

SCIENTIFIC INVESTIGATIONS

Performance of a Portable Sleep Monitoring Device in Individuals with High Versus Low Sleep Efficiency

Rachel R. Markwald, PhD¹; Sara C. Bessman, MS¹; Seth A. Reini, PhD¹; Sean P.A. Drummond, PhD^{2,3}

¹Warfighter Performance Department, Naval Health Research Center, San Diego, CA; ²Psychiatry Department, University of California, San Diego, CA; ³Psychology Service, VA San Diego Healthcare System, San Diego, CA

Study Objectives: Portable and automated sleep monitoring technology is becoming widely available to consumers, and one wireless system (WS) has recently surfaced as a research tool for sleep and sleep staging assessment outside the hospital/laboratory; however, previous research findings indicate low sensitivity for wakefulness detection. Because difficulty discriminating between wake and sleep is likely to affect staging performance, we sought to further evaluate the WS by comparing it to the gold-standard polysomnography (PSG) and actigraphy (ACT) for overall sleep/wakefulness detection and sleep staging, within high and low sleep efficiency sleepers.

Methods: Twenty-nine healthy adults (eight females) underwent concurrent WS, PSG, and ACT assessment in an overnight laboratory study. Epoch-by-epoch agreement was determined by comparing sleep/wakefulness decisions between the WS to both PSG and ACT, and for detection of light, deep, and rapid eye movement (REM) sleep stages between the WS and PSG.

Results: Sensitivity for wakefulness was low (40%), and an overestimation of total sleep time and underestimation of wake after sleep onset was observed. Prevalence and bias adjusted kappa statistic indicated moderate-to-high agreement between the WS and PSG for sleep staging. However, upon further inspection, WS performance varied by sleep efficiency, with the best performance during high sleep efficiency.

Conclusions: The benefit of the WS as a sleep monitoring device over ACT is the ability to assess sleep stages, and our findings suggest this benefit is only realized within high sleep efficiency. Care should be taken to collect data under conditions where this is expected.

Keywords: automated analyses, mobile sleep assessment, sleep monitoring, sleep staging

Citation: Markwald RR, Bessman SC, Reini SA, Drummond SP. Performance of a portable sleep monitoring device in individuals with high versus low sleep efficiency. *J Clin Sleep Med* 2016;12(1):95–103.

INTRODUCTION

Public awareness of the importance of sleep as a modifiable risk factor in health and disease and in human cognitive performance has increased. This has driven the development of sleep monitoring devices designed to provide sleep measures to personal consumers.¹ Historically, the gold standard for sleep assessment is the laboratory- or hospital-based polysomnogram (PSG), a costly overnight test that requires trained specialists to conduct and analyze the data. From a research standpoint, a high-performance and low-cost portable sleep monitoring device would allow investigation of sleep, particularly sleep staging, in environments where PSG is not practical (e.g., various home and military settings), not affordable (long-term tracking of sleep and sleep patterns), or both. To date, only one research-grade alternative has been available to measure sleep/wakefulness outside of the laboratory environment. Validated in several populations against the gold-standard PSG, actigraphy (ACT) utilizes accelerometer data to determine sleep/wakefulness state based on body movement.^{2,3} However, the device cannot provide sleep stage information, which limits its utility. One commercially developed wireless sleep system (WS) has shown promise in assessing sleep/wakefulness as well as collecting sleep staging information for research intent.

BRIEF SUMMARY

Current Knowledge/Study Rationale: A commercially developed wireless system for monitoring sleep has been utilized as the primary sleep assessment in sleep research studies; however, a previously identified weakness in device performance needs to be further explored. We conducted a study to evaluate the device for sleep/wakefulness identification and sleep staging in a sample of heterogeneous sleep efficiency sleepers.

Study Impact: These findings indicate that the wireless system is useful at sleep staging when high-efficiency sleep is expected and when traditional polysomnography is precluded due to cost or location restrictions. Importantly, although the wireless system performs well at dichotomous sleep/wake identification in low-efficiency sleepers, the sleep stage information becomes less reliable and does not offer any advantages over standard actigraphy for remote monitoring.

Four laboratory studies (one company sponsored) have thus far evaluated the performance of the WS versus PSG.^{4–7} In the company-sponsored study, the WS was additionally compared side-by-side with ACT performance in relation to PSG-determined sleep/wakefulness.⁶ One additional home-based study assessed performance of the WS versus ACT for sleep/wakefulness detection for 5 nights.⁸ The PSG validation studies found moderate to moderate-to-high sleep staging agreement between the WS and PSG; however, one

consistent finding was that WS sensitivity for PSG-defined wakefulness is low.

Further examination is thus necessary to fully characterize the utility of the device. This is especially warranted because the WS has begun to surface as the primary sleep measurement device in research studies.^{9–12} Here, we aimed to further evaluate the WS for overnight sleep/wakefulness and sleep staging performance by comparing it with PSG and with ACT, and, for the first time, evaluate sleep staging performance in sleepers with high ($\geq 85\%$) and low ($< 85\%$) sleep efficiency. We hypothesized that WS performance would be better under conditions of high sleep efficiency due to the previously identified low sensitivity to wakefulness detection.

MATERIALS AND METHODS

Participants

Data for this study were collected from the habituation night of a larger sleep protocol. Participants were healthy, with no mental health or sleep disorders. Exclusion criteria included self-reported personal or familial diagnosis of a Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) Axis I disorder, lifetime use of psychotropic medications, serious neurological and medical disorders, a loss of consciousness exceeding 2 min, sleep disorders, daily caffeine consumption > 300 mg, current drug use, body mass index exceeding 30, and erratic sleep–wake cycles. Inclusion criteria were (1) age: 18–39 y, (2) education ≥ 12 y, and (3) consistent sleep–wake schedule with 7–9 h of overnight sleep/night. Participants provided written, informed consent as outlined by the VA San Diego Healthcare System Institutional Review Board. Twenty-nine participants (eight female) aged 24.0 ± 5.3 y (mean \pm standard deviation) completed the study. Data were collected on 44 participants; however, 15 participants were removed due to the WS malfunctioning ($n = 10$) or the headband falling off ($n = 5$). Additionally, three participants were missing ACT event markers, which prevented proper alignment for sleep analysis.

Polysomnography

PSGs were obtained with Grass Comet Plus (Grass Technologies, Middleton, WI, USA) digital sleep recorders using monopolar electroencephalogram (EEG) channels referenced to contralateral mastoids according to the international 10–20 system (C3-A2, C4-A1, O1-A2, and F3-A2)¹³ right and left electro-oculograms (EOGs), chin electromyogram, and electrocardiogram. Impedances were < 5 k ohms. EEG data were digitized at a sampling rate of 200 Hz. Low- and high-pass digital filters for EEG and EOG were set at 0.3 Hz and 70 Hz, respectively. Sleep was scored from brain region C3-A2 in 30-sec epochs according to 2007 American Academy of Sleep Medicine criteria (AASM).¹⁴ All PSG recordings were scored by a single registered polysomnographic technologist with an intrarater reliability of 0.95. Scored epochs were identified as one of the following stages: rapid eye movement (REM), non-REM stages N1, N2, N3, and wake.

Actigraphy

The Actiwatch-64 actigraph (Philips Respironics, Bend, OR, USA) was configured to collect data in 30-sec epochs, with a medium sensitivity level. The automated Actiware software (version 3.3; Philips Respironics) scored sleep and wakefulness for the interval between down and rise time, corresponding to scheduled lights on and lights off.

Wireless System

The WS (Zeo, Inc., Newton, MA, USA) consisted of a headband containing three dry fabric leads that communicated wirelessly to a bedside base station at 2.4 GHz. The headband collected electrophysiological signals from the EEG, eye movements, and frontalis muscle at the location of the forehead. Real-time sleep stage information was determined by a microprocessor within the base station using proprietary algorithms based on Rechtschaffen and Kales scoring criteria. For more information on the WS technical specifications please see Shambroom et al.⁶ The WS categorized sleep into four stages (wake, light, deep, and REM) at 30-sec intervals.

Procedure

Participants slept overnight in the sleep laboratory at their habitual bed and wake times while undergoing PSG, WS, and ACT recordings. Prior to bedtime, a sleep technician performed a PSG hook-up, and then fitted the WS headband onto the forehead and the Actiwatch onto the wrist. All devices were time synchronized, and event markers and technical notes were used to align recordings for analysis of sleep measures from lights off until lights on. A sleep technician monitored the entire study session and was responsible for initiating and terminating each recording device at the scheduled times.

Data Analysis

Sleep Measures

The following sleep summary measures were calculated for each sleep recording device: total sleep time (TST); sleep onset latency (SOL), defined as the interval between lights off and the first epoch scored as non-wake; sleep efficiency (SE), the ratio between TST and total time in bed; wake after sleep onset (WASO), the interval between SOL and final awakening; and the number of awakenings (NA) > 2 min. For the sleep staging and agreement analysis between the WS and PSG, it was necessary to classify the PSG sleep stages as those defined by the WS. Therefore, PSG stages N1 and N2 were grouped together as light sleep. Sleep staging summary measures included time in wake (T_{Wake}), time in light (T_{Light}), time in deep (T_{Deep}), and time in REM (T_{REM}). Statistical analyses were performed using STATISTICA (version 12.0; Stata-Corp LP, College Station, Texas, USA). Data are expressed as mean \pm standard deviation.

Sleep Summary Outcomes: WS, ACT, and PSG

The measures TST, SOL, SE, WASO, and NA > 2 min were determined to have marginally non-normal distributions upon inspection of normal QQ plots and of a significant Shapiro–Wilk

Table 1—Sleep summary and staging measures.

Outcome Measures: All	PSG (n = 29)	WS (n = 29)	ACT (n = 26)
Total sleep time (min)	441.3 ± 39.3	461.5 ± 49.4	469.0 ± 28.9 ^a
Sleep onset latency (min)	14.7 ± 17.3	7.3 ± 9.5 ^a	2.4 ± 1.8 ^a
WASO (min)	48.3 ± 37.9	26.7 ± 26.8 ^{a,b}	33.8 ± 20.4
# awakenings ≥ 2min	4.0 ± 3.3	4.0 ± 3.2	4.4 ± 3.7
Sleep efficiency (%)	87.1 ± 7.2	90.9 ± 7.8	92.6 ± 4.3 ^a
Sleep stage (min)			
T _{Wake}	66.2 ± 38.5	34.5 ± 29.1 ^a	—
T _{Light}	263.5 ± 41.3	245.1 ± 62.0	—
T _{Deep}	81.1 ± 31.7	97.1 ± 47.5	—
T _{REM}	96.6 ± 26.9	119.2 ± 42.2 ^a	—

Sleep summary outcome measures for polysomnogram (PSG), wireless system (WS) and actigraphy (ACT) and sleep staging for PSG and WS. ^aSignificant difference from PSG ($p < 0.05$). ^bSignificant difference from ACT ($p < 0.05$).

W test result. The results of a one-way analysis of variance (ANOVA) were compared with the results from the nonparametric Kruskal–Wallis H test. Because the results did not differ and the conclusions drawn were the same, we chose to report the ANOVA results only. When the model was significant, Fisher least significant difference tests, combined with Bonferroni correction for multiple comparisons, were used to examine group differences.

Sleep Staging: WS and PSG

For the WS versus PSG sleep stage summary comparison, non-normally distributed metrics (T_{Deep} , T_{Wake}) were assessed using the nonparametric Wilcoxon matched-pairs rank-sum tests and results were compared with paired t -tests, with no differences in conclusions. Therefore, the results from the paired t -tests are reported for all these measures.

Epoch-by-Epoch Agreement

To determine agreement between the devices with PSG for sleep/wakefulness decisions, a 30-sec epoch-by-epoch analysis was conducted to determine accuracy, sensitivity, specificity, and positive predictive value (PPV) using the definitions set forth by Tilmanne et al.¹⁵ A more detailed description of the calculations involved with this procedure is available elsewhere.¹⁶ Briefly, a confusion matrix is generated accounting for true positives, false positives, true negatives, and false negatives. For comparison purposes, binary scores were produced for both the PSG and WS (wakefulness = 0, all sleep stages = 1). For the analysis between the WS and PSG for sleep staging decisions, epoch-by-epoch agreement was determined for light, deep, and REM sleep stages.

Traditionally the kappa statistic has been reported in ACT and WS validation studies for determining agreement. However, we chose to calculate the prevalence and bias-adjusted kappa (PABAK) because the high proportion of sleep that results in unequal numbers of sleep and wake epochs makes kappa less reliable for overnight sleep measurement.¹⁷ The PABAK gives balanced weight to sleep and wake epochs.¹⁸ The standard definitions were adopted (0–0.2 = little or no agreement, 0.2–0.4 = low, 0.4–0.6 = moderate, 0.6–0.8 = high,

and 0.8–1.0 = near-perfect agreement).¹⁹ Finally, Bland–Altman plots for TST and WASO were generated in order to visualize the absolute agreement between the sleep devices and because of known inadequacies using correlation and t -test analysis to evaluate agreement between two measurement methods.²⁰

SE Subgroup Analysis

Data from the current study were derived from the habituation night of a larger protocol and thus comprised a wide range of SE values. The WS, ACT, and PSG records were divided into low < 85% ($n = 13$) and high ≥ 85% ($n = 16$) SE subgroups as determined by the PSG. Criteria were chosen based on the commonly used clinical and research cutoff of 85% for identifying insomnia, although it is noted some clinicians and researchers use a cutoff of 90%.²¹ Because the two SE groups did not result in any further non-normality in sleep summary measures, and results from the three-way sleep device comparison were identical whether the parametric ANOVA or nonparametric alternative were used, we chose to run a two-way ANOVA to determine differences for the factors SE group (low, high) and sleep device (PSG, ACT, WS). PSG and WS SE group means for sleep staging measures (T_{Wake} , T_{Light} , T_{REM} , and T_{Deep}) were significant for Levene's test for equality of variance. Therefore, we ran the non-parametric Mann–Whitney U test to examine SE group differences in sleep staging measures and the Wilcoxon signed-rank test for paired comparisons between the WS and PSG. Finally, the 30-sec epoch-by-epoch analysis for sleep/wakefulness and sleep staging decisions was repeated for the SE groups. Cohen's kappa was computed to examine agreement within each SE group.

RESULTS

Sleep Summary Outcomes: WS, ACT, and PSG

Statistical differences were found among the sleep devices for TST, $F(2,81) = 3.518$, $p = 0.034$; WASO, $F(2,81) = 4.009$, $p = 0.02$; SOL, $F(2,81) = 7.739$, $p = 0.001$; SE, $F(2,81) = 5.019$, $p = 0.009$. For results of the *post hoc* analyses, see **Table 1**.

Table 2—Epoch-by-epoch agreement analysis.

All Subjects	Accuracy (%)	Specificity (%)	Sensitivity (%)	PPV (%)	PABAK
WS vs. PSG (n = 29)					
Overall wake	90.7	98.2	40.4	77.6	0.81
Overall sleep	89.5	44.6	96.3	92.0	0.79
Stage light	73.1	75.8	70.6	76.0	0.46
Stage deep	86.4	90.0	67.3	56.0	0.73
Stage REM	85.8	88.5	74.5	60.0	0.72
WS vs. ACT (n = 26)					
Overall wake	90.7	95.7	28.6	35.2	0.81
Overall sleep	89.5	32.2	94.2	95.0	0.79
ACT vs. PSG (n = 26)					
Overall wake	89.3	96.7	37.0	61.1	0.79
Overall sleep	89.3	37.0	96.7	92.0	0.79
SE < 85%	Accuracy (%)	Specificity (%)	Sensitivity (%)	PPV (%)	PABAK
WS vs. PSG (n = 29)					
Overall wake	87.4	98.0	43.3	88.0	0.75
Overall sleep	86.0	46.7	95.6	88.1	0.72
Stage light	71.3	78.8	63.4	74.0	0.43
Stage deep	83.4	86.7	63.4	45.0	0.67
Stage REM	84.4	85.9	77.5	53.0	0.69
WS vs. ACT (n = 26)					
Overall wake	88.5	93.5	37.8	36.3	0.77
Overall sleep	87.6	40.1	92.3	94.0	0.75
ACT vs. PSG (n = 26)					
Overall wake	85.6	97.1	35.5	73.4	0.71
Overall sleep	85.6	35.5	97.1	87.0	0.71
SE ≥ 85%	Accuracy (%)	Specificity (%)	Sensitivity (%)	PPV (%)	PABAK
WS vs. PSG (n = 29)					
Overall wake	93.5	98.4	34.0	63.9	0.87
Overall sleep	92.5	40.1	96.8	95.1	0.85
Stage light	74.6	72.9	76.0	77.0	0.49
Stage deep	89.0	92.9	70.1	67.0	0.78
Stage REM	87.0	90.8	72.5	67.0	0.74
WS vs. ACT (n = 26)					
Overall wake	92.5	97.4	18.5	32.9	0.85
Overall sleep	91.0	23.7	95.6	94.9	0.82
ACT vs. PSG (n = 26)					
Overall wake	92.2	96.4	39.9	47.6	0.84
Overall sleep	92.2	39.9	96.4	95.2	0.84

Agreement analysis examined overall sleep (dichotomous sleep/wake scoring), and each sleep stage (versus all other stages) for all subjects and split into low < 85% sleep efficiency (SE) and high SE ≥ 85%. Prevalence and bias-adjusted kappa (PABAK) is the prevalence and bias adjusted kappa and is weighted for the total number of epochs. ACT, actigraphy; PPV, positive predictive value; PSG, polysomnogram; REM, rapid eye movement; WS, wireless system.

Sleep Staging: WS and PSG

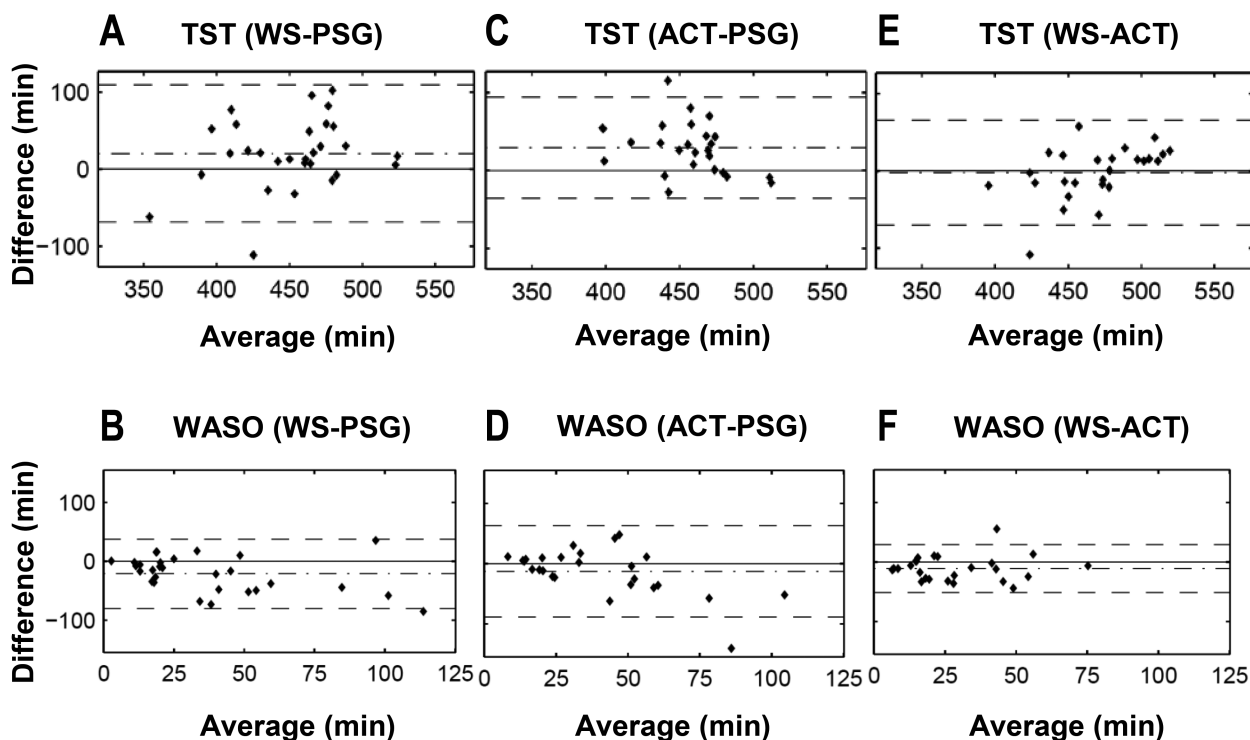
Paired *t*-tests showed significant differences in T_{Wake} ($t_{28} = 6.895$; $p < 0.001$) and T_{REM} ($t_{28} = -2.599$; $p < 0.05$), and a nonsignificant trend was identified for T_{Light} ($t_{28} = 6.895$; $p < 0.1$). Sleep staging data are reported in **Table 1**.

Epoch-by-Epoch Agreement

All accuracy, sensitivity, specificity, PPV, and PABAK scores are reported in **Table 2** (All subjects) for the entire dataset. The

overall agreement with PSG (all 29,435 epochs from all subjects) was 66.9% ($\kappa = 0.50$). **Figure 1** shows the Bland–Altman plots for further visualization of the agreement between the WS and PSG, ACT and PSG, and the WS and ACT. The WS overestimated TST (**Figure 1A**) and underestimated WASO (**Figure 1B**) when compared with PSG. ACT also overestimated TST (**Figure 1C**) and, to a lesser extent, underestimated WASO (**Figure 1D**) relative to PSG. There was little difference in TST estimation between the WS and ACT (**Figure 1E**).

Figure 1—Bland-Altman plots for total sleep time (TST) and wake after sleep onset (WASO).



Bland-Altman plots for TST and WASO determined by WS versus PSG (A,B); ACT versus PSG (C,D); and WS versus ACT (E,F); solid line represents difference of zero; dashed-dotted line represents mean difference; dashed line represent ± 2 standard deviations (SD) from the mean; bias represents average difference where positive bias indicates overestimation and negative bias indicates underestimation. (A) Bias 20.2, SD 33.0. (B) Bias -21.6, SD 22.7. (C) Bias 25.3, SD 31.4. (D) -10.2, SD 30.0. (E) Bias -2.3, SD 34.5. (F) Bias -11.0, SD 20.7. ACT, actigraphy; PSG, polysomnogram; WS, wireless system.

However, when compared with ACT, the WS underestimated WASO (Figure 1F) as was expected because ACT was closer to PSG than the WS in estimating this measure.

SE Group Analysis

For the SE group analysis, sleep summary measures are reported in Table 3. There was a significant interaction between SE group and sleep device for TST, $F(2,78) = 3.1440$, $p = 0.0486$; WASO, $F(2,78) = 5.134$, $p = 0.008$; and SE, $F(2,78) = 4.970$, $p = 0.009$. There were significant differences among the devices within the low SE group for TST, $F(2,78) = 6.725$, $p = 0.002$; WASO, $F(2,78) = 10.041$, $p < 0.001$; and SE, $F(2,78) = 6.725$, $p = 0.002$. No significant differences were found among the three sleep devices in the high SE group for any measure. For results of *post hoc* tests, see Table 3. There was a main effect of sleep device for SOL in the low SE group, $F(2,78) = 8.918$, $p < 0.001$, with significant differences between PSG and both ACT and WS (all $p \leq 0.05$). We found the expected between-group differences for sleep staging measures (Table 3). The Wilcoxon signed-rank test indicated significant differences between the WS and PSG in T_{Wake} , T_{Light} , and T_{REM} in the low SE group, and T_{Wake} in the high SE group (all $p < 0.05$).

Results from the SE group epoch-by-epoch analysis are reported in Table 2. Overall, WS sleep staging improved in the

high SE group relative to the low SE group; however, sensitivity and PPV with PSG for wakefulness was lower despite PABAK-determined near-perfect agreement. For further visualization of WS performance based on SE, Figure 2 shows two representative subjects for a PSG-defined high and low SE sleeper. The associated contingency tables represent the agreement in sleep/wakefulness and sleep staging decisions between PSG and the WS.

DISCUSSION

The results of this validation study indicate that the WS has limitations in wakefulness detection that are further exacerbated in low SE sleep. Further, WS sleep staging performance varies by SE such that the best performance, relative to the gold-standard PSG, was observed in high SE sleep only. The WS performed moderately well for determining sleep summary and sleep staging measures in our sample of heterogeneous sleep efficiencies. However, upon more detailed inspection, the important novel finding here is that WS performance was worse under conditions of low SE sleep.

Overall, our findings for sleep/wakefulness detection were in line with prior reports indicating that although sensitivity

Table 3—Sleep summary measures and staging by sleep efficiency group.

Outcome Measures: SE < 85	PSG (n = 13)	WS (n = 13)	ACT (n = 11)
Total sleep time (min)	416.8 ± 33.5	452.0 ± 56.1	473.7 ± 28.8 ^a
Sleep onset latency (min)	21.0 ± 23.4	7.3 ± 10.9 ^a	2.2 ± 1.4 ^a
WASO (min)	78.4 ± 36.9	42.9 ± 30.3 ^a	42.3 ± 22.4 ^a
# awakenings ≥ 2 min	6.4 ± 3.2	5.7 ± 3.4	5.9 ± 3.8
Sleep efficiency (%)	80.5 ± 4.9	87.2 ± 8.2 ^a	91.1 ± 4.7 ^a
Sleep stage (min)			
T _{Wake}	100.9 ± 27.7	52.0 ± 31.8 ^a	–
T _{Light}	252.8 ± 24.1	216.4 ± 47.9 ^a	–
T _{Deep}	75.1 ± 28.1	106.2 ± 58.8	–
T _{REM}	88.9 ± 18.3	129.4 ± 32.2 ^a	–
Outcome Measures: SE ≥ 85	PSG (n = 16)	WS (n = 16)	ACT (n = 11)
Total sleep time (min)	461.2 ± 32.3	469.1 ± 43.5	465.6 ± 29.4
Sleep onset latency (min)	9.6 ± 7.6	7.4 ± 8.6	2.6 ± 2.1
WASO (min)	23.9 ± 13.0	13.6 ± 14.3	27.5 ± 16.9
# awakenings ≥ 2 min	2.1 ± 1.8	2.6 ± 2.2	3.3 ± 3.4
Sleep efficiency (%)	92.4 ± 3.2	93.9 ± 6.2	93.7 ± 3.7
Sleep stage (min)			
T _{Wake}	38.0 ± 16.1	20.2 ± 16.9 ^a	–
T _{Light}	272.2 ± 50.4	268.5 ± 63.5	–
T _{Deep}	86.1 ± 34.3	89.7 ± 36.3	–
T _{REM}	102.9 ± 31.5	111.0 ± 48.4	–

Sleep summary outcome measures for polysomnogram (PSG), wireless system (WS) and actigraphy (ACT) and sleep staging for PSG and WS for each sleep efficiency (SE) group. ^aSignificant difference from PSG ($p < 0.05$).

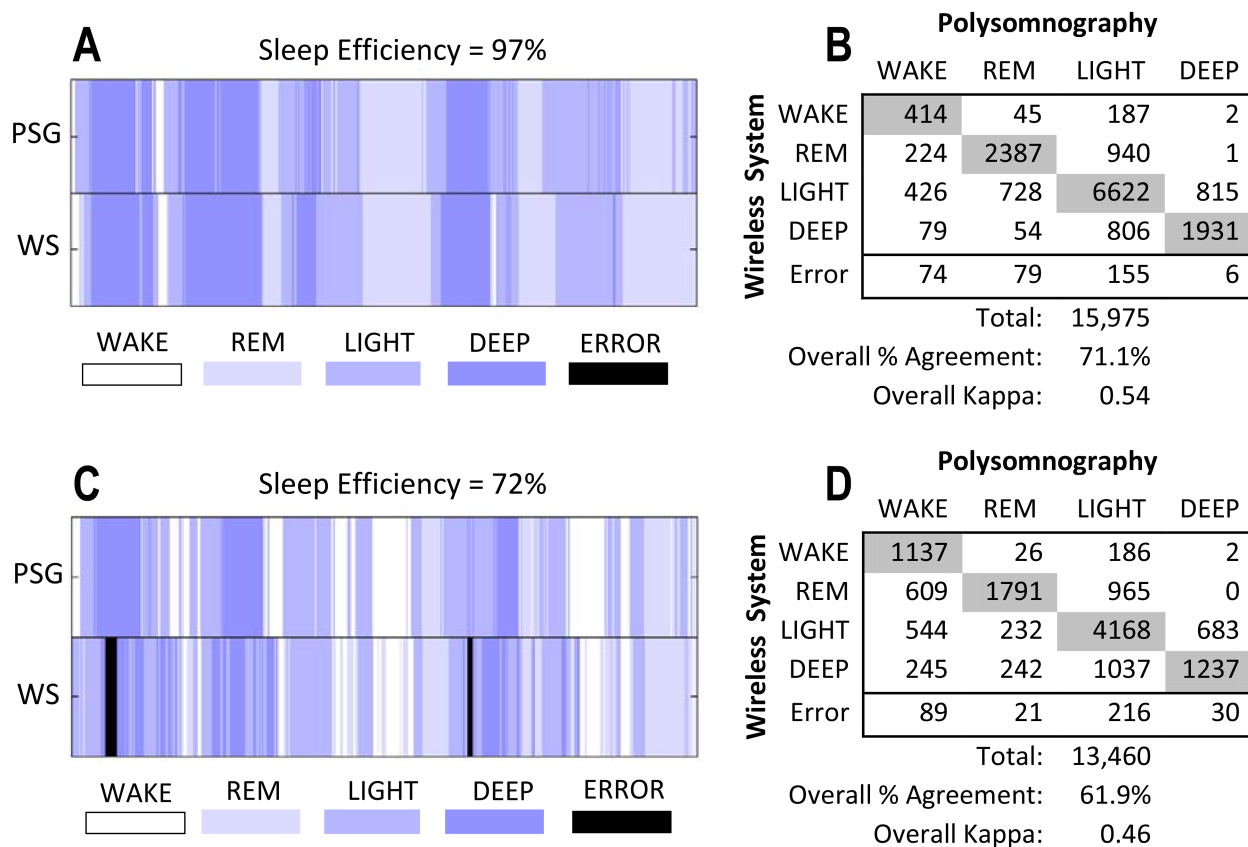
for sleep is high (96.3%), specificity for sleep is low (44.6%), demonstrating that the WS is not proficient in discriminating between PSG-determined sleep and wakefulness due to misidentification of epochs of wakefulness as sleep.

Not surprisingly, when compared with PSG, we found that the WS underestimated WASO and overestimated TST (see **Table 1**). For sleep-staging performance, the WS was previously shown to have high agreement with PSG scorers (all > 0.60) in the company-sponsored study by Shambroom et al.⁶ Cohen's kappa agreement determined from that study was higher than that observed in our whole sample analysis (0.46–0.58), but was closer to that of the high SE group (0.49–0.62) (see **Table 2**). We also had the same findings as prior studies in which the WS overscored REM^{5,6} and underestimated the amount of deep sleep.⁷ However, in our sample, the effect for deep sleep was not statistically significant, which may be attributed to the inclusion of AASM scoring criteria since these findings were similar to those of Cellini et al.⁴

Because of the previously identified low sensitivity to wakefulness, we chose to further examine WS performance during high and low SE sleep because SE is often used to represent sleep quality and low SE contains more epochs of wakefulness. Additionally, we wanted to compare the performance of the WS with that of ACT because it has been reported that concordance of ACT with PSG is lower in nights with lower SE.^{2,22} When the study sample was divided by SE, we found that performance on sleep summary measures TST, WASO,

and SE varied most likely because WS performance for wakefulness detection was further worsened by low SE. This was observed by the lower PPV and sensitivity for wakefulness in high SE sleep compared with low SE sleep (see **Table 2**). Further, when compared with PSG, WS-determined WASO, SOL, and SE differed in the low SE group only. Importantly, sleep stage identification by the WS improved in the high relative to low SE group, as indicated by an increase in PABAK values (from 0.43–0.69 to 0.49–0.78). For these reasons, the WS may not be as reliable for sleep staging under low SE conditions.

When examining the whole range of SE sleep, performance of the ACT versus WS as compared with PSG was mixed. For example, ACT performed better than the WS for WASO; however, the WS performed better at estimating TST and SE. When broken down by SE, we observed no significant differences between both ACT and WS, with PSG indicating that both devices performed similarly well during high SE sleep (≥ 85%). During low SE sleep (< 85%), the performance of both devices was worse for estimating (WASO, SOL, and SE) as evidenced by significant differences from PSG; however, the WS performed better than ACT for estimating TST. This was likely a result of the device's improved ability to identify epochs of wakefulness during low SE sleep (sensitivity% 43.4; PPV% 88.0) as compared to the ACT (sensitivity% 35.5; PPV% 73.4); see **Table 2**. These results indicate that both devices perform well with high SE sleep; however, if low SE sleep is expected the WS may offer a slight advantage over ACT for assessing

Figure 2—Representative subjects and associated contingency tables for sleep staging.

Representation of the wireless system (WS) and polysomnogram (PSG) for a high (A) and low (C) PSG-defined sleep efficiency (SE) sleeper. Associated contingency tables for the high (B) and low (D) SE sleeper represent full night comparisons among PSG and WS for sleep/wake and sleep staging decisions for every 30-sec epoch; diagonal represents agreement between WS and PSG; error defined as WS score of zero which is an undefined measurement by WS; percent agreement for each SE sleeper and corresponding Cohen's kappa statistic are included for comparison.

sleep summary statistics, especially when the emphasis is on sleep duration.

This study and the WS technology had some limitations. First, the WS was compared with PSG records that were scored by a single expert sleep scorer rather than to a consensus score, as was done in the study by Shambroom and colleagues.⁶ In most clinical and research settings however, only one sleep scoring expert will typically score a PSG record. It is possible that if we had included another expert scorer we may have found that agreement slightly varied between the WS and PSG. For the SE group analysis, small group sample size, specifically in the < 85% group, may have led to insufficient power to detect significant differences in some of the sleep summary and staging measures; thus, more research will need to be conducted to further evaluate WS performance under these conditions. We selected a medium sensitivity algorithm for ACT analysis based on our inclusion of a healthy sleeping population. However, ACT performance may have improved if a higher sensitivity threshold had been selected as has been shown in insomnia populations²³ and daytime naps models.¹⁶ Prescreening using a sleep diary and/or validated questionnaires to assess sleep disturbance may

be useful in determining if the population of interest is likely to have high SE and thus appropriate for assessment with the WS or similar technology.

The ~34% data loss that occurred in our study due to WS technical problems should be mentioned. In five participants the headband came off during recording and this was likely because the headband was not fitted tightly to avoid headaches or other participant discomfort. In 10 participants, the WS malfunctioned, leading to a failure to record data. Unfortunately, this amount of data loss, either due to the headband falling off or WS malfunction (15/44 participants ~ 34% of sample), has occurred in other studies utilizing the WS. Technical difficulties with the WS contributed to the ~32% data loss observed in the study by Tonetti et al. 2013⁷ and ~42% in the study by Griessenberger et al. 2013.⁵ Clearly, this is a disadvantage of this system and a factor to consider during the sleep assessment selection process. Additionally, the location of recording of the WS likely affected performance. PSG collects brain electrical information from several sites on the scalp, not just the frontal locations, as well as the eyes and muscular activity of the chin using electrodes placed beside each eye and on the mentalis muscle. This information is then

utilized by expert scorers to determine sleep stages based off established criteria during visual inspection of the records.¹⁴ The WS proprietary algorithms attempt to assess and categorize sleep stages utilizing brain activity, eye movements, and frontalis muscle activity from a frontal location site on the scalp only. Alpha activity measured from the forehead is low and therefore may contribute to the WS's difficulty detecting quiet wakefulness as has been suggested before.⁷ Although the WS in this current form is no longer manufactured, future technology may benefit from adding specific sensors to more precisely collect the additional eye movement and muscle activity information.

Together with findings from previous validation studies, the current findings suggest that the WS has strengths and weaknesses that need to be fully considered in the context of each research setting. The real benefit of using the WS over ACT is the ability to access sleep staging measures, which is better recognized in high SE conditions. Due to the larger problem of overscoring sleep, sleep staging measures may be less reliable in low SE conditions (e.g., insomnia, daytime sleep), and for this reason, effort should be taken to collect only high SE data to improve accuracy. Therefore, the WS may be considered a useful tool for assessing sleep staging in healthy sleeping individuals, under environmental conditions that would support high SE sleep, when PSG examination is precluded due to cost or time restrictions, or population-specific limitations (e.g., certain military settings). If the primary research objective is not sleep staging, but is instead sleep/wakefulness determination, then there is little benefit to using the WS over the standard ACT because device performance was comparable. Further studies examining the performance of the WS in low SE conditions will be necessary to determine at what point the limitations in the WS algorithm outweigh the potential usefulness of the device.

ABBREVIATIONS

AASM, American Academy of Sleep Medicine

ACT, actigraphy

ANOVA, analysis of variance

EEG, electroencephalogram

EMG, electromyogram

EOG, electro-oculograms

NA, number of Awakenings

PABAK, prevalence and bias adjusted kappa

PSG, polysomnography

REM, rapid eye movement sleep

SE, sleep efficiency

SOL, sleep onset latency

TIB, time in bed

T_{Deep}, time in deep sleep

T_{Light}, time in light sleep

TST, total sleep time

T_{Wake}, time in Wake

VA, Veteran's Affairs

WASO, wakefulness after sleep onset

WS, wireless system

REFERENCES

- Kelly JM, Strecker RE, Bianchi MT. Recent developments in home sleep-monitoring devices. *ISRN Neurol* 2012;2012:768794.
- Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. *Sleep* 2003;26:342–92.
- Pollak CP, Tryon WW, Nagaraja H, Dzwonczyk R. How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep* 2001;24:957–65.
- Cellini N, McDevitt EA, Ricker AA, Rowe KM, Mednick SC. Validation of an automated wireless system for sleep monitoring during daytime naps. *Behav Sleep Med* 2015;13:157–68.
- Griessenberger H, Heib DP, Kunz AB, Hoedlmoser K, Schabus M. Assessment of a wireless headband for automatic sleep scoring. *Sleep Breath* 2013;17:747–52.
- Shambroom JR, Fabregas SE, Johnstone J. Validation of an automated wireless system to monitor sleep in healthy adults. *J Sleep Res* 2012;21:221–30.
- Tonetti L, Cellini N, de Zambotti M, et al. Polysomnographic validation of a wireless dry headband technology for sleep monitoring in healthy young adults. *Physiol Behav* 2013;118C:185–8.
- Tonetti L, Fabregas SE, Fabbri M, et al. Comparison of a wireless dry headband technology for sleep monitoring with actigraphy in healthy adults. *Biol Rhythm Res* 2013;44:333–8.
- Gumenyuk V, Roth T, Korzyukov O, Jefferson C, Bowyer S, Drake CL. Habitual short sleep impacts frontal switch mechanism in attention to novelty. *Sleep* 2011;34:1659–70.
- Kudesia RS, Bianchi MT. Decreased nocturnal awakenings in young adults performing bikram yoga: a low-constraint home sleep monitoring study. *ISRN Neurol* 2012;2012:153745.
- Marshall JC, Malerba JR, Schroeder JA. Use of personal EEG monitors in a behavioral neuroscience course to investigate natural setting sleep patterns and the factors affecting them in college students. *J Undergrad Neurosci Educ* 2011;10:A65–70.
- Scullin MK. Sleep, memory, and aging: the link between slow-wave sleep and episodic memory changes from younger to older adults. *Psychol Aging* 2013;28:105–14.
- Jasper HH, Radmussen T. Studies of clinical and electrical responses to deep temporal stimulation in men with some considerations of functional anatomy. *Assoc Res Nerv Ment Dis* 1958;36:316–34.
- Iber C, Ancoli-Israel S, Chesson A, Quan SF. The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications. 1st ed. Westchester, IL: American Academy of Sleep Medicine, 2007.
- Tilmanne J, Urbain J, Kothare MV, Wouwer AV, Kothare SV. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. *J Sleep Res* 2009;18:85–98.
- Kanady JC, Drummond SP, Mednick SC. Actigraphic assessment of a polysomnographic-recorded nap: a validation study. *J Sleep Res* 2011;20:214–22.
- Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol* 1990;43:543–9.
- Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46:423–9.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- Altman DG, Bland JM. Measurement in medicine - the analysis of method comparison studies. *Statistician* 1983;32:307–17.
- Perlis ML, Swinkels CM, Gehrman PR, Pigeon WR, Matteson-Rusby SE, Jungquist CR. The incidence and temporal patterning of insomnia: a pilot study. *J Sleep Res* 2010;19:31–5.
- 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;2:389–96.
- Lichstein KL, Stone KC, Donaldson J et al. Actigraphy validation with insomnia. *Sleep* 2006;29:232–9.

ACKNOWLEDGMENTS

The authors thank the VA San Diego Healthcare System Center for Excellence in Stress and Mental Health and the Defense Medical Research and Development Program for funding this study and Marc Taylor for careful review of the manuscript.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication February, 2015

Submitted in final revised form July, 2015

Accepted for publication July, 2015

Address correspondence to: Rachel R. Markwald, PhD, Warfighter Performance Department, Naval Health Research Center, 140 Sylvester Road, San Diego, CA 92147; Tel: (619) 553-6380; Fax: (619) 767-4494; Email: rachel.r.markwald.civ@mail.mil

DISCLOSURE STATEMENT

This was not an industry supported study. Support was provided by the VA San Diego Healthcare System Center for Excellence in Stress and Mental Health Work Unit No. 1024 and Defense Medical Research and Development Program. The authors have indicated no financial conflicts of interest. The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, or the US Government. Approved for public release; distribution is unlimited. US Government Work (17 USC 105). Not copyrighted in the US. This research has been conducted in compliance with all applicable federal regulations governing the protection of human subjects in research (Protocol VASDHS.101303).