Comparison of Motionlogger Watch and Actiwatch actigraphs to polysomnography for sleep/wake estimation in healthy young adults

Tracy L. Rupp · Thomas J. Balkin

Abstract Sleep/wake identification and sleep parameter estimates from Motionlogger Watch and Actiwatch-64 actigraphs were compared to polysomnography (PSG). Following one night of baseline sleep, 29 volunteers remained awake for 36 h, followed by 11 h of recovery sleep in the laboratory. Two sets of analyses were performed: (1) epoch-by-epoch agreement and discriminability index (d') calculations, and (2) sleep parameter concordance with repeated measures ANOVAs. Sensitivity (sleep identification), specificity (wake detection), and overall agreement with PSG, as well as d', were higher for the Motionlogger than for Actiwatch. Relative to PSG, the Actiwatch-estimated total sleep time and sleep efficiency were underestimated and the number of awakenings was overestimated for baseline and recovery; sleep latency was underestimated on the baseline night. On the other hand, the Motionlogger-estimated total sleep time and sleep efficiency estimates were underestimated, and the sleep latency was overestimated on recovery, versus PSG. Despite these misestimations, it was concluded that the Motionlogger provided nominally better agreement with PSG, and that actigraphy generally constitutes a reasonably reliable tool for producing objective measurements of sleep/wake, but that users should remain mindful of its limitations.

Keywords Actigraphy · Polysomnography · Sleep deprivation

Wrist actigraphy provides an objective measure of sleep/wake behavior in a naturalistic setting (i.e., at home or in an operational field environment). Generally, actigraph devices record movement using accelerometers (movement detectors), sampling several times per second. Activity (or inactivity) data are estimated as sleep or wake for each "epoch" (a time period generally defined at 1 min, in the case of actigraphy) such that inactivity is associated with sleep and activity is associated with wake (the thresholds vary depending on hardware and software settings). Actigraphy has been used as an alternative to polysomnography (PSG), the gold standard for sleep/wake identification, due to its comparative convenience and cost-effectiveness. Briefly, PSG uses electroencephalography (EEG) to record brain activity using scalp electrodes. PSG tracings can be used to characterize and quantify sleep characteristics (e.g., sleep onset latency, number of awakenings) and sleep stages (e.g., wake, Stages 1 and 2, and slow-wave sleep).

Although validation studies have legitimized the use of actigraphy (e.g., Mullaney, Kripke, & Messin, 1980), few studies have directly compared the different commercially available actigraphs with PSG. Actigraphs vary in both hardware (e.g., sensitivity and specifications of the accelerometer) and software (e.g., definitions of sleep measures). Actigraph units of similar design are often assumed to yield similar data, and are thus used interchangeably, though no standard has been defined. Comparisons of different devices used simultaneously are desirable in order to inform decisions regarding actigraph use and interpretation.

The Basic Mini-Motionlogger (Ambulatory Monitoring, Ardsley, NY) and the Actiwatch L (Mini-Mitter, Bend, OR) are two commonly used actigraph devices. One direct comparison of the devices for two nights (worn simultaneously on the same arm), showed that the devices performed similarly overall when the Actiwatch was set to medium sensitivity (wake sensitivity at 40 activity counts per epoch), with no mean differences between devices

T. L. Rupp (✉) · T. J. Balkin
Behavioral Biology Branch, Center for Military Psychiatry
and Neuroscience, Walter Reed Army Institute of Research,
503 Robert Grant Avenue,
Silver Spring, MD 20910, USA
e-mail: tracy.rupp@amedd.army.mil

evident for the sleep measures (Benson et al., 2004). A more recent comparison of the same devices (Basic Mini-Motionlogger and Actiwatch) worn simultaneously for seven nights in the laboratory with PSG revealed that sleep parameter estimates from the actigraphs were similar to each other, but sleep latency was underestimated by both relative to PSG (Tonetti, Pasquini, Fabbri, Belluzzi, & Natale, 2008). Based on these findings, it was concluded that both devices were reliable and valid tools to evaluate sleep parameters (except for sleep onset latency) in healthy individuals.

In some previous reports, comparisons were made by correlating actigraph and PSG sleep outcome variables (e.g., Benson et al., 2004). Doing so, especially if most of the data are collected during the sleep period, may overestimate agreement, because correlations speak to the strength of the relationship between two variables (which would expected to be quite high in this case), but not the agreement between them (discussed in Bland & Altman, 1986). For example, the correlation between actigraphy and PSG total sleep time could be high but the epoch-by-epoch agreement low. Therefore, approaches in which epoch-by-epoch comparisons are made across the entire sleep/wake cycle are preferable for assessments of sensitivity, specificity, and overall agreement (e.g., Tryon, 1991). In some recent studies in which PSG and actigraph comparisons were made, both the correspondence of sleep parameters and the epoch-by-epoch agreement have been determined (Paquet, Kawinska, & Carrier, 2007), but to date such comparisons have not been made using two different commercially available actigraphs.

A new model of the Motionlogger actigraph has recently been introduced, the Motionlogger Watch (MW; Ambulatory Monitoring, Ardsley, NY), which includes wireless single sensor units that allow multiparameter data collection to be downloaded with a common interface. To date, the new device has not been directly compared to the Actiwatch (nor to any other actigraph), so the comparability of their activity measurements is currently unknown, and it is not yet clear whether the new device confers any advantages.

The objectives of this study were to directly compare the MW to the Actiwatch-64 (AW; Mini Mitter, Bend, OR) and to compare both to polysomnography (the current objective "gold standard" for recording sleep/wake) on baseline and recovery sleep nights in the laboratory (as part of a larger study). Epoch-by-epoch agreement analyses (for dichotomous assessment of wake vs. sleep) and sleep parameter concordance analyses (for assessment of continuous variables across the night [e.g., TST]) were performed in order to provide a more comprehensive comparison than in previous studies, in which only one comparison method was utilized.

Method

Subjects

Civilian and active-duty military men and women 18–39 years of age were recruited via flyers posted at local colleges, universities, and military installations as part of a larger study on the effects of personality and social experience on performance during sleep loss (Rupp, Killgore, & Balkin, 2010). After providing informed consent, subjects completed questionnaires to determine their eligibility on the basis of physical state, psychological state, sleep habits, and chronotype. Exclusion criteria included the following: habitual daytime napping; average nighttime lights-out times earlier than 21:00 Sunday through Thursday; average morning wake-up times later than 9:00 AM Monday through Friday; travel across more than three time zones within the last month; cardiovascular disease; hypertension; resting pulse greater than 95 beats per minute; past or present neurologic, psychiatric, or sleep disorder; present or past use of over-the-counter substances with purported psychoactive properties; asthma or other reactive airway diseases; prior history of cancer; allergies; regular nicotine use (or addiction) within the last 1 year; current heavy alcohol use; current use of illicit drugs, liver disease, or liver abnormalities; self-reported history of high daily caffeine use; anxiety (Spielberger & Vagg, 1984); depression (Beck & Steer, 1993; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961); extreme morning or evening preference (Horne & Ostberg, 1976); and current pregnancy. From the initial 470 volunteers responding to study recruitment flyers, 356 screened for the study, 56 enrolled, and 48 subjects completed the larger study.

Testing facilities

During testing and sleep periods, each subject was housed individually in a private sound-attenuated 8 × 10 foot room that included a bed and a computer workstation. The ambient temperature was approximately 23°C, and lighting was approximately 500 lux (with lights off during sleep periods). Background white noise was 60 dB at all times. When not engaged in testing or sleep, subjects remained in a common living area to play games, eat, read, or watch television and movies. The subjects were monitored continuously by at least one laboratory technician.

Procedure

As depicted in Fig. 1, the volunteers obtained one night of baseline sleep of 8 h time in bed (TIB) from 23:00 to 07:00, and then remained awake for a total of 36 h. Volunteers

were given 11 h TIB from 20:00 to 07:00 for recovery sleep immediately following the 36 h awake. Volunteers remained in the laboratory for the entire duration of the study.

Measures

Actigraphy Wrist movement and activity was recorded simultaneously (on the same, nondominant wrist) using both the MW and the AW during baseline and recovery sleep periods. MW data were collected in 30-s epochs using the "zero-crossing mode" with otherwise default settings. The 30-s epoch length was selected to be consistent with standard PSG scoring using 30-s epochs. The MW data were scored automatically for sleep/wake using Action-W Version 2, software using the Cole–Kripke algorithm (Cole & Kripke, 1988; Cole, Kripke, Gruen, Mullaney, & Gillan, 1992). AW data were similarly collected in 30-s epochs and scored automatically for sleep/wake using Actiware-Sleep, Version 3.4 (Mini Mitter, Bend, OR). All AW and MW data scored for sleep/wake were exported to Excel in an epoch-by-epoch format for analyses.

Polysomnography The PSG measurements included electroencephalogram (C3 and C4), electrooculogram (outer canthus of each eye), and electromyogram (mental/submental). Contralateral mastoid leads served as references for all unipolar measurements (electroencephalography and electrooculography). The PSG data were scored by a trained research technician and a 30-s epoch length was used, in accordance with Rechtschaffen and Kales's criteria (Rechtschaffen & Kales, 1968), and were displayed with Alice 4 Sleepware software (Respironics, Murraysville, PA). The dependent measures for nighttime sleep periods (defined as lights out to lights on) included minutes of the individual stages (wake, Stage 1, Stage 2, slow-wave sleep,

and rapid eye movement) and total sleep time (sum of minutes spent in all sleep stages) but were transformed to simply sleep or wake for the purpose of the present analyses.

Results

Demographics

A total of 29 volunteers (20 males, 9 females; mean (SD) age = 24.3 (5.4); 25 right-handed, 3 left-handed, 1 ambidextrous) were included in the present analysis. Data from a volunteer were not included if any actigraph or PSG data were missing due to technician error or technical problems (for AW, 4 volunteers were missing baseline and recovery; for MW, 3 volunteers were missing baseline and recovery, 1 volunteer missing baseline, and 1 volunteer missing recovery; for PSG, 6 volunteers were missing baseline and recovery, 3 volunteers missing baseline, and 1 volunteer missing recovery). Taking into account the missing data, 29 volunteers with complete data remained (of the 48 included in the larger study).

Actigraphy

The actigraph data from both devices were downloaded and automatically scored for wake and sleep for each 30-s epoch and were time synchronized with the PSG (also scored in 30-s epochs). Two sets of comparisons and analyses were performed: (1) epoch-by-epoch agreement with discriminability index (d') calculations and (2) sleep parameter concordance. For the epoch-by-epoch agreement measures and sleep parameters, repeated measures ANOVAs were performed for all variables, using Metric

Fig. 1 Schematic of the study design and procedures. Hours awake is on the top *x*-axis, and clock time is on the bottom *x*-axis. Allocated time in bed is indicated by black shading on the baseline and recovery nights. Motionlogger Watch, Actiwatch-64, and polysomnographic sleep/wake data were collected as indicated on baseline and recovery nights of sleep. PSG, polysomnography
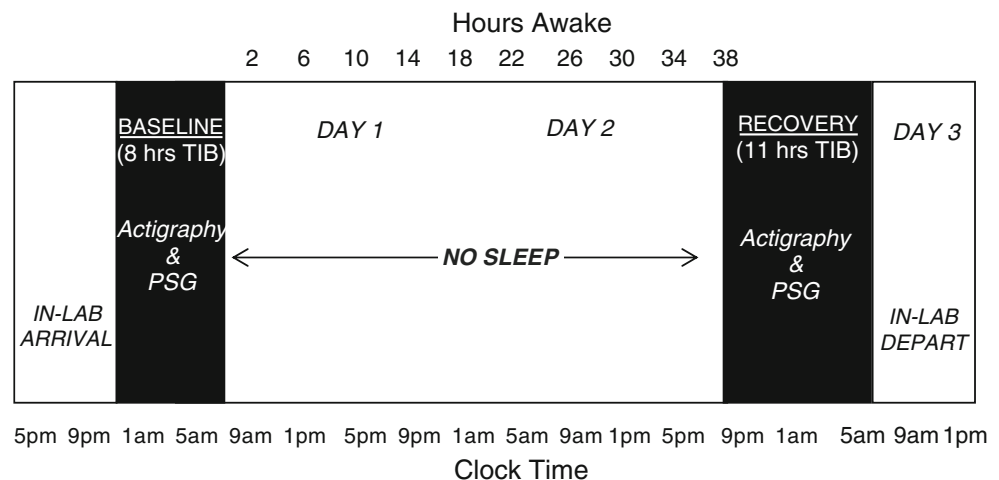
(MW or AW) and Night (baseline or recovery) as within-subjects factors. All analyses were performed using SPSS software, Version 12 (SPSS, Chicago, IL). For the sleep parameter analyses, post-hoc paired t tests with Bonferroni corrections were used to follow up on significant main effects of metric and significant interactions of metric and night. Greenhouse–Geisser corrections and significance levels set at $p < .05$ were used for all analyses.

Epoch-by-epoch Agreement For the epoch-by-epoch analysis, the percentages of matching epochs were calculated among the two different actigraphs and PSG using Tyron's (1991) method of calculating and reporting sensitivity, specificity, and overall agreement. *Sensitivity* was defined as the proportion of PSG sleep epochs also identified as sleep by actigraphy; *specificity* was defined as the proportion of nonsleep (wake) epochs correctly identified by actigraphy, and *agreement* was defined as the proportion of PSG epochs correctly identified by actigraphy (true sleep epochs + true wake epochs / all epochs).

The discriminability index (d') and criterion c were calculated in order to further assess device sensitivity and bias toward scoring sleep. Our reported measure of sensitivity based on Tyron's (1991) method is equivalent in terms of signal detection theory to a "hit," with the remaining proportion equivalent to the proportion of "misses" (e.g., epochs identified by PSG as sleep but by actigraphy as wake). Our measure of specificity is equivalent to a "correct rejection" in signal detection theory, with the remaining proportion equivalent to "false alarms" (e.g., epochs scored by PSG as wake but scored by an actigraph as sleep). As such, our d' and criterion c calculations were performed from these measurements as follows:

$$d' = Z(\text{sensitivity}) - Z(1 - \text{specificity}), \text{ equivalent to } d'$$
$$= Z(\text{hit rate}) - Z(\text{false alarm rate}).$$
$$c = -0.5 * [Z(\text{sensitivity}) + Z(1 - \text{specificity})], \text{equivalent to } c$$
$$= -0.5*[Z(\text{hit rate}) + Z(\text{false alarm rate})]$$

The means (with *SD*s in parentheses) and repeated measures ANOVA results (e.g., test statistics, degrees of freedom) for specificity, sensitivity, and overall agreement with PSG, as well as for d' and criterion c, are presented for both MW and AW in Tables 1 and 2.

Sensitivity, specificity, and overall agreement with PSG were significantly higher for MW than for AW, although sensitivity and overall agreement were reasonably high for both actigraphs (>89%). Specificity, although higher for MW than for AW, was generally low (66% and 56%, respectively) relative to sensitivity and overall agreement. In addition, overall agreement was higher on the recovery

versus on the baseline night [mean (*SD*) baseline = 91.6 (4.2), mean (*SD*) recovery = 93.5 (3.5)]. Figure 2 shows sensitivity (sleep detection) and specificity (wake detection) plotted for each subject averaged over the baseline and recovery nights for MW and AW. As is demonstrated in the figure, although sensitivity was generally high, specificity values greater than 80% (threshold indicated by dashed line) were few (and more numerous for MW).

The d' calculations revealed significantly better discrimination between sleep and wake for MW than for AW [mean (*SD*) MW = 2.7 (1.0), mean (*SD*) AW = 1.68 (0.90)]. No effects were significant for criterion c (ps > .05; see Table 2). Figure 3 provides a comparison of sensitivity, specificity, overall agreement, and d' and criterion c values for MW and AW, averaged over baseline and recovery nights.

Sleep Parameter Concordance The second set of analyses was conducted to compare PSG-derived sleep parameters with actigraphically estimated sleep parameters. Four sleep parameters were calculated using the following definitions for both actigraphy and PSG data: sleep onset latency (SL: minutes from lights out to the first epoch of sleep); total sleep time (TST: total minutes of sleep from lights out to lights on), number of awakenings (NW: number of continuous blocks of 30-s epochs of wake from the end of sleep latency to lights on), and sleep

Table 1 Epoch-by-epoch agreement results for descriptive statistics

Variable	MW Mean (*SD*)	AW Mean (*SD*)
Sensitivity (%)[a]		
Baseline	96.2 (3.6)	92.2 (2.9)
Recovery	97.0 (2.0)	92.4 (2.3)
Specificity (%)[a]		
Baseline	63.6 (28.1)	57.6 (23.8)
Recovery	68.5 (27.3)	54.7 (26.3)
Overall Agreement (%)[a,b]		
Baseline	93.6 (4.0)	89.6 (3.5)
Recovery	95.9 (2.2)	91.2 (2.8)
Discriminability index (d')[a]		
Baseline	2.52 (1.0)	1.72 (0.8)
Recovery	2.79 (1.0)	1.64 (1.0)
Criterion[c]		
Baseline	–0.61 (0.6)	–0.58 (0.4)
Recovery	–0.55 (0.5)	–0.64 (1.2)

MW, Motionlogger Watch; AW, Actiwatch-64.

[a] Significant difference between metrics, $p < .05$.

[b] Significant difference between nights, $p < .05$.

[c] Significant metric x night interaction, $p < .05$.

Table 2 Epoch-by-epoch agreement results for repeated measures ANOVA results

Effect	F	DF_{num}	DF_{denom}	p
Metric (MW or AW)				
Sensitivity*	142.4	1	28	<.001
Specificity*	5.9	1	28	.022
Overall agreement*	120.2	1	28	<.001
d'*	61.8	1	28	<.001
c	0.2	1	28	.672
Night (baseline or recovery)				
Sensitivity	0.9	1	28	.340
Specificity	0.1	1	28	.812
Overall agreement*	8.5	1	28	.007
d'	0.4	1	28	.521
c	0.001	1	28	.978
Metric x Night				
Sensitivity	1.5	1	28	.229
Specificity	3.3	1	28	.080
Overall agreement	1.5	1	28	.237
d'	3.9	1	28	.058
c	2.8	1	28	.104

MW, Motionlogger Watch; AW, Actiwatch-64. *Significant effect, $p < .05$

Fig. 2 Scatterplot of sensitivity (*y*-axis, detection of sleep) and specificity (*x*-axis, detection of wakefulness) for all subjects averaged across baseline and recovery nights for Motionlogger Watch (MW, filled circles) and Actiwatch-64 (AW, open circles). Data points to the right of the vertical dashed line represent values showing sensitivity and specificity values ≥80%. As illustrated here, both devices produced sensitivity values >80%, but specificity was much lower than sensitivity (though the MW values generally show both greater sensitivity and specificity, as compared to AW)

efficiency (SE: percentage of sleep between sleep onset and awakening).

The means (*SD*s in parentheses) and the repeated measures ANOVA results (test statistics and degrees of freedom) for SL, TST, NW, and SE for MW, AW, and PSG are presented in Tables 3 and 4.

The results from post-hoc paired *t* tests (Bonferroni corrected) for significant main effects of metric and significant interactions of metric and night for each estimated sleep parameter are summarized as follows, and also illustrated in Fig. 4.

Sleep Latency On the baseline night, the AW-estimated SL was significantly shorter than the MW-estimated SL [mean (*SE*) difference AW – MW = −7.95 (1.58), $p < .001$] and than the PSG-derived SL [mean (*SE*) difference AW – PSG = −6.85 (1.71), $p = .001$]; MW and PSG did not differ. On the recovery night, the MW-estimated SL was significantly longer than those for both AW [mean (*SE*) difference MW – AW = 4.19 (0.97), $p = .001$] and PSG [mean (*SE*) difference MW – PSG = 4.05 (0.87), $p < .001$]; AW and PSG did not differ. SL was shorter on the recovery night overall.

Total Sleep Time On the baseline night, AW-estimated TST was significantly shorter than either MW-estimated or PSG-derived TST [mean (*SE*) difference AW – MW = −15.66
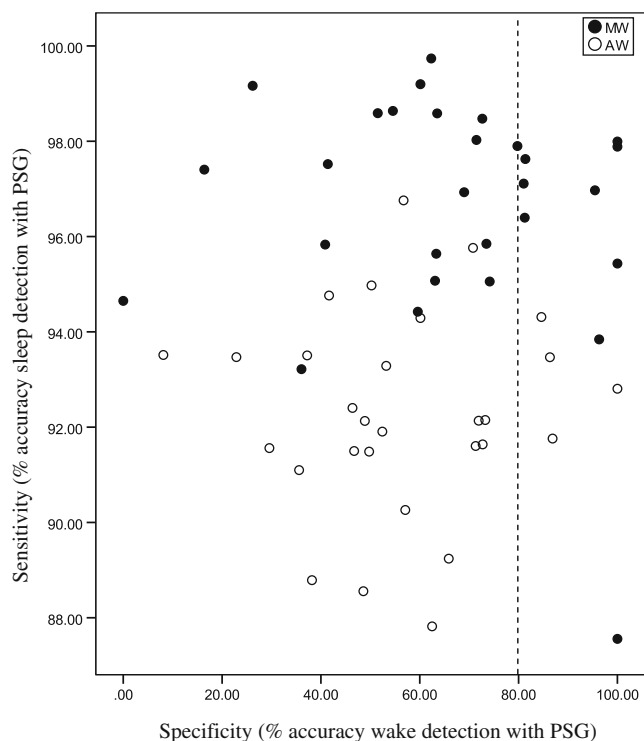
(2.17), $p < .001$; AW – PSG = −20.35 (3.45), $p < .001$]. MW and PSG did not differ. The calculations of TST for all metrics were significantly different from one another on the recovery night [mean (*SE*) difference MW – AW = 25.19 (2.45), $p < .001$; MW – PSG = −14.69 (2.90), $p < 0.001$; AW – PSG = −39.88 (2.60), $p < .001$], with AW-estimated TST being the lowest, PSG-derived TST the greatest, and MW-estimated TST in between. The TST estimates overall were greater on the recovery night than on the baseline night.

Number of Awakenings More awakenings were estimated with AW than with MW or than were derived from PSG on both the baseline nights [mean (*SE*) difference AW – MW = 29.38 (1.75), $p < .001$; AW – PSG = 25.90 (2.64), $p < .001$] and the recovery nights [mean (*SE*) difference AW – MW = 39.48 (2.34), $p < .001$; AW – PSG = 35.00 (2.57), $p < 0.001$]. MW and PSG did not differ significantly. NW was greater overall on the recovery night than on the baseline night.

Fig. 3 Sensitivity, specificity, overall agreement, d', and criterion c values averaged across baseline and recovery nights for Motionlogger Watch (MW, gray bars) and Actiwatch-64 (AW, black bars). The sensitivity, specificity, and overall agreement values were divided by 100 for comparison with d' and c. Asterisks indicate significant differences between the metrics
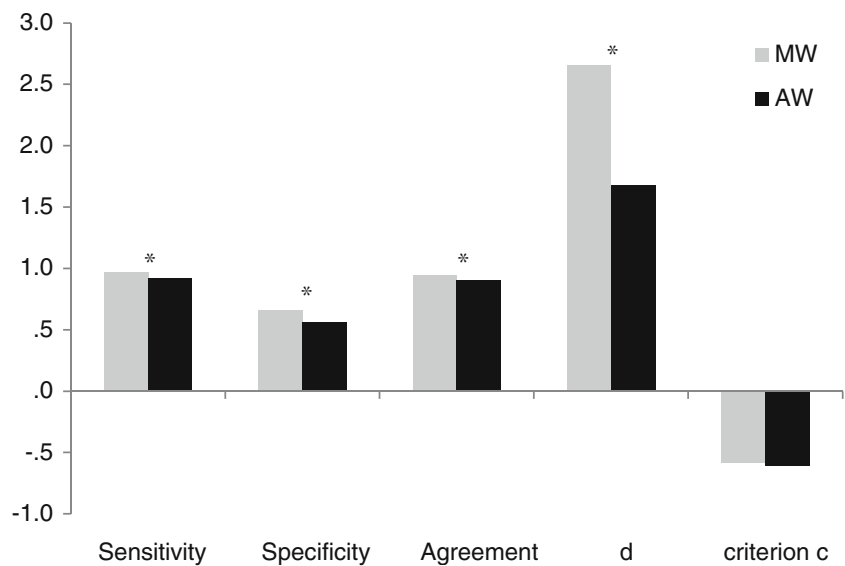
Sleep Efficiency On the baseline night, the AW-estimated SE was lower than either the MW-estimated or the PSG-derived SE [mean (SE) difference AW – MW = −3.28 (0.45), $p < .001$; AW – PSG = 4.41 (0.75), $p < .001$]. All metrics were significantly different on the recovery night [mean (SE) difference MW– AW = 3.82 (0.37), $p < .001$; MW – PSG = −2.23 (0.44), $p < .001$; AW – PSG = −6.05 (0.39), $p < .001$], with AW-estimated SE the lowest, PSG-derived SE the greatest, and MW-estimated SE in between. SE was higher overall on the recovery night than on the baseline night.

Discussion

Sleep/wake identification and sleep parameters obtained with MW, AW, and PSG were compared on the basis of epoch-by-epoch agreement and sleep parameter concordance during baseline and recovery nights of sleep in the laboratory. The epoch-by-epoch agreement analyses revealed significantly higher sensitivity (sleep identification), specificity (wake detection), and overall agreement with PSG for the MW, as compared to the AW (though sensitivity and overall agreement were high and specificity

Table 3 Sleep parameter comparison results for descriptive statistics

Variable	MW Mean (SD)	AW Mean (SD)	PSG Mean (SD)
SL (min)[a,b]			
Baseline	14.8 (9.6)	6.9 (5.6)[c]	13.7 (10.6)
Recovery	9.1 (7.0)[c]	4.9 (3.9)	5.0 (5.5)
TST (min)[a,b]			
Baseline	441.7 (21.7)	426.0 (16.7)[c]	446.4 (21.9)
Recovery	622.8 (22.2)[c]	597.6 (21.3)[c]	637.5 (22.4)
NW (#)[a,b]			
Baseline	7.4 (5.1)	36.8 (11.8)[c]	10.9 (9.9)
Recovery	8.4 (5.5)	47.9 (14.2)[c]	12.9 (13.1)
SE (%)[a,b]			
Baseline	92.5 (4.4)	89.3 (3.4)[c]	93.7 (4.0)
Recovery	94.6 (3.4)[c]	90.8 (3.4)[c]	96.8 (3.4)

MW, Motionlogger Watch; AW, Actiwatch-64; PSG, polysomnography; SL, sleep latency; TST, total sleep time; NW, number of awakenings; SE, sleep efficiency. [a] Significant difference between metrics, $p < .05$.

[b] Significant difference between nights, $p < .05$.

[c] Significantly different from PSG, $p < .05$.

Table 4 Sleep parameter comparison results for repeated measures ANOVA results

Effect	F	DF_{num}	DF_{denom}	p
Metric (MW, AW, PSG)				
SL*	13.5	2	55	<.001
TST*	63.9	2	44	<.001
NW*	193.2	2	51	<.001
SE*	53.9	2	41	<.001
Night (baseline or recovery)				
SL*	24.0	1	28	<.001
TST*	2,009.4	1	28	<.001
NW*	16.3	1	28	<.001
SE*	10.9	1	28	.003
Metric x Night				
SL*	10.3	2	52	<.001
TST*	25.9	2	49	<.001
NW*	12.0	1	40	<.001
SE*	4.0	2	42	.037

MW, Motionlogger Watch; AW, Actiwatch-64; PSG, polysomnography; SL, sleep latency; TST, total sleep time; NW, number of awakenings; SE, sleep efficiency. *Significant effects, $p < .05$.
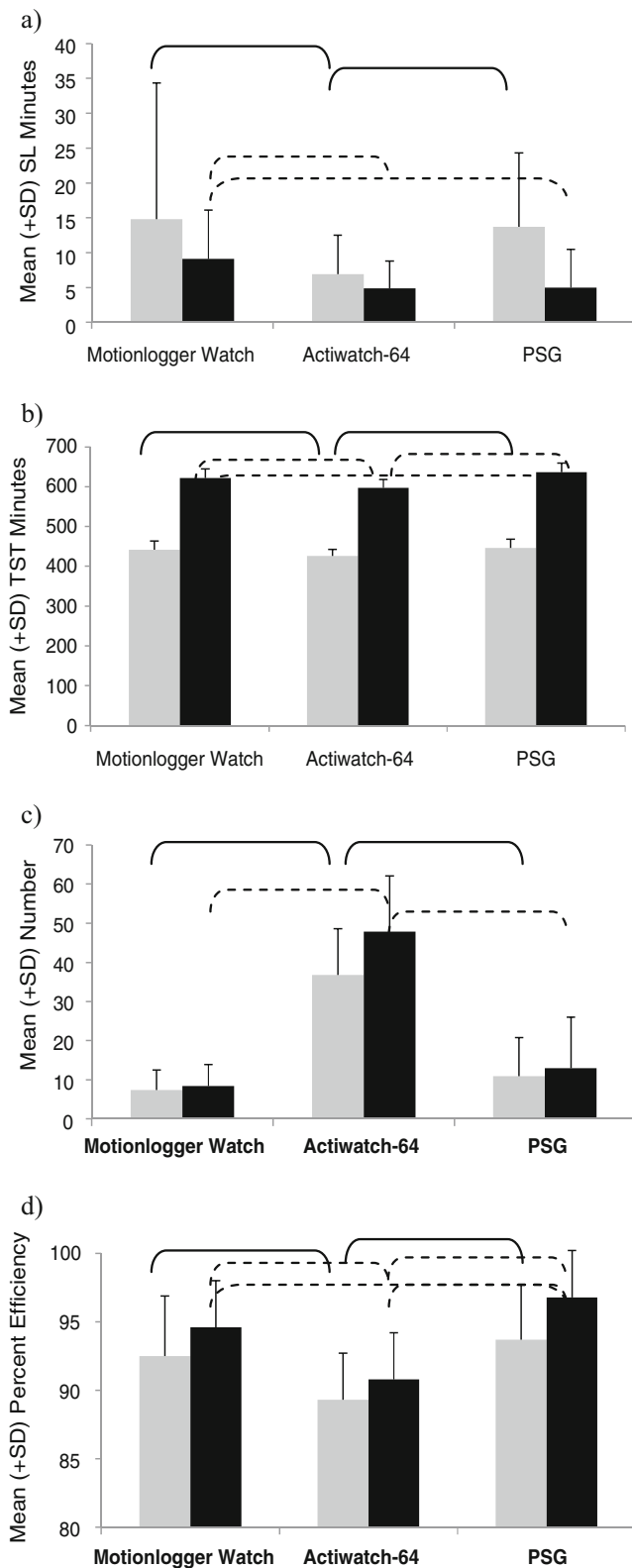
Fig. 4 Mean (+ SD) values for baseline (gray bars) and recovery (black bars) nights for (a) sleep latency, (b) total sleep time, (c) number of awakenings, and (d) sleep efficiency. Solid lines represent significant differences between metrics on the baseline night, and dashed lines represent significant difference between metrics on the recovery night

was low for both actigraphs). Discrimination index (d′) calculations revealed better signal (sleep) detection for MW than for AW. Overall, agreement was higher for recovery than for baseline nights. The sleep parameter concordance analyses showed that relative to PSG, the AW underestimated total sleep time (TST) and sleep efficiency (SE) and overestimated number of awakenings (NW) on both nights, as well as underestimating sleep latency (SL) on the baseline night; MW, on the other hand, under-estimated both TST and SE overall, and overestimated SE on the recovery night.

Sleep/wake identification using both actigraph devices was sensitive, and overall agreement with PSG was >89%, consistent with previous findings (i.e., Paquet et al., 2007). However, specificity (ability to detect wakefulness) was much lower (66% for MW and 56% for AW), a finding that is also consistent with those of previous studies (i.e., de Souza et al., 2003). In the present study, the sleep/wake comparison was performed over two nights of in-laboratory sleep, when volunteers were generally sedentary. Lack of physical activity during this time likely produced a bias toward detection of sleep. The low rates of specificity, however, showed that the actigraphs were not as accurate as PSG for identifying wakefulness in these relatively sedentary volunteers.

To assess the devices' discriminatory ability for sleep detection, d′ was calculated. The d′ index takes into account the actigraphic estimation of sleep for epochs also defined as sleep by PSG ("hits") and actigraphic estimation of sleep for epochs defined as wake by PSG ("false alarms" or "false positives"), with higher d′ values being indicative of better discrimination. The d′ value for AW was lower than that for MW, indicating worse discrimination of sleep versus wake by the AW. These data suggest that the MW was more sensitive at scoring sleep in this experimental situation (i.e., in laboratory, during defined sleep periods). Our analyses for bias, as quantified by criterion c values, did not show any significant differences in bias between metrics or nights. These results might differ if periods of measurement outside of the sleep period were included (with a higher proportion of epochs defined as wake).

Although sensitivity and agreement with PSG were high for both actigraph devices, sleep/wake identification using the MW was significantly better, with higher agreement overall versus the AW. Analyses of the data with this approach (epoch by epoch) did not reveal a moderating effect of baseline or recovery night on sensitivity or specificity, but overall, agreement was higher on the recovery night. Differences between the nights may be explained by longer TST and greater SE on the recovery night, considering that volunteers were generally immobile and relatively sleepy—conditions under which an actigraphic bias toward identification of sleep would tend to produce generally favorable comparisons with PSG.

Comparing the actigraphs to PSG on the basis of sleep parameters revealed additional differences between actigraphy and PSG. In contrast with prior studies in which it was reported that the actigraphs performed similarly, with no meaningful differences between devices for sleep measures (Benson et al., 2004; Tonetti et al., 2008), the present findings revealed that the AW-estimated TST, NW, and SE differed significantly from the PSG-derived TST, NW, and SE on both baseline and recovery nights, and that SL differed on the baseline night. The MW-estimated SL, TST, and SE were also significantly different from PSG-derived calculations on the recovery night. Thus, in the present study, MW-estimated sleep/wake was found to be more consistent with PSG-derived sleep/wake than were the AW estimates. In part, this may be because the newer MW was used, whereas the Basic Mini-Motionlogger had been used in prior studies. Findings of an interaction between in-laboratory night (baseline vs. recovery) and device for SL, TST, and SE showed that the MW-estimated sleep parameters were more consistent with PSG-derived parameters on the baseline night. The reason for this interaction is unclear, but it suggests that the reliability of the MW for sleep estimation may vary depending on sleep/wake history (in this case, on baseline night vs. recovery night following sleep deprivation).

Consistent with some previous reports (i.e., Paquet et al., 2007; Tonetti et al., 2008), AW-estimated SL was underestimated relative to the PSG-derived SL on the baseline night. MW-estimated SL was overestimated relative to PSG-derived SL on the recovery night—a finding that is consistent with those of previous reports (de Souza et al., 2003).

Discrepancies between PSG-derived and actigraphy-estimated SL are understandable, and in some cases they may be related to how SL is defined. For example, in previous studies (e.g., Cole et al., 1992) in which SL was defined as the first epoch of actigraph-estimated sleep (as was also done in the present study), the correlation between actigraphy and PSG was .53, but when sleep onset was defined as the beginning of the first period containing 20 min of actigraph-identified sleep with no more than 1 min of wake intervening, the agreement improved to .94 (Cole et al., 1992). As discussed in a review by Ancoli-Israel et al. (2003), the first-minute definition continues to be commonly used, which may account for differences between PSG versus actigraphic scoring—differences that affect not only SL, but also SE and wake after sleep onset.

Previous reports have tended to reveal overestimations of TST and SE and underestimations of NW using actigraphy (e.g., de Souza et al., 2003). In contrast, the present study revealed that TST and SE were significantly underestimated on baseline and recovery nights with AW, and that both measures were underestimated using the MW on the recovery night. The Basic Mini-Motionlogger was used in one such study (de Souza et al., 2003); however, the

data were collected in 1-min epochs, allowing for less precision than the 30-s epochs used in the present study. De Souza et al. also reported that sleep parameter estimation using actigraphy underestimated NW. In contrast, NW was overestimated using the AW in the present study, but MW-estimated and PSG-derived values for NW did not differ significantly. Of note, the number of awakenings recorded in the present study was relatively low. One explanation for this might have been greater sleep efficiency on the baseline night due to preexisting sleep debts of individuals prior to entering the sleep study and on the recovery night due to the prior night of sleep deprivation.

Because the present subject sample was limited to young, healthy adults without sleep complaints, generalizability to older adults or children, with or without sleep complaints, has not been established. Also, because all data collection was performed in the controlled conditions of the laboratory during periods designated for sleep, there is a possibility that different results would be obtained outside of the lab (i.e., at home) or during self-selected sleep periods. In addition, it should be noted that because volunteers were assessed during an externally imposed sleep period, this may have artificially increased the occurrence of sedentary wakefulness. Thus, further study in clinical populations outside of the laboratory is warranted. An additional limitation of the present study and a consideration for future studies is that placement of the devices (i.e., closer to the hand or to the elbow) was not balanced (volunteers were only instructed to wear the devices on the same wrist). Movement detection might differ depending on whether the actigraph is closer to the hand than to the elbow. Finally, the technical specifications for the devices used in the study were consistent (30-s epoch length and Cole–Kripke scoring algorithm); these specifications or different scoring algorithms for other devices might show better (or worse) agreement with PSG. Indeed, lowering the wake sensitivity for the AW might significantly improve wake detection.

In summary, the present findings help delineate the extent to which actigraphy is a useful and reliable alternative to PSG for sleep/wake identification and sleep parameter estimation in healthy young adults in the laboratory. In the present study, the MW was found to provide some advantages relative to the AW. An important consideration, however, is that while sensitivity to detecting sleep and overall agreement was high for both actigraphs, specificity for detecting wake was much lower. Taken together, these data suggest that PSG remains the preferred method for estimates of sleep and wakefulness transitions (e.g., sleep onset latency, number of awakenings, and sleep efficiency). Actigraphy remains a useful tool for measures of total sleep time. Thus, PSG would be recommended for overnight clinical assessments regarding diagnosing insomnia or in research settings or studies for which sleep/wake transitions are important (e.g., sleep fragmentation). Actigraphy remains a valuable tool for characterizing

sleep/wake patterns overall, and may be especially useful in research or clinical settings in which confirmation of usual sleep habits or patterns is needed. Given the convenience and cost effectiveness of actigraphy relative to PSG, researchers and/or clinicians may still choose to use it in situations in which PSG might be preferred but is not feasible for practical considerations. In such cases, our data suggest that the MW is the more reliable tool for sleep/wake estimation, as compared to AW, especially given the potentially greater bias toward scoring sleep for the AW. Room for improvement remains, but as long as researchers and clinicians remain mindful of its limitations, actigraphy serves as a useful and reliable tool.

References

Ancoli-Israel, S., Cole, C., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. P. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep, 26*, 342–392.

Beck, A. T., & Steer, R. A. (1993). *Manual for the beck depression inventory*. San Antonio, TX: Psychological Corp.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561–671.

Benson, K., Friedman, L., Noda, A., Wicks, D., Wakabayashi, E., & Yesavage, J. (2004). The measurement of sleep by actigraphy: Direct comparison of 2 commercially available actigraphs in a nonclinical population. *Sleep, 27*, 986–989.

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 327*, 307–310.

Cole, R. J., & Kripke, D. F. (1988). Progress in automatic sleep/wake scoring by wrist actigraph. *Sleep Research, 17*, 331.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep, 15*, 461–469.

de Souza, L., Benedito-Silva, A. A., Pires, M. L., Poyares, D., Tufik, S., & Calil, H. M. (2003). Further validation of actigraphy for sleep studies. *Sleep, 26*, 81–85.

Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*, 97–110.

Mullaney, D. J., Kripke, D. F., & Messin, S. (1980). Wrist-actigraphic estimation of sleep time. *Sleep, 3*, 83–92.

Paquet, J., Kawinska, A., & Carrier, J. (2007). Wake detection capacity of actigraphy during sleep. *Sleep, 30*, 1362–1369.

Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages in human subjects*. Washington, DC: U.S. Government Printing Office.

Rupp, T. L., Killgore, W. D. S., & Balkin, T. J. (2010). Socializing by day may affect performance by night: Vulnerability to sleep deprivation is differentially mediated by social exposure in extraverts vs. introverts. *Sleep, 33*, 1475–1485.

Spielberger, C. D., & Vagg, P. R. (1984). Psychometric properties of the STAI: A reply to Ramanaiah, Franzen, and Schill. *Journal of Personality Assessment, 48*, 95–97.

Tonetti, L., Pasquini, F., Fabbri, M., Belluzzi, M., & Natale, V. (2008). Comparison of two different actigraphs with polysomnography in healthy young subjects. *Chronobiology International, 25*, 145–153.

Tyron, W. W. (1991). Activity and sleep. In *Activity measurement in psychology and medicine* (pp. 149–195). New York: Plenum Press.