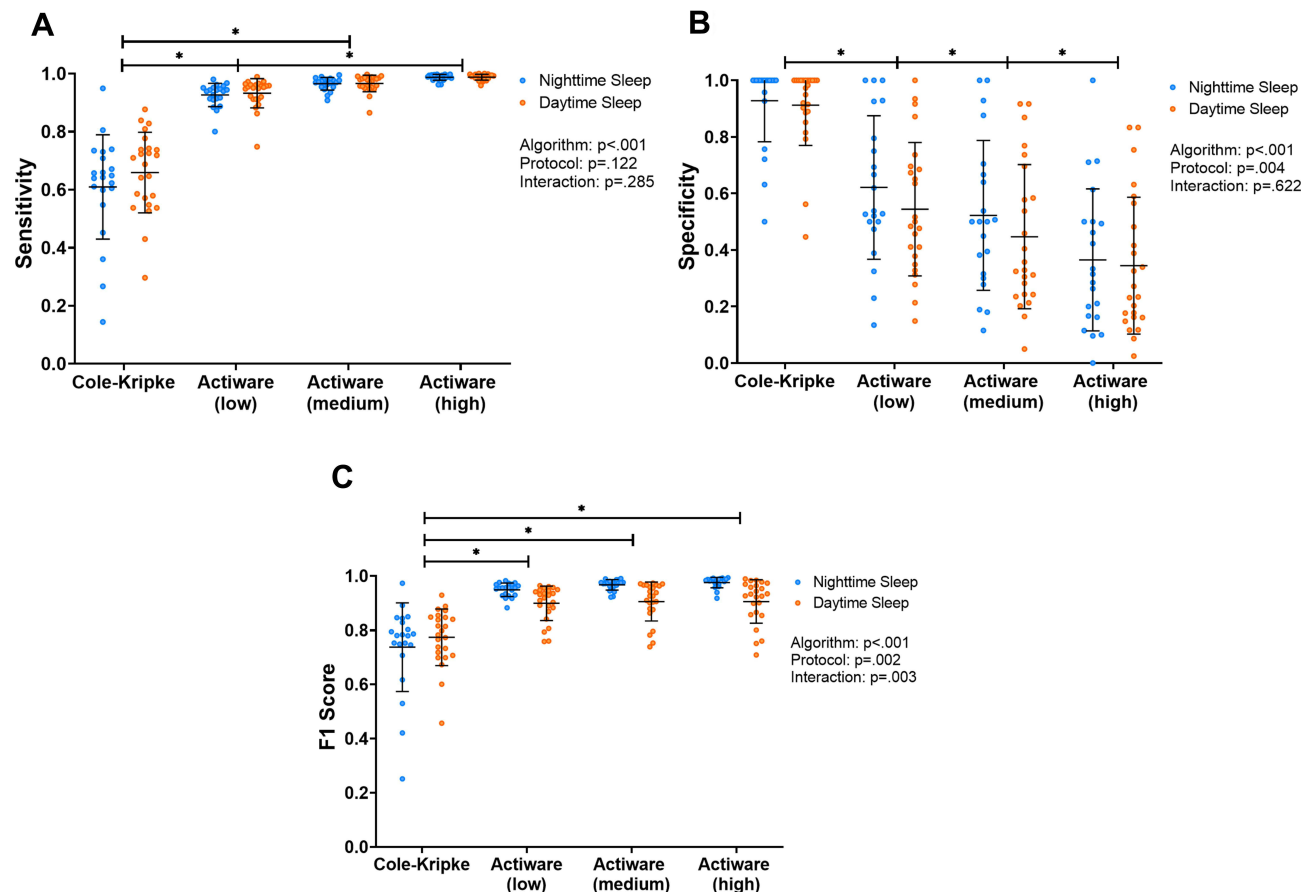
	Experiment 1 (Non-Shift Workers)	Experiment 2 (Shift Workers)	Comparison
N	14	11	
Age (years)	28.16 (9.20), 22.08–32.30	34.45 (7.63), 30.10–39.30	$t=1.87, p=0.074$
Gender (female)	8 (57.14%)	6 (54.55%)	$\chi=0.02, p=0.897$
<b>Circadian Alignment Protocol (Nighttime sleep)</b>			
Scheduled sleep opportunity with actigraphy and PSG (h)	16	8	
PSG total sleep time/night (min)	438.19 (19.68), 424.06–454.25	452.33 (27.77), 445.00–469.00	$t=1.37, p=0.187$
NREM light sleep (N1 and N2; min)	217.13 (31.27), 192.69–230.69	220.61 (28.88), 207.00–244.00	$t=0.26, p=0.797$
NREM light sleep (N1 and N2; %)	49.72 (8.02), 42.52–56.78	48.66 (4.70), 44.49–53.64	$t=0.35, p=0.728$
Slow-wave sleep (min)	104.96 (28.18), 81.50–124.81	108.72 (31.06), 84.25–140.75	$t=0.29, p=0.775$
Slow-wave sleep (%)	23.89 (6.06), 19.67–28.20	24.26 (7.57), 18.46–32.06	$t=0.13, p=0.901$
Rapid eye movement sleep (min)	116.10 (23.85), 90.75–129.88	123.00 (37.52), 94.00–156.75	$t=0.52, p=0.612$
Rapid eye movement sleep (%)	26.39 (4.69), 21.27–29.70	27.08 (7.70), 21.66–34.00	$t=0.25, p=0.802$
<b>Circadian Misalignment Protocol (Daytime sleep)</b>			
Scheduled sleep opportunity with actigraphy and PSG (h)	16	8	
PSG total sleep time/day (min)	386.46 (44.45), 357.63–417.50	368.82 (86.26), 273.50–451.00	$t=0.65, p=0.526$
NREM light sleep (N1 and N2; min)	190.83 (31.54), 163.50–212.25	178.05 (56.92), 112.00–218.00	$t=0.70, p=0.495$
NREM light sleep (N1 and N2; %)	49.68 (8.13), 42.34–55.52	47.75 (7.49), 46.37–53.34	$t=0.60, p=0.554$
Slow-wave sleep (min)	109.31 (27.50), 89.13–130.88	92.50 (30.80), 62.50–116.00	$t=1.41, p=0.172$
Slow-wave sleep (%)	28.16 (5.75), 25.00–31.60	25.51 (7.25), 20.39–29.98	$t=1.00, p=0.328$
Rapid eye movement sleep (min)	86.33 (28.47), 70.38–111.25	98.27 (29.24), 71.50–124.50	$t=1.01, p=0.323$
Rapid eye movement sleep (%)	22.16 (5.92), 18.85–26.35	26.74 (4.76), 25.59–30.12	$t=2.06, p=0.051$

**Notes:** Data presented as n (%) or mean (standard deviation), interquartile ranges. Each scheduled sleep opportunity episode was 8 hours.

**Abbreviations:** NREM, non-rapid eye movement; PSG, polysomnography.

$F=2.12, p=0.100$ ). Furthermore, there was no difference in activity counts per epoch between circadian alignment and misalignment protocols during wakefulness, NREM light sleep, or SWS ([Supplemental Table 4](#)), but activity counts were higher during nighttime than daytime REM sleep ( $p=0.010$ ). During REM sleep, higher activity counts per epoch were associated with lower sensitivity in all algorithms ( $ps<0.001$ ).

Specificity was lower for daytime than nighttime sleep ( $F=8.37, p=0.004$ ), differed across algorithms ( $F=124.05, p<0.001$ ), and showed no Protocol  $\times$  Algorithm interaction ( $F=0.59, p=0.622$ ). Post-hoc analyses revealed that specificity was the highest in Cole-Kripke, and it decreased as the Actiware threshold increased from low to high (all pairwise  $ps<0.05$ ). To explore the factors that contributed to the difference in specificity between daytime and nighttime sleep, we further examined those PSG-based wake epochs within sleep episodes ([Supplemental Table 4](#)). We found that, despite longer duration of wake bouts during daytime sleep, mean activity counts in each wake bout during daytime sleep were much lower than those during nighttime sleep ( $p<0.001$ ). In addition, lower activity accounts were associated with lower specificity within both daytime and nighttime sleep episodes for all algorithms ( $ps<0.001$ ).



**Figure 2** (A) Sensitivity, (B) specificity, and (C) F1 score comparison of actigraphy-based sleep scoring algorithms for scheduled sleep episodes. Each circle represents the value from an individual study protocol of a participant. The horizontal lines represent the mean and 1 standard deviation above/below the mean. \*Indicates  $p < 0.05$ .

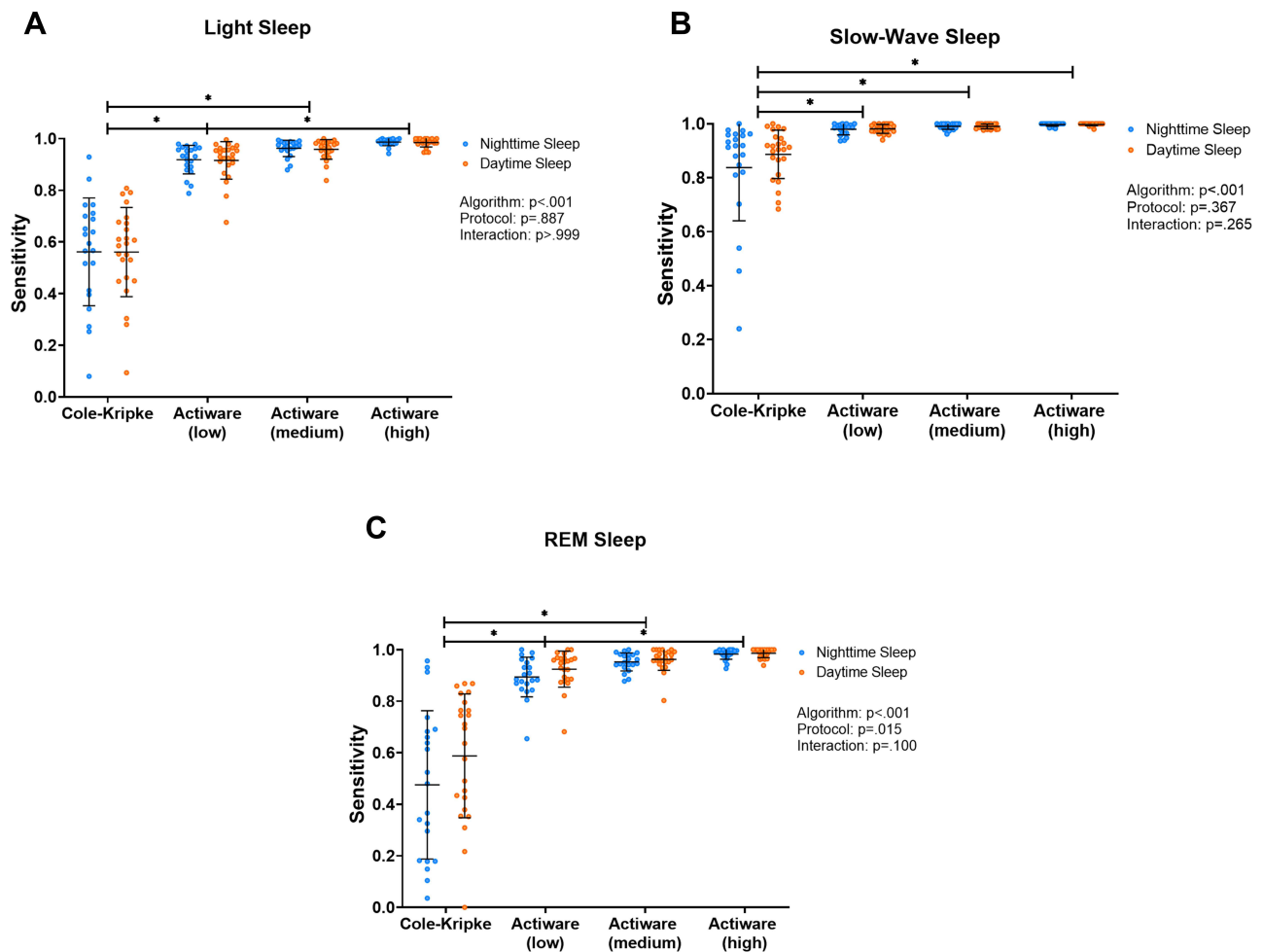
F1 differed across algorithms ( $F=65.72$ ,  $p < 0.001$ ; [Supplemental Table 1](#); [Figure 2C](#)) and the difference appeared to depend on scheduled sleep time (Protocol  $\times$  Algorithm interaction:  $F=4.96$ ,  $p=0.003$ ). Post-hoc analyses showed that Cole-Kripke yielded lower F1 than Actiware with different thresholds for both daytime and nighttime sleep ( $ps < 0.05$ ), F1 of Cole-Kripke showed an improving trend for scoring daytime sleep, and F1 of Actiware showed an improving trend for scoring nighttime sleep. Consistently, the difference in F1 between Cole-Kripke and Actiware was more pronounced for nighttime sleep.

## Scheduled Sleep vs 24-Hour Analysis Window

Overall specificity increased when using 24-h time windows as compared to that based on scheduled sleep episodes ( $F=49.18$ ,  $p < 0.001$ ; [Figure 4A](#)). This effect was mainly driven by Actiware rather than Cole-Kripke (Analysis Window  $\times$  Algorithm interaction:  $F=5.57$ ,  $p=0.001$ ), ie, Actiware had improved specificity for 24-hour analysis window than scheduled sleep episodes ( $ps < 0.05$ ), while no improvement was observed in Cole-Kripke ( $p > 0.05$ ). Additionally, the difference in specificity between 24-hour and scheduled sleep episodes was similar for daytime and nighttime sleep (Analysis Window  $\times$  Protocol interaction:  $F=1.49$ ,  $p=0.223$ ; Analysis Window  $\times$  Algorithm  $\times$  Protocol three-way interaction:  $F=0.22$ ,  $p=0.882$ ). Though the change of specificity was different for different algorithms when using 24-h windows, Cole-Kripke still had the highest specificity and Actiware (high) had the lowest specificity for both nighttime and daytime sleep.

F1 score decreased when using data across 24 h as compared to that using data during scheduled sleep episodes ( $F=789.19$ ,  $p < 0.001$ ). This decrease in F1 was observed in all algorithms ( $ps < 0.05$ ), but it was larger in Actiware than Cole-Kripke (Analysis Window  $\times$  Algorithm interaction:  $F=44.74$ ,  $p < 0.001$ ; [Figure 4B](#)). Protocol (daytime/nighttime





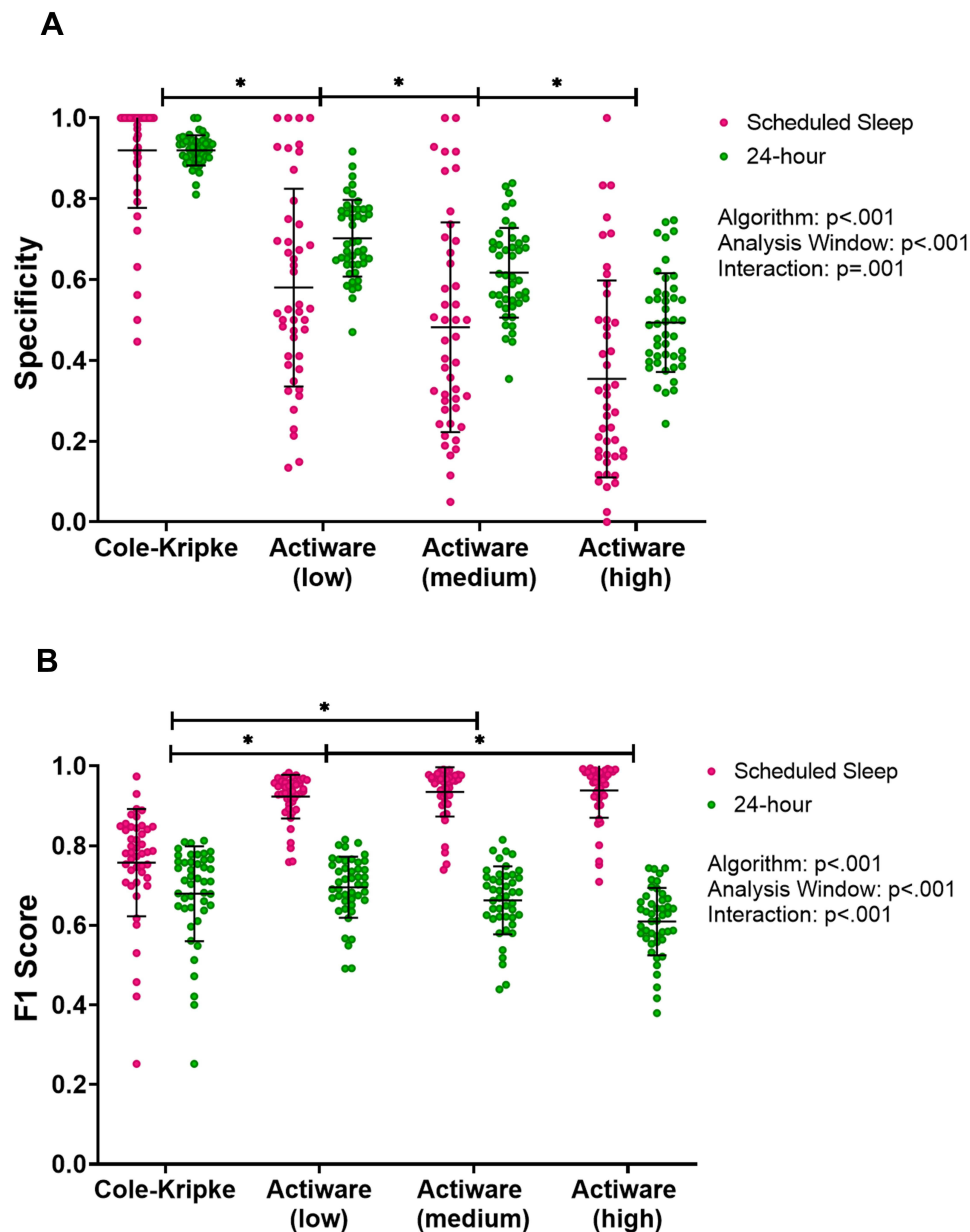
**Figure 3** Comparison of sensitivity of actigraphy-based sleep scoring algorithms in detecting sleep during (A) light sleep (ie, N1 and N2), (B) slow-wave sleep (N3), and (C) REM sleep stages. Each circle represents the value from an individual study protocol of a participant. The horizontal lines represent the mean and 1 standard deviation above/below the mean. \* $p < 0.05$ . \*Indicates  $p < 0.05$ .

**Abbreviation:** REM, rapid eye movement.

sleep) had no significant influences on these changes in F1 score (Analysis Window  $\times$  Protocol interaction:  $F = 0.28$ ,  $p = 0.594$ ; Analysis Window  $\times$  Algorithm  $\times$  Protocol three-way interaction:  $F = 0.07$ ,  $p = 0.974$ ).

## Summary Sleep Parameters

Complementary to the epoch-by-epoch analyses, we examined whether TST scored by actigraphy algorithms agreed with TST measured by PSG. As shown in Table 2 and Figure 5, during the scheduled sleep episodes, Cole-Kripke agreed with PSG poorly (ICC = 0.29) for nighttime sleep and moderately (ICC = 0.57) for daytime sleep (Supplemental Figure 1A and B). Actiware (low) showed moderate agreement with PSG for nighttime sleep (ICC = 0.65) and poor agreement for daytime sleep (ICC = 0.49; Supplemental Figure 1C and D). Actiware (medium) scored nighttime sleep with moderate agreement (ICC = 0.72) and daytime sleep with poor agreement with PSG (ICC = 0.47; Supplemental Figure 1E and F). Actiware (high) showed moderate agreement with PSG for nighttime (ICC = 0.66) and poor agreement for daytime sleep (ICC = 0.38; Supplemental Figure 1G and H). In summary, Cole-Kripke underestimated TST for both nighttime and daytime sleep (Table 2). Actiware (medium) showed the least bias and highest agreement with PSG for nighttime sleep; Actiware (low) showed the least bias and Cole-Kripke showed highest agreement with PSG for daytime sleep.



**Figure 4 (A)** Specificity and **(B)** F1 score comparison of actigraphy-based sleep scoring algorithms using data during scheduled sleep episodes and using 24-h data. Each circle represents the value from an individual study Protocol of a participant. The horizontal lines represent the mean and 1 standard deviation above/below the mean. \*Indicates  $p < 0.05$ .

Next, during 24-hour time windows, all algorithms showed poor agreement with PSG for both nighttime and daytime sleep (ICCs  $< 0.50$ ; Table 2; Figure 6; Supplemental Figure 2), but Cole-Kripke yielded the least bias. All Actiware thresholds overestimated TST, with greater overestimation in higher thresholds.

As shown in Supplemental Table 6, all algorithms yielded poor agreement with PSG in scoring WASO for nighttime and daytime sleep (ICCs  $< 0.50$ ).

## Discussion

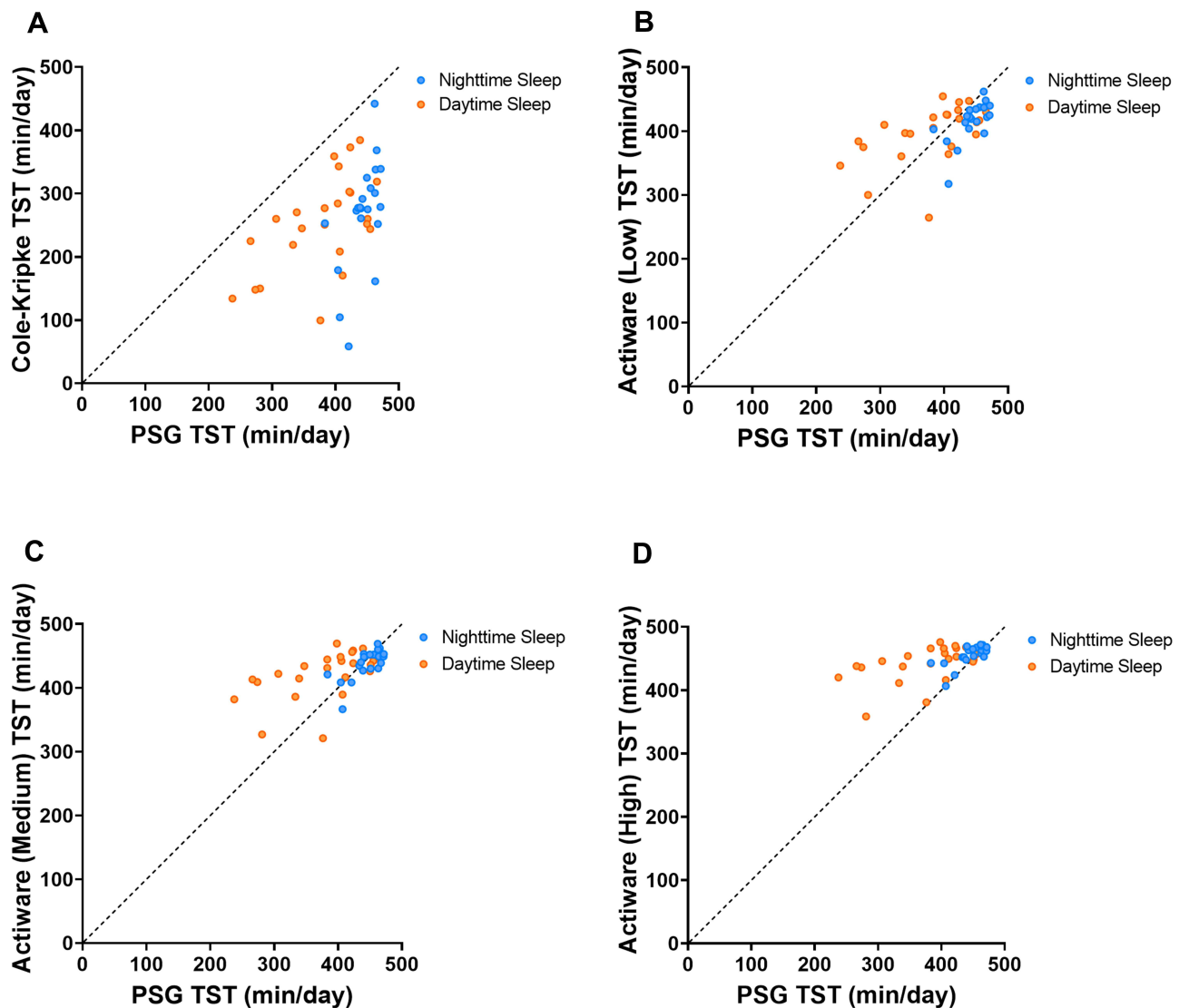
Using simulated shift work data collected from the same individuals with both daytime and nighttime sleep, we compared the performance of Cole-Kripke algorithm and Actiware software in identifying sleep/wake epochs and achieved three objectives.

**Table 2** Agreement Between the Cole-Kripke Algorithm, Actiware, and PSG on Total Sleep Time (TST)

		Scheduled Sleep Episodes			24-Hour Periods			16-h Non-Scheduled Sleep Periods
	PSG TST Min/Night (Mean, SD, Interquartile Range)	Bias (Mean, SD)	Absolute Difference (Mean, SD)	PSG-Actigraphy Agreement (ICC, 95% CI)	Bias (Mean, SD)	Absolute Difference (Mean, SD)	PSG-Actigraphy Agreement (ICC, 95% CI)	Bias (Mean, SD)
<b>Circadian Alignment Protocol; Nighttime Sleep</b>								
Cole-Kripke	444.25 (23.93), 434.50–463.13	−175.58 (75.18)***	175.58 (75.18)***	0.29 (−0.15, 0.64)	−103.18 (106.71)***	118.49 (88.45)***	0.22 (−0.23, 0.59)	72.40 (39.28)***
Actiware (low)	444.25 (23.93), 434.50–463.13	−27.85 (23.32)***	29.70 (20.77)***	0.65 (0.31, 0.84)	232.08 (109.93)***	233.42 (106.92)***	0.18 (−0.27, 0.56)	259.93 (94.96)***
Actiware (medium)	444.25 (23.93), 434.50–463.13	−5.99 (17.75)	14.08 (12.03)***	0.72 (0.42, 0.87)	332.99 (116.56)***	332.99 (116.56)***	0.16 (−0.28, 0.55)	338.98 (110.14)***
Actiware (high)	444.25 (23.93), 434.50–463.13	11.63 (17.08)**	14.70 (14.38)***	0.66 (0.32, 0.84)	467.96 (120.45)***	467.96 (120.45)***	0.16 (−0.28, 0.54)	456.33 (118.05)***
<b>Circadian Misalignment Protocol; Daytime Sleep</b>								
Cole-Kripke	378.38 (65.93), 334.50–423.69	−124.98 (65.72)***	124.98 (65.72)***	0.57 (0.23, 0.79)	−50.21 (79.00)**	74.90 (54.93)***	0.49 (0.12, 0.74)	74.77 (28.80)***
Actiware (low)	378.38 (65.93), 334.50–423.69	17.75 (57.29)	47.75 (35.09)***	0.49 (0.11, 0.74)	286.13 (116.62)***	286.13 (116.62)***	0.13 (−0.29, 0.50)	268.38 (75.74)***
Actiware (medium)	378.38 (65.93), 334.50–423.69	43.06 (55.27)***	55.08 (42.71)***	0.47 (0.09, 0.73)	387.71 (128.86)***	387.71 (128.86)***	0.06 (−0.35, 0.44)	344.65 (92.71)***
Actiware (high)	378.38 (65.93), 334.50–423.69	64.38 (56.43)***	65.33 (55.26)***	0.38 (−0.2, 0.68)	522.25 (141.82)***	522.25 (141.82)***	0.01 (−0.39, 0.41)	457.88 (109.41)***

**Notes:** Bias was computed by subtracting PSG-measured TST from TST scored by the Cole-Kripke or Actiware algorithm. Positive values indicate that actigraphy-based algorithm overestimated TST; negative values indicate that actigraphy-based algorithm underestimated TST. \*\*Indicates  $p < 0.01$ , \*\*\*Indicates  $p < 0.001$ .

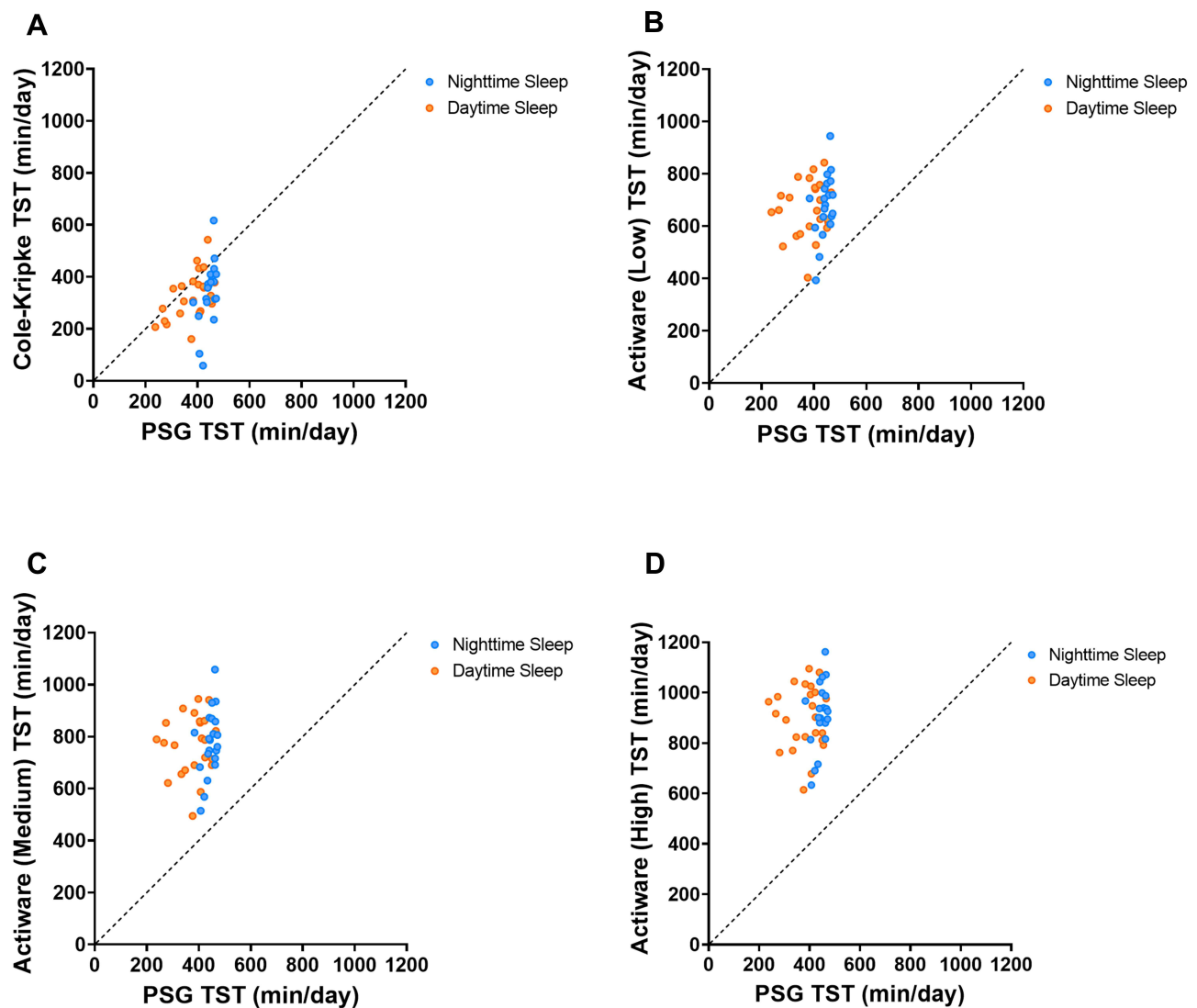
**Abbreviations:** ICC, intraclass correlation coefficient; PSG, polysomnography; TST, total sleep time.



**Figure 5** Agreement in total sleep time during scheduled sleep episodes between PSG- and (A) Cole-Kripke algorithm, (B) Actiware with low threshold, (C) Actiware with medium threshold, and (D) Actiware with high threshold. The dashed lines represent perfect agreement.

**Abbreviation:** TST, total sleep time.

For objective 1 (nighttime vs daytime sleep), overall sensitivity for both algorithms was similar between daytime and nighttime except that sensitivity in detecting sleep epochs within REM sleep was higher for daytime sleep than that for nighttime sleep. The lower sensitivity during nighttime for REM sleep may be explained by greater activity counts across REM epochs during nighttime than those of daytime because actigraphy-based sleep scoring is mainly based on activity levels (ie, low levels for sleep). Different from sensitivity, specificity was lower for daytime sleep, which is consistent for both algorithms. This finding is consistent with previous studies. For example, using a within-subject design, Paquet found that specificity of Actiware (as compared to PSG) was lower for two daytime recovery sleep episodes than that during a nocturnal sleep episode,<sup>45</sup> and another actigraphy study of a diverse sample of participants showed lower specificity of the Cole-Kripke algorithm during daytime sleep in night workers as compared to that during nighttime sleep in sleep-restricted individuals.<sup>46</sup> We showed that the daytime-nighttime difference in specificity was caused by lower mean activity levels within wake bouts during daytime sleep that led to lower specificity despite longer duration of wake bouts during daytime sleep. Regarding TST, we found that Actiware (medium) showed the least bias for nighttime sleep and Actiware (low) showed the least bias for daytime sleep. Though the Cole-Kripke algorithm showed the highest



**Figure 6** Agreement in total sleep time during 24-hour periods between PSG- and (A) Cole-Kripke algorithm, (B) Actiware with low threshold, (C) Actiware with medium threshold, and (D) Actiware with high threshold. The dashed lines represent perfect agreement.

**Abbreviation:** TST, total sleep time.

agreement with PSG for daytime sleep, it showed significant bias. In conjunction with previous studies showing the least bias of Actiware low threshold (compared to medium/high) in estimation of TST among patients with sleep disorders,<sup>7,10</sup> Actiware low-threshold (rather than medium/high) should be recommended for estimating TST when sleep may be disrupted.

Our objective 2 was to compare the Cole-Kripke algorithm with Actiware. The Cole-Kripke algorithm yielded lower sensitivity, higher specificity, and lower F1 scores than Actiware (all three thresholds). Our findings on Actiware (ie, high sensitivity and low specificity) were consistent with past observations.<sup>6</sup> However, our findings on the Cole-Kripke algorithm were different from past observations of high sensitivity, low specificity, and accurate estimation of TST.<sup>13,15,19</sup> The differences may be caused by different accelerometer devices (eg, Motionlogger Actigraph,<sup>13</sup> Motionlogger Micro Watch Actigraph,<sup>15</sup> Mini Motionlogger Actigraph – Basic 32 C<sup>19</sup> vs Actiwatch Spectrum in this study) with different configurations to obtain activity counts (eg, zero-crossing mode<sup>13,15,19</sup> vs proportional integrated mode in this study), PSG scoring criteria (eg, standard criteria of Rechtschaffen and Kales<sup>13,19</sup> vs AASM guidelines in this study), populations (eg, middle-aged adults,<sup>13</sup> healthy young adults,<sup>15,19</sup> vs healthy shift workers and non-shift workers in this study), and/or the study environment (eg, at home<sup>15</sup> vs laboratory environment in this study).

Our objective 3 was to compare the algorithms' performance between scheduled sleep episodes and 24-h time windows. For the Cole-Kripke algorithm, specificity was similar for scheduled sleep and 24-hour periods, but F1 was lower for 24-hour windows. For Actiware, specificity was higher but F1 was lower for 24-hour windows than scheduled sleep episodes. The improvement in specificity is likely due to the inclusion of more wake epochs with higher activity counts for 24-hour windows. Researchers and clinicians often employ different methods to assess sleep via actigraphy. For example, in some studies, actigraphy was collected continuously across multiple days and nights (eg, Gao et al 2019),<sup>47</sup> whereas in other studies, actigraphy was only collected during scheduled sleep episodes at night (eg, Alsaadi et al 2014;<sup>11</sup> common among studies that only used actigraphy for one night).<sup>2</sup> The proportion of sleep epochs across all available epochs would influence the estimation of specificity and accuracy of actigraphy-based algorithms, especially when participants were relatively sedentary. Therefore, limiting analyses to sleep episodes, for example via sleep diary, appears to improve the accuracy of sleep assessment.

Strengths of the current study include the diverse samples of non-shift workers and shift workers, within-subject design for both circadian alignment and misalignment protocols (nighttime and daytime sleep), and the comparison of algorithms using the same data sets. In a laboratory environment, we were also able to closely monitor sleep/wake using both video and PSG while minimizing potential confounders during simulated shift work (eg, differences in sleep environment). The compromise of both algorithms in focusing on only one performance measure (either sensitivity or specificity) while sacrificing the other calls for the development of new approaches that can balance and/or improve both measures. One of the limitations is that all participants were young and healthy. The performance of the algorithms should also be tested in older populations and clinical populations with sleep, circadian, or movement disorders and during sedentary periods. We combined non-shift workers and shift workers in our experiments as they were demographically similar, but future studies with sufficient sample sizes should investigate the populations separately. Moreover, in the current study, the scheduled sleep opportunities were regular across days and the duration of each sleep opportunity was 8 hours during daytime/nighttime while no spontaneous sleep was allowed outside of this window. Our findings based on 8-h scheduled sleep episodes in the laboratory may not be generalizable to short sleep episodes (eg, napping). For instance, Cole-Kripke's re-scoring rules should be reconsidered for detection of short naps. Given that naps are linked to diseases (eg, dementia),<sup>48</sup> algorithms and validation studies are needed on detecting naps via actiwatches to facilitate investigations of objective naps and diseases. Last, we only evaluated two actigraphy algorithms using activity counts collected with the Actiwatch Spectrum devices that automatically detect and exclude off-wrist periods.<sup>49</sup> Future studies are needed to examine other actigraphy-scoring algorithms (eg, the Sadeh algorithm)<sup>50</sup> or other sleep-monitoring devices (eg, other wristband accelerometers or non-wearables)<sup>51</sup> especially those without on-wrist/in-use detection.

In conclusion, across Actiware and Cole-Kripke algorithms, sensitivity of sleep scoring was similar between daytime and nighttime sleep and specificity was lower for daytime sleep. Actiware outperformed the Cole-Kripke algorithm in sleep scoring of actigraphic recordings for both daytime and nighttime sleep episodes scheduled in laboratory. On average, Actiware with low-threshold yielded the least bias in estimating daytime TST and Actiware with medium-threshold yielded the least bias in estimating nighttime TST. Additionally, using sleep diary and limiting the actigraphy-based analyses to known periods in bed should improve sleep assessment.

## Abbreviations

C-K, Cole-Kripke; ECG, electrocardiography; EEG, electroencephalography; EOG, electrooculography; EMG, electromyography; ICC, intraclass correlation coefficient; NREM, non-rapid eye movement; PPV, positive predictive value; PSG, polysomnography; REM, rapid eye movement; SWS, slow-wave sleep; TST, total sleep time.

## Funding

This research was supported by NIH R01HL094806, RF1AG064312, RF1AG059867. C.G. and P.L. are also supported by the BrightFocus Foundation Alzheimer's Disease Research Program (A2020886S). C.G. is additionally supported by the Alzheimer's Association (AARFD-22-928372). L.G. is also supported by the National Institute on Aging (NIA) grant (R03AG067985). F.A.J.L.S. was further supported by NIH R01HL153969.

## Disclosure

C.J.M is an employee of and holds stock/stock options in Biogen. This employment is not related to the current work. CG reports grants from NIH, BrightFocus Foundation, and National Institute on Aging, during the conduct of the study. PL reports grants from The BrightFocus Foundation and grants from NIH, outside the submitted work. FAJLS serves on the Board of Directors for the Sleep Research Society and has received consulting fees from the University of Alabama at Birmingham. FAJLS interests were reviewed and managed by Brigham and Women's Hospital and Partners HealthCare in accordance with their conflict of interest policies. FAJLS consultancies are not related to the current work. The authors report no other conflicts of interest in this work.

## References

1. Stone KL, Ancoli-Israel S. Actigraphy. In: *Principles and Practice of Sleep Medicine*. 6th ed. Elsevier; 2016:1671–1678.
2. Conley S, Knies A, Batten J, et al. Agreement between actigraphic and polysomnographic measures of sleep in adults with and without chronic conditions: a systematic review and meta-analysis. *Sleep Med Rev*. 2019;46:151–160. doi:10.1016/j.smrv.2019.05.001
3. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine clinical practice guideline. *J Clin Sleep Med*. 2018;14(7):1231–1237. doi:10.5664/jcsm.7230
4. Agnew HW, Webb WB, Williams RL. The first night effect: an EEG study of sleep. *Psychophysiology*. 1966;2(3):263–266. doi:10.1111/j.1469-8986.1966.tb02650.x
5. Morgenthaler T, Alessi C, Friedman L, et al. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep*. 2007;30(4):519–529. doi:10.1093/sleep/30.4.519
6. Smith MT, McCrae CS, Cheung J, et al. Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: an American Academy of Sleep Medicine systematic review, meta-analysis, and GRADE assessment. *J Clin Sleep Med*. 2018;14(7):1209–1230. doi:10.5664/jcsm.7228
7. Taibi DM, Landis CA, Vitiello MV. Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. *J Clin Sleep Med*. 2013;9(3):217–225. doi:10.5664/jcsm.2482
8. Beecroft JM, Ward M, Younes M, Crombach S, Smith O, Hanly PJ. Sleep monitoring in the intensive care unit: comparison of nurse assessment, actigraphy and polysomnography. *Intensive Care Med*. 2008;34(11):2076–2083. doi:10.1007/s00134-008-1180-y
9. Chakar B, Senny F, Poirrier AL, Cambron L, Fanielle J, Poirrier R. Validation of midsagittal jaw movements to measure sleep in healthy adults by comparison with actigraphy and polysomnography. *Sleep Sci*. 2017;10(3):122. doi:10.5935/1984-0063.20170021
10. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*. 2001;2(5):389–396. doi:10.1016/S1389-9457(00)00098-8
11. Alsaadi SM, McAuley JH, Hush JM, et al. Assessing sleep disturbance in low back pain: the validity of portable instruments. *PLoS One*. 2014;9(4):e95824. doi:10.1371/journal.pone.0095824
12. Cellini N, Buman MP, McDevitt EA, Ricker AA, Mednick SC. Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiol Int*. 2013;30(5):691–698. doi:10.3109/07420528.2013.782312
13. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep*. 1992;15(5):461–469. doi:10.1093/sleep/15.5.461
14. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. *Sleep*. 1982;5(4):389–399. doi:10.1093/sleep/5.4.389
15. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance comparison of different interpretative algorithms utilized to derive sleep parameters from wrist actigraphy data. *Chronobiol Int*. 2019;36(12):1752–1760. doi:10.1080/07420528.2019.1679826
16. Gu C, Brereton N, Schweitzer A, et al. Metabolic effects of late dinner in healthy volunteers—a randomized crossover clinical trial. *J Clin Endocrinol Metab*. 2020;105(8):2789–2802. doi:10.1210/clinem/dgaa354
17. Widome R, Berger AT, Iber C, et al. Association of delaying school start time with sleep duration, timing, and quality among adolescents. *JAMA Pediatr*. 2020;174(7):697–704. doi:10.1001/jamapediatrics.2020.0344
18. Jackson CL, Ward JB, Johnson DA, Sims M, Wilson J, Redline S. Concordance between self-reported and actigraphy-assessed sleep duration among African-American adults: findings from the Jackson Heart Sleep Study. *Sleep*. 2020;43(3):zsz246. doi:10.1093/sleep/zsz246
19. de Souza L, Benedito-Silva AA, Pires MLN, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. *Sleep*. 2003;26(1):81–85. doi:10.1093/sleep/26.1.81
20. Lo JC, Leong RL, Ng AS, et al. Cognitive effects of split and continuous sleep schedules in adolescents differ according to total sleep opportunity. *Sleep*. 2020;43(12):zsaa129. doi:10.1093/sleep/zsaa129
21. Kanady JC, Drummond SP, Mednick SC. Actigraphic assessment of a polysomnographic-recorded nap: a validation study. *J Sleep Res*. 2011;20(1pt2):214–222. doi:10.1111/j.1365-2869.2010.00858.x
22. Littner M, Kushida CA, Anderson WM, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep*. 2003;26(3):337–341. doi:10.1093/sleep/26.3.337
23. Maich KH, Lachowski AM, Carney CE. Psychometric properties of the consensus sleep diary in those with insomnia disorder. *Behav Sleep Med*. 2018;16(2):117–134. doi:10.1080/15402002.2016.1173556
24. Thurman SM, Wasylyshyn N, Roy H, et al. Individual differences in compliance and agreement for sleep logs and wrist actigraphy: a longitudinal study of naturalistic sleep in healthy adults. *PLoS One*. 2018;13(1):e0191883. doi:10.1371/journal.pone.0191883
25. Li P, Morris CJ, Patxot M, et al. Reduced tolerance to night shift in chronic shift workers: insight from fractal regulation. *Sleep*. 2017;40(7):zsz092. doi:10.1093/sleep/zsz092
26. Morris CJ, Purvis TE, Hu K, Scheer FA. Circadian misalignment increases cardiovascular disease risk factors in humans. *Proc Natl Acad Sci*. 2016;113(10):E1402–E1411. doi:10.1073/pnas.1516953113

27. Morris CJ, Purvis TE, Mistretta J, Hu K, Scheer FA. Circadian misalignment increases C-reactive protein and blood pressure in chronic shift workers. *J Biol Rhythms*. 2017;32(2):154–164. doi:10.1177/0748730417697537
28. Chellappa SL, Morris CJ, Scheer FA. Circadian misalignment increases mood vulnerability in simulated shift work. *Sci Rep*. 2020;10(1):1–10. doi:10.1038/s41598-020-75245-9
29. Tucker MA, Morris CJ, Morgan A, et al. The relative impact of sleep and circadian drive on motor skill acquisition and memory consolidation. *Sleep*. 2017;40(4):zsx036. doi:10.1093/sleep/zsx036
30. Chellappa SL, Morris CJ, Scheer FAJL. Daily circadian misalignment impairs human cognitive performance task-dependently. *Sci Rep*. 2018;8(1):3041. doi:10.1038/s41598-018-20707-4
31. Chellappa SL, Morris CJ, Scheer FAJL. Effects of circadian misalignment on cognition in chronic shift workers. *Sci Rep*. 2019;9(1):699. doi:10.1038/s41598-018-36762-w
32. Qian J, Morris CJ, Caputo R, Wang W, Garaulet M, Scheer FA. Sex differences in the circadian misalignment effects on energy regulation. *Proc Natl Acad Sci*. 2019;116(47):23806–23812. doi:10.1073/pnas.1914003116
33. Morris CJ, Garcia JI, Myers S, Yang JN, Trienekens N, Scheer FA. The human circadian system has a dominating role in causing the morning/evening difference in diet-induced thermogenesis. *Obesity*. 2015;23(10):2053–2058. doi:10.1002/oby.21189
34. Qian J, Dalla Man C, Morris CJ, Cobelli C, Scheer FA. Differential effects of the circadian system and circadian misalignment on insulin sensitivity and insulin secretion in humans. *Diabetes Obes Metab*. 2018;20(10):2481–2485. doi:10.1111/dom.13391
35. Qian J, Morris CJ, Caputo R, Garaulet M, Scheer FA. Ghrelin is impacted by the endogenous circadian system and by circadian misalignment in humans. *Int J Obes*. 2019;43(8):1644–1649. doi:10.1038/s41366-018-0208-9
36. Qian J, Morris CJ, Phillips AJK, et al. Unanticipated daytime melatonin secretion on a simulated night shift schedule generates a distinctive 24-h melatonin rhythm with antiphasic daytime and nighttime peaks. *J Pineal Res*. 2022;72(3):e12791. doi:10.1111/jpi.12791
37. Morris CJ, Purvis TE, Mistretta J, Scheer FA. Effects of the internal circadian system and circadian misalignment on glucose tolerance in chronic shift workers. *J Clin Endocrinol Metab*. 2016;101(3):1066–1074. doi:10.1210/jc.2015-3924
38. Morris CJ, Yang JN, Garcia JI, et al. Endogenous circadian system and circadian misalignment impact glucose tolerance via separate mechanisms in humans. *Proc Natl Acad Sci U S A*. 2015;112(17):E2225–2234. doi:10.1073/pnas.1418955112
39. Iber C, Ancoli-Israel S, Chesson A, Quan SF. *The AASM Manual for the Scoring of Sleep and Associated Events*. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
40. Oakley NR. *Validation with Polysomnography of the Sleepwatch Sleep/Wake Scoring Algorithm Used by the Actiwatch Activity Monitoring System*. Bend: Mini Mitter, Cambridge Neurotechnology; 1997.
41. Parikh R, Mathai A, Parikh S, Sekhar GC, Thomas R. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*. 2008;56(1):45. doi:10.4103/0301-4738.37595
42. Sasaki Y, Yamamoto W, Wada M. The truth of the F-measure. *Palliative Supportive Care*. 2007;5:411–414. doi:10.1017/s1478951507000612
43. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Medica*. 2012;22(3):276–282. doi:10.11613/BM.2012.031
44. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012
45. Paquet J, Kawinska A, Carrier J. Wake detection capacity of actigraphy during sleep. *Sleep*. 2007;30(10):1362–1369. doi:10.1093/sleep/30.10.1362
46. Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep*. 2013;36(11):1747–1755. doi:10.5665/sleep.3142
47. Gao C, Terlizzese T, Scullin MK. Short sleep and late bedtimes are detrimental to educational learning and knowledge transfer: an investigation of individual differences in susceptibility. *Chronobiol Int*. 2019;36(3):307–318. doi:10.1080/07420528.2018.1539401
48. Li P, Gao L, Yu L, et al. Daytime napping and Alzheimer's dementia: a potential bidirectional relationship. *Alzheimer's Dementia*. 2022. doi:10.1002/alz.12636
49. Philips. Actiwatch spectrum activity monitor. Available from: <https://www.usa.philips.com/healthcare/product/HC1046964/actiwatch-spectrum-activity-monitor>. Accessed September 29, 2022.
50. Sadeh A, Sharkey M, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep*. 1994;17(3):201–207. doi:10.1093/sleep/17.3.201
51. Hsiou DA, Gao C, Matlock RC, Scullin MK. Validation of a nonwearable device in healthy adults with normal and short sleep durations. *J Clin Sleep Med*. 2022;18(3):751–757. doi:10.5664/jcsm.9700

## Nature and Science of Sleep

Dovepress

### Publish your work in this journal

Nature and Science of Sleep is an international, peer-reviewed, open access journal covering all aspects of sleep science and sleep medicine, including the neurophysiology and functions of sleep, the genetics of sleep, sleep and society, biological rhythms, dreaming, sleep disorders and therapy, and strategies to optimize healthy sleep. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/nature-and-science-of-sleep-journal>