

REPORT

Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults

Nicola Cellini¹, Matthew P. Buman², Elizabeth A. McDevitt³, Ashley A. Ricker³, and Sara C. Mednick³

¹*Department of General Psychology, University of Padua, Padua, Italy*, ²*School of Nutrition and Health Promotion, Arizona State University, Phoenix, Arizona, USA*, and ³*Department of Psychology, University of California, Riverside, California, USA*

The last 20 yrs have seen a marked increase in studies utilizing actigraphy in free-living environments. The aim of the present study is to directly compare two commercially available actigraph devices with concurrent polysomnography (PSG) during a daytime nap in healthy young adults. Thirty healthy young adults, ages 18–31 (mean 20.77 yrs, SD 3.14 yrs) simultaneously wore AW-64 and GT3X+ devices during a polysomnographically recorded nap. Mann-Whitney *U* (M-U) test, intraclass correlation coefficients, and Bland-Altman statistic were used to compare total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency (SE) between the two actigraphs and PSG. Epoch-by-epoch (EBE) agreement was calculated to determine accuracy, sensitivity, specificity, predictive values for sleep (PVS) and wake (PVW), and kappa and prevalence- and bias-adjusted kappa (PABAK) coefficients. All frequency settings provided by the devices were examined. For both actigraphs, EBE analysis found accuracy, sensitivity, specificity, PVS, and PVW comparable to previous reports of other similar devices. Kappa and PABAK coefficients showed moderate to high agreement with PSG depending on device settings. The GT3X+ overestimated TST and SE, and underestimated SOL and WASO, whereas no significant difference was found between AW-64 and PSG. However, GT3X+ showed overall better EBE agreements to PSG than AW-64. We conclude that both actigraphs are valid and reliable devices for detecting sleep/wake diurnal patterns. The choice between devices should be based on several parameters as reliability, cost of the device, scoring algorithm, target population, experimental condition, and aims of the study (e.g., sleep and/or physical activity). (Author correspondence: smednick@ucr.edu)

Keywords: Actigraphy, Bland-Altman, nap, polysomnography, validation

INTRODUCTION

Actigraphy is a noninvasive, objective monitoring device that evaluates sleep/wake patterns based on the recording of movement.

This ecological and noninvasive technology is a cost-effective substitute for polysomnography (PSG), the gold standard for sleep assessment, to objectively monitor the sleep/wake rhythm in home setting (Meltzer et al., 2012a). Actigraphy devices (actigraphs) are worn on the wrist, ankle, or hip and contain an accelerometer that monitors and collects body movement. Actigraphy does not require specialized technicians, permits several days of recordings, and is useful for studying specific populations such as infants and children (Meltzer et al., 2012a), older adults (see Blackwell et al., 2008), and psychiatric populations (Poyurovsky et al., 2000). Previous validation studies comparing various actigraph device models to PSG have reported agreement rates ranging between 85% and 95% for the identification of sleep/wake epochs (Sadeh, 2008).

The last 20 yrs have seen a marked increase in studies utilizing actigraphy in free-living environments (Sadeh, 2011). Several commercially available actigraphy devices are currently being used in research settings. For example, a recent review of pediatric sleep research reported that eight different actigraphy brands have been utilized (Meltzer et al., 2012a). Validation studies that examine interdevice agreement and comparisons with PSG are needed to correctly interpret results from studies involving different devices and sleep/wake scoring algorithms. To date, few studies have conducted direct comparisons of different actigraphy devices (Benson et al., 2004; Pollak et al., 1998), with only a subset of these utilizing comparisons with PSG recordings (Meltzer et al., 2012b; Rupp & Balkin, 2011; Tonetti et al., 2008; Weiss et al., 2010). These studies reveal that different brands could give discrepant results based on differences in manufacturer specifications, settings, and algorithms used, and specific devices could be more useful for a given population. For example, Meltzer and

Submitted November 13, 2012, Returned for revision February 15, 2013, Accepted March 1, 2013

Correspondence: Sara C. Mednick, Department of Psychology, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA. Tel.: 951-827-5259. Fax: + 951-827-3985. E-mail: smednick@ucr.edu

colleagues (2012b) found different outcomes comparing two actigraphy brands, suggesting caution should be used when comparing results from research involving different actigraphy brands.

An important limitation of these studies is that they exclusively measured nighttime sleep periods. However, one of the common problems with actigraphy is its poor ability to detect wakefulness (Paquet et al., 2007). This problem could lead to an important bias: as noted by de Souza et al. (2003), nighttime recordings in healthy participants are almost exclusively composed of sleep. Consequently, an actigraph scoring all epochs as sleep could give 92% agreement with PSG (Sadeh et al., 1989). This could explain differences in the accuracy of actigraphic assessment of sleep between healthy adults and populations with fragmented sleep, such as insomniacs (Chesson et al., 2007; Sadeh & Acebo, 2002; Sadeh et al., 1995).

To address this limitation, we compared the reliability of different actigraph devices with PSG recordings of daytime sleep (naps). As reported by Milner & Cote (2009), there is a lack of studies investigating how the architecture of daytime sleep differs from nocturnal sleep architecture. However, nap studies in healthy sleepers report lower sleep efficiency (SE) than nocturnal sleep (e.g., 68% of SE in Tucker et al. (2006), 77% in Kanady et al. (2011), and 77% in Nishida & Walker (2007)), mainly due to the “weight” of the sleep onset latency in a short sleep period. Thus, the higher likelihood of wakefulness in the sleep period during daytime sleep provides a better context for validation studies of actigraphy devices.

The aim of this study was to directly compare two widely used actigraph devices, Actiwatch-64 (AW-64; Phillips Respironics, Bend, Oregon, USA) and the GT3X+ (Actigraph, Pensacola, Florida, USA), with a PSG-recorded nap. AW-64 is commonly used in sleep research (Meltzer & Westin, 2011) and has been validated both for nocturnal sleep (Rupp & Balkin, 2011) and daytime naps (Kanady et al., 2011). The GT3X+ is regularly used to measure daytime activity (Rowlands & Stiles, 2012), but it also provides the possibility to record and score sleep/wake patterns. Measuring sleep and physical activity at the same time with a single device is a useful and efficient way to obtain information about well-being and health behaviors. However, to the best of our knowledge, this device has never been validated for sleep.

MATERIALS AND METHODS

Participants

Thirty-four nonsmoking adults (19 female) between the ages of 18 and 31 gave informed consent to participate in the experiment, which was approved by the University of California at Riverside Human Research Protections Program and in accordance with accepted international ethical standards (Portaluppi et al., 2010).

Exclusionary criteria included (a) not having a regular sleep/wake schedule (defined as habitually spending 7–9 h of time in bed per night); (b) having a sleep disorder (sleep disorders were screened by interviewing the subject and asking about potential symptoms of insomnia, apnea, narcolepsy, and restless leg syndrome/periodic leg movements); (c) any personal or immediate family (i.e., first-degree relative) history of diagnosed significant psychopathology; (d) personal history of head injury with loss of consciousness greater than 2 min or seizures; (e) history of substance dependence; (f) current use of any psychotropic medications; and (g) any cardiac, respiratory, or other medical condition that may affect metabolism. Participants were asked to maintain their usual sleep/wake schedule the week prior to the study, and compliance was monitored with sleep diaries and actigraphy. They received financial compensation and course credit for participating in the study.

Due to technical problems with actigraphs, we had to exclude four participants due to inaccessible data. The final sample consisted on 30 adults (16 female; mean 20.77 yrs, SD 3.14 yrs).

Procedure

The study took place at the Sleep and Cognition Lab in the Department of Psychology at the University of California, Riverside. All participants wore two different actigraphs on their nondominant wrist during a PSG-recorded daytime nap. Naps occurred between 13:00 and 15:00 h. Participants were allowed to sleep up to 120 min in bed. Each participant was recorded once. The devices' clocks were synchronized in order to compare the output of the two actigraphs with the PSG.

Polysomnography

All PSG data were collected using Astro-Med Grass Heritage model 15 amplifiers (Astro-Med, Inc., West Warwick, RI, USA) with Grass Gamma software (Astro-Med, Inc., West Warwick, RI, USA). Three unipolar electroencephalogram (EEG) channels (C4-A1, C3-A2, O1-A2) were recorded according to the International 10–20 system (Jasper, 1958). Two unipolar electrooculograms referenced to opposite mastoids and chin bipolar electromyogram were recorded. Raw data were digitalized at a sampling rate of 256 Hz and passed to the Grass Gamma software, where the data were filtered and visually scored in 30-s epochs following American Academy of Sleep Medicine (AASM) rules for sleep staging (Iber et al., 2007).

Actigraphy

The AW-64 (Phillips Respironics) is $2.8 \times 2.7 \times 1$ cm in size and weighs 16 g. The hardware consists of a piezoelectric accelerometer with a sensitivity of 0.02 g, a sampling rate of 32 Hz, and a storage capacity of 64 kb. Recording at 1-min epochs allows for about 45 d of data collection. The GT3X+ (Actigraph) is $4.6 \times 3.3 \times 1.5$ cm in size and weighs 19 g. It contains a triaxial accelerometer

with a sensitivity of 0.05 g and the sampling rate ranges from 30 to 100 Hz. Recording at a 30-Hz sampling rate allows for more than 40 d of data collection.

The GT3X+ is commercially available. The current price (2013) is \$249.* Although AW-64 is still largely used in sleep research, it has been replaced by newer models from the same brand. However, a Respiroics report (Philips Respiroics, 2008) showed that AW-64 and the newer Actiwatch devices, which cost about \$2300 per device, may be used interchangeably to compute sleep statistics when used with the Actiware software algorithms.

Data processing

AW-64 nap-recordings were analyzed using Actiware 5.52.0003 (Philips Respiroics) software. Data were scored using three different sensitivity threshold levels (high, medium, low) provided by the software. Here sensitivity refers to the number of activity counts used to define wake (i.e., medium sensitivity threshold used 40 activity counts per epoch to score wake, whereas low used 20 counts and high used 80 counts per epoch).

GT3X+ data were analyzed using ActiLife 6.4.3 (Actigraph) software. This software provides two different algorithms for sleep scoring: the Sadeh algorithm (Sadeh et al., 1994) and the Cole-Kripke algorithm (Cole et al., 1992). Given the lower age of our sample, we used the Sadeh algorithm, which was validated in a young population (age range from 10 to 25 yrs), whereas the Cole-Kripke algorithm was validated in an older sample (age range from 35 to 65 yrs). Additionally, the software provides a low-frequency extension (LFE) option designed to lower the band-pass filter threshold for signal detection. We scored our data with and without this additional option.

ActiLife 6.4.3 automatically scores sleep using 60-s epoch cycles, even if the GT3X+ is initialized to collect data in 30-s epochs or shorter (in which case the data is averaged into 1-min epochs). Due to this fact, both devices were initialized to collect data in 1-min epochs.

All analyses were confined to the period between lights off and lights on, which was set by the PSG recording and synchronized with the two software programs. In order to conduct epoch-by-epoch comparison between PSG data and the two actigraph recordings, the 30-s epochs of PSG scoring were transformed to 1-min epochs to match the 1-min epochs of the actigraphs using the following rules: an epoch was marked as wake if one or both of the two 30-s epochs were scored as wake, otherwise the epoch was scored as sleep (Meltzer et al., 2012b; Sadeh et al., 1994).

Data analysis

Sleep parameters concordance analysis

Four sleep parameters were examined in this study: total sleep time (TST), defined as the number of minutes

scored as sleep between lights off and lights on; sleep onset latency (SOL), the number of minutes between lights out and the first epoch scored as sleep; wake after sleep onset (WASO), the number of minutes scored as wake after sleep onset; and sleep efficiency (SE), the ratio between TST and total time spent in bed (TIB; from lights off to lights on). For the two actigraphs, these parameters were directly extracted from the output of the respective software packages. Mann-Whitney U (M-U) test and intraclass correlation coefficients (ICCs) were used to compare sleep parameters between the different devices and PSG. A $p < 0.05$ was considered statistically significant for all analyses.

Bland-Altman plots

Due to inadequacies using correlation and t -test analyses to evaluate the agreement between two measurement methods (Altman & Bland, 1983; Bland & Altman, 1999), we also used Bland-Altman plots to evaluate the concordance between the sleep parameters derived from actigraph software and PSG recordings. This technique plots the difference score between two measures (i.e., TST for actigraphy minus TST for PSG) against their averages. In order to determine the significance of the “bias” (the mean of the differences between the two methods), we used the upper and the lower limits of this mean based on 95% confidence intervals.

Epoch-by-epoch agreement analysis

In order to calculate accuracy, sensitivity, and specificity (Table 1) for all AW-64 (low, medium, high) and GT3X+ (none, LFE) settings, we conducted an epoch-by-epoch (EBE) analysis with PSG scorings coded as binary scores (0 = sleep and 1 = wake).

Accuracy represents the agreement between PSG and actigraph scoring, sensitivity is the percentage of epochs identified correctly as sleep, whereas specificity is the percentage of epochs identified correctly as wake by the actigraph. As previously required, we computed the predicted value for sleep (PVS), which reflects the percentage of epochs scored as sleep by the actigraph that were also scored as sleep via PSG, and the predicted value for wakefulness (PVW), which reflects the percentage of epochs scored as wake by the actigraph that were also scored as wake via PSG (Ancoli-Israel et al., 2003).

We calculated Cohen’s kappa coefficient value for all settings of the two actigraphs, which is an index of interrater reliability that reflects the percentage of scoring agreement of two methods not due to chance. Given the high proportion of sleep epochs in our data, we could fall into “the first paradox of kappa statistic” (Feinstein & Cicchetti, 1990), that is, when two measures show high agreement but a low kappa. In order to correct this bias, we also computed a prevalence- and

*This price does not include the wrist strap (\$15) and the ActiLife 6.4.3 software (\$1295).

TABLE 1. Epoch-by-epoch comparison.

Actigraph	PSG		Total
	Sleep	Wake	
Sleep	True sleep (TS)	False wake (FW)	TS + FW
Wake	False sleep (FS)	True wake (TW)	FS + TW
Total	TS + FS	FW + TW	TS + FS + TW + FW

Sensitivity = $TS/(TS + FS)$; Specificity = $TW/(FW + TW)$; Accuracy = $(TS + TW)/(TS + FS + TW + FW)$; Predictive value for sleep (PVS) = $TS/(TS + FW)$; Predictive value for wake (PVW) = $TW/(TW + FS)$.

TABLE 2. Summary sleep measures.

Measure	PSG	AW-64			GT3X+	
		Low	Med	High	ACT	LFE
TST (min)	48.23 ± 25.63	43.53 ± 22.79	47.13 ± 23.68	48.4 ± 25.52	57.03 ± 20.13	56.27 ± 20.19
SOL (min)	10.60 ± 5.88	14.37 ± 15.36	14.37 ± 15.36	14.37 ± 15.36	7.67 ± 5.44	7.8 ± 5.40
WASO (min)	15.27 ± 13.63	16.37 ± 13.63	12.8 ± 12.31	11.57 ± 12.59	9.37 ± 11.59*	10 ± 12.04
SE (%)	62.61 ± 27.52	57.46 ± 27.48	62.37 ± 27.23	63.37 ± 28.91	77.14 ± 21.07*	76.15 ± 21.62*

* $p < 0.05$. PSG = polysomnography; Low = AW-64 low setting; Med = AW-64 medium setting; High = AW-64 high setting; ACT = GT3X + default setting; LFE = GT3X + low-frequency extension setting; TST = total sleep time; SOL = sleep onset latency; WASO = wake after sleep onset; SE = sleep efficiency.

bias-adjusted kappa (PABAK), which gives balanced weight to sleep and wake epochs (Byrt et al., 1993).

According to the Landis and Koch scale (Landis & Koch, 1977), we considered a kappa coefficient of 0–0.2 as slight agreement, 0.2–0.4 as fair agreement, 0.4–0.6 as moderate agreement, 0.6–0.8 as substantial agreement, and 0.8–1.0 almost perfect agreement.

RESULTS

Sleep parameters concordance analysis

As reported in Table 2, the three settings of the AW-64 (high, medium, and low) showed no significant differences compared with PSG for any sleep parameter. The M-U test revealed a significant overestimation of SE for both settings of the GT3X+ and a significant underestimation of WASO only for the default setting (ACT).

ICCs highlighted a modest (for WASO, SE) to substantial (for TST) relationship between each setting of the two devices and PSG (Table 3). ICCs also showed a low correlation between AW-64 settings and SOL, whereas coefficient values were modest for the GT3X+ settings.

Bland-Altman plots

The Bland-Altman technique revealed a systematic bias of the GT3X+ settings for each sleep parameter (Figure 1). Based on confidence intervals, as reported in Table 4, the two GT3X+ settings significantly overestimated TST and SE, and underestimated SOL and WASO. No significant over- or underestimation was shown for the AW-64 settings.

Epoch-by-epoch agreement analysis

Accuracy, sensitivity, specificity, PVS, and PVW values are showed in Table 5. As expected, the actigraphs

TABLE 3. Intraclass correlations between devices and PSG.

Measure	AW-64			GT3X	
	Low	Med	High	ACT	LFE
TST	0.79	0.80	0.80	0.76	0.76
SOL	0.29	0.29	0.29	0.55	0.56
WASO	0.51	0.49	0.46	0.51	0.54
SE	0.70	0.68	0.69	0.49	0.50

Low = AW-64 low setting; Med = AW-64 medium setting; High = AW-64 high setting; ACT = GT3X+ default setting; LFE = GT3X+ low-frequency extension setting; TST = total sleep time; SOL = sleep onset latency; WASO = wake after sleep onset; SE = sleep efficiency.

showed high sensitivity and low to medium specificity. In general, the GT3X+ showed higher values than the AW-64.

Kappa values showed a fair agreement (from 0.2 to 0.4) for the high-sensitivity setting of the AW-64, and a moderate agreement (from 0.4 to 0.6) for the other settings/devices. As expected, the PABAK showed higher κ values for each device (Table 5). Both kappa and PABAK showed higher agreement rates for the GT3X+ compared with the AW-64.

DISCUSSION

The present study examined the validity of two widely used actigraphy devices to measure sleep/wake activity in healthy young adults. To date, only few studies have compared different brands of actigraphy devices (Benson et al., 2004; Meltzer et al., 2012b; Pollak et al., 1998; Rupp & Balkin, 2011; Tonetti et al., 2008;

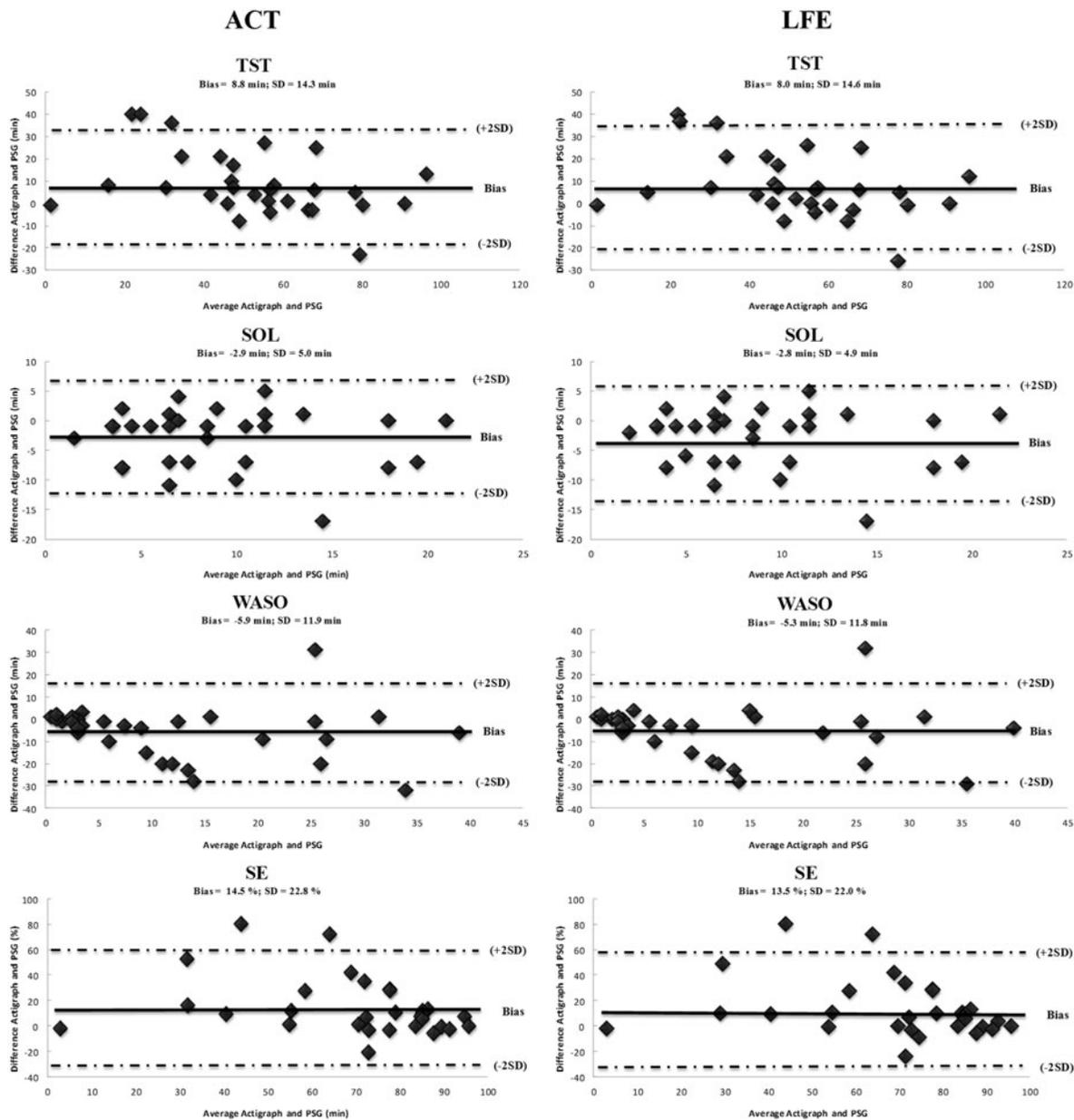


FIGURE 1. Bland-Altman plots of PSG and GT3X+ for each sleep parameters (ACT setting to the left, LFE setting to the right). Differences are shown in the y-axis and averages in the x-axis. Mean bias (Bias) and standard deviation (SD) are shown in the right corner. TST = total sleep time; SOL = sleep onset latency; WASO = wake after sleep onset; SE = sleep efficiency.

Weiss et al., 2010). In this study, we directly compared sleep parameter concordance and epoch-by-epoch agreements between two actigraph devices, the AW-64 and the GT3X+, using concurrent PSG recordings of daytime naps. Both devices, with some differences, demonstrated good reliability compared with the “gold standard.”

The GT3X+ is a recent model of the Actigraph accelerometer line, the most commonly used and extensively validated monitor for physical activity (Robusto & Trost, 2012; Trost et al., 2005). However, as far as we know, this is the only study investigating the reliability of this device for sleep/wake detection. Compared with PSG, both GT3X+ settings (ACT and LFE) overestimated SE as a consequence of

underestimation of WASO and SOL. Bland-Altman confidence limits further highlighted these differences, showing a significant underestimation of SOL and WASO, and consequently an overestimation of TST and SE for both GT3X+ settings. These results are in line with literature showing that difficulty detecting wake is the most critical problem of actigraphy validity (Paquet et al., 2007). However, ICCs showed strong correlations for TST and a moderate association for SOL, SE, and WASO for both settings. Notwithstanding the difference in the sleep parameters, EBE comparison revealed sensitivity, specificity, and accuracy values for both GT3X+ settings similar to previous studies (Martin & Hakim, 2011). Also, PVS and PVW showed good agreement between the GT3X and PSG. Even though different

TABLE 4. Bland-Altman statistics between actigraphs and PSG.

Measure	AW-64			GT3X+	
	Low TST	Med TST	High TST	Act TST	LFE TST
Mean difference	-4.70	-1.10	3.77	8.80	8.03
Standard deviation	15.52	15.96	13.76	14.33	14.56
Upper limit	0.85	4.61	8.69	13.93	13.24
Lower limit	-10.85	-6.81	-1.16	3.67	2.28
Over- or underestimation				Over	Over
	Low SOL	Med SOL	High SOL	Act SOL	LFE SOL
Mean difference	3.77	3.77	3.77	-2.93	-2.80
Standard deviation	13.76	13.76	13.76	4.98	4.94
Upper limit	8.69	8.69	8.69	-1.16	-1.03
Lower limit	-1.16	-1.16	-1.16	-4.71	-4.57
Over- or underestimation				Under	Under
	Low WASO	Med WASO	High WASO	Act WASO	LFE WASO
Mean difference	1.10	-2.47	-3.70	-5.90	-5.27
Standard deviation	13.64	13.09	13.53	11.87	11.75
Upper limit	5.98	2.22	1.14	-1.65	-1.30
Lower limit	-3.78	-7.15	-8.54	-10.15	-9.51
Over- or underestimation				Under	Under
	Low SE	Med SE	High SE	Act SE	LFE SE
Mean difference	-5.15	-.24	0.76	14.53	13.54
Standard deviation	21.14	22.25	22.47	22.78	22.98
Upper limit	2.42	7.73	8.80	22.68	21.76
Lower limit	-12.72	-8.20	-7.28	6.38	5.31
Over- or underestimation				Over	Over

The mean difference (estimated bias of actigraphy), standard deviation (fluctuation around mean), and upper and lower limits (95% confidence interval) for all the device/settings. Bold type indicates significance and the last row states whether the setting significantly over- or underestimates the specific sleep parameter.

TABLE 5. Epoch-by-epoch analyses.

Measure	AW-64			GT3X	
	Low	Med	High	ACT	LFE
Accuracy (%)	78.97	78.52	74.44	82.78	82.78
Sensitivity (%)	86.76	92.22	94.47	94.88	94.06
Specificity (%)	64.05	52.29	36.08	59.61	61.18
PVS (%)	82.21	78.73	73.89	81.81	82.27
PVW (%)	71.64	77.82	77.31	85.88	84.32
Kappa	0.46	0.40	0.28	0.52	0.52
PABAK	0.61	0.57	0.48	0.66	0.66

PSG = polysomnography; Low = AW-64 low setting; Med = AW-64 medium setting; High = AW-64 high setting; ACT = GT3X+ default setting; LFE = GT3X+ low-frequency extension setting; PVS = predictive value for sleep; PVW = predictive value for wake; PABAK = prevalence- and bias-adjusted kappa.

physical activity estimates have been reported in a free-living research design using the ACT and the LFE settings (Ried-Larsen et al., 2012), our data show the two settings are relatively equivalent for recording daytime sleep. Future research should investigate this using longer sleep periods.

The AW-64 is an actigraph commonly used in sleep research and has been validated for daytime naps (Kanady et al., 2011). Compared with that study, our data showed worse agreement for both summary sleep parameters and epoch-by-epoch comparison. However, the length of the sleep period differed between the two studies: in the current study the mean TST was 48.23 min, whereas Kanady et al. report mean TST

of 67.6 min. Given that the accuracy of the actigraph appears to increase when the sleep duration increases (Sadeh, 2011), this could explain the differences between studies. Another difference between the studies is the PSG scoring rules applied. Kanady and colleagues used Rechtschaffen and Kales sleep staging criteria (Rechtschaffen & Kales, 1968), whereas this study used the AASM criteria (Iber et al., 2007). It has been reported that AASM criteria may score increased WASO compared with Rechtschaffen and Kales (Moser et al., 2009). Even considering these differences, the AW-64 showed satisfactory overall agreement with PSG. The M-U test and the Bland-Altman statistic showed no differences for sleep outcomes between AW-64 and PSG. The EBE found accuracy ranging between 74.44% and 78.52% for the three settings, with high sensitivity values for the medium (92.22%) and high settings (94.47%), and moderate specificity for the low setting (64.05%). Our results also confirmed that the high setting is the best to determine sleep, whereas the low setting is better for detecting wakefulness. The medium setting gives the best compromise between detecting sleep and wake, also confirmed by PVW and PVS values, and the overall agreement values. These results suggest that the choice of a specific setting should depend on the aim of the study.

In spite of high agreements values, we found very low kappa values for all settings for the two devices. This paradox (Feinstein & Cicchetti, 1990) is mainly due to the asymmetry in number of sleep and wake epochs.

As expected, the PABAK, which takes into account the different prevalence of sleep and wake epochs, showed higher values than the unweighted kappa. The GT3X+ settings reported PABAK values falling in the high agreement range (0.6–0.8), with no difference between them. In contrast, the AW-64 showed moderate agreement for all settings, with the low setting showing the best agreement, followed by the medium and then the high setting.

Our results also confirmed the poor ability of actigraphy to detect wakefulness (Sadeh, 2011). Given that actigraphy defines sleep based on lack of movements, lying in bed awake but motionless will likely be coded as sleep (Martin & Hakim, 2011). This inability to detect motionless wakefulness may be increased during naps in which participants lie motionless while trying to fall asleep, affecting the estimation of SOL and WASO.

To summarize, both actigraphy devices appear to be valid and reliable for detecting sleep/wake diurnal patterns, with some notable differences: (1) the GT3X+ showed better EBE agreements than AW-64 for all the variables examined; and (2) the AW-64 was superior for defining overall sleep parameters (i.e., SOL, TST, WASO, and SE). The latter difference in sleep parameter estimation also results in substantially better wake detection (i.e., no significant differences with PSG-derived SOL and WASO). This is likely explained by the algorithms used by each software. ActiLife derives sleep parameters as the sum of the single epochs defined as sleep or wake by the Sadeh algorithm, whereas Actiware uses a proprietary algorithm that applies PSG-derived correction factors to improve sleep statistics beyond that of just summing the sleep/wake epochs (see Actiware Software Actiwatch Instruction Manual). Our sample is composed of young adult students. This limits the generalization of our results beyond this population. Future research should test the performance of the GT3X+ using the Cole-Kripke algorithm in an older population (Cole et al., 1992). Additionally, we only recorded daytime naps, and our results cannot be extended to nighttime sleep periods. However, given the lower amounts of wake during nighttime sleep, we expect similar or better performance of both devices in nighttime recordings.

Finally, there is a growing need for devices that can concurrently measure both physical activity and sleep, hence representing the full 24-h cycle (Buman & King, 2010). The Actigraph has extensive reliability and validity evidence to measure physical activity and sedentary behavior in a variety of populations, in both laboratory and free-living settings, when fixed to the hip (Freedson et al., 1998; Hendelman et al., 2000; Kozey-Keadle et al., 2011; Trost et al., 2005). Although Actigraph has developed an algorithm to scale wrist-worn data to hip-worn estimations, there is little evidence currently available to support valid estimation of physical activity using the GT3X+ when worn on the wrist. Moreover, a recent study using the Actigraph in children has found,

for sleep estimation, that wrist and hip placements are not interchangeable (Hjorth et al., 2012). One important contribution of the present study is that a wrist-worn GT3X+ is reliable and valid to measure sleep. Therefore, it is now possible to objectively measure 24-h sleep/wake/activity behaviors if the GT3X+ is worn on the hip during the day and wrist while sleeping. Future studies should explore the feasibility and participant burden of this measurement strategy.

In conclusion, the present study compared two widely used actigraph devices to measure sleep/wake patterns relative to PSG. We further confirmed that AW-64 is a valid and reliable device for measuring sleep in daytime naps. In addition, compare with the GT3X+, the AW-64 showed a better performance on sleep statistics, likely due to the scoring algorithms used by this device. The GT3X+ showed a systematic bias when detecting daytime sleep parameters, but accuracy, sensitivity, and specificity values were similar to those in previous reports (Martin & Hakim, 2011). This suggests that the GT3X+ may be a useful device to monitor sleep/wake activity, although it is important to use this actigraph with an understanding of its limitations. However, further validation studies focusing on nighttime sleep and different populations (older adults, children, people with sleep disorders) are warranted. In addition, improving scoring algorithms in order to better assess sleep/wake transition and motionless wakefulness could increase the validity and reliability of this device. We suggest that the choice of a specific actigraph for research or clinical purposes should be based on the reliability and cost of the device, the scoring algorithm, the target population (e.g., young adults vs. children), the experimental condition (e.g., daytime vs. nighttime recording), and the aims of the study (e.g., sleep and/or physical activity).

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES

- Altman DG, Bland JM. (1983). Measurement in medicine: the analysis of method comparison studies. *Statistician*, 32, 307–17.
- Ancoli-Israel, S, Cole, R, Alessi, C, et al. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *American Academy of Sleep Medicine Review Paper*. *Sleep*, 26, 342–92.
- Benson, K, Friedman, L, Noda, A, et al. (2004). The measurement of sleep by actigraphy: direct comparison of 2 commercially available actigraphs in a nonclinical population. *Sleep*, 27, 986–9.
- Blackwell, T, Redline, S, Ancoli-Israel, S, et al. (2008). Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. *Sleep*, 31, 283–91.
- Bland, JM, Altman, DG. (1999). Measuring agreement in method comparison studies. *Stat Methods Med Res*, 8, 135–60.

- Buman, MP, King, AC. (2010). Exercise as a treatment to enhance sleep. *Am J Lifestyle Med*, 4, 500–14.
- Byrt, T, Bishop, J, Carlin, JB. (1993). Bias, prevalence and kappa. *J Clin Epidemiol*, 46, 423–9.
- Chesson Jr M, Coleman, M, Lee-Chiong, M, Pancer, D. (2007). Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep*, 30, 519–29.
- Cole, R, Kripke, D, Gruen, W, et al. (1992). Automatic sleep/wake identification from wrist activity. *Sleep*, 15, 461–9.
- de Souza, L, Benedito-Silva, A, Pires, M, et al. (2003). Further validation of actigraphy for sleep studies. *Sleep*, 26, 81–5.
- Feinstein, AR, Cicchetti, DV. (1990). High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*, 43, 543–9.
- Freedson, PS, Melanson, E, Sirard, J. (1998). Calibration of the Computer Science and Applications, Inc. accelerometer. *Med Sci Sports Exerc*, 30, 777–81.
- Hendelman, D, Miller, K, Baggett, C, et al. (2000). Validity of accelerometry for the assessment of moderate intensity physical activity in the field. *Med Sci Sports Exerc*, 32, S442–9.
- Hjorth, MF, Chaput, JP, Damsgaard, CT, et al. (2012). Measure of sleep and physical activity by a single accelerometer: can a waist-worn Actigraph adequately measure sleep in children? *Sleep Biol. Rhythms*, 10, 328–35.
- Iber, C, Ancoli-Israel, S, Chesson, A, Quan, SF. (2007). The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications. Westchester, IL: American Academy of Sleep Medicine.
- Jasper, H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol*, 10, 371–5.
- Kanady, J, Drummond, S, Mednick, S. (2011). Actigraphic assessment of a polysomnographic-recorded nap: a validation study. *J Sleep Res*, 20, 214–222.
- Kozey-Keadle, S, Libertine, A, Lyden, K, et al. (2011). Validation of wearable monitors for assessing sedentary behavior. *Med Sci Sports Exerc*, 43, 1561–7.
- Landis, JR, Koch, GG. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–74.
- Martin, JL, Hakim, AD. (2011). Wrist actigraphy. *CHEST J*, 139, 1514–27.
- Meltzer, L, Montgomery-Downs, H, Insana, S, Walsh, C. (2012a). Use of actigraphy for assessment in pediatric sleep research. *Sleep Med Rev*, 16, 463–75.
- Meltzer, L, Walsh, C, Traylor, J, Westin, A. (2012b). Direct comparison of two new actigraphs and polysomnography in children and adolescents. *Sleep*, 35, 159–66.
- Milner, CE, Cote, KA. (2009). Benefits of napping in healthy adults: impact of nap length, time of day, age, and experience with napping. *J Sleep Res*, 18, 272–81.
- Moser, D, Anderer, P, Gruber, G, et al. (2009). Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. *Sleep*, 32, 139–49.
- Nishida, M, Walker, MP. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. *PLoS ONE*, 2, e341.
- Paquet, J, Kawinska, A, Carrier, J. (2007). Wake detection capacity of actigraphy during sleep. *Sleep*, 30, 1362–9.
- Philips Respironics. (2005). Actiware software: Actiwatch instruction manual, software version 5.52.0003. Bend, Oregon: Mini Mitter Company.
- Philips Respironics. (2008). *Equivalence of activity recording and derived sleep statistics*. Koninklijke Philips Electronics N.V. http://actigraphy.respironics.com/downloads/AWDataEquivalenceReport_0309_final.pdf. Accessed May 14, 2013, 1–6.
- Pollak, C, Stokes, P, Wagner, D. (1998). Direct comparison of two widely used activity recorders. *Sleep*, 21, 207–12.
- Portaluppi, F, Smolensky, MH, Touitou, Y. (2010). Ethics and methods for biological rhythm research on animals and human beings. *Chronobiol Int*, 27, 911–1929.
- Poyurovsky, M, Nave, R, Epstein, R, et al. (2000). Actigraphic monitoring (actigraphy) of circadian locomotor activity in schizophrenic patients with acute neuroleptic-induced akathisia. *Eur Neuropsychopharmacol*, 10, 171–6.
- Rechtschaffen, A, Kales, A. (1968). A manual of standardized terminology, techniques, and scoring system for sleep stages of human participants. Los Angeles, CA: University of California, Brain Information Service/Brain Research Institute.
- Ried-Larsen, M, Brønd, JC, Brage, S, et al. (2012). Mechanical and free living comparisons of four generations of the Actigraph activity monitor. *Int J Behav Nutr Phys Act*, 9, 113. <http://link.springer.com/article/10.1186%2F1479-5868-9-113>.
- Robusto, KM, Trost, SG. (2012). Comparison of three generations of ActiGraph™ activity monitors in children and adolescents. *J Sports Sci*, 30, 1429–35.
- Rowlands, A, Stiles, V. (2012). Accelerometer counts and raw acceleration output in relation to mechanical loading. *J Biomech*, 45, 448–54.
- Rupp, TL, Balkin, TJ. (2011). Comparison of Motionlogger Watch and Actiwatch actigraphs to polysomnography for sleep/wake estimation in healthy young adults. *Behav Res Methods*, 43, 1152–60.
- Sadeh, A. (2008). Commentary: comparing actigraphy and parental report as measures of children's sleep. *J Pediatr Psychol*, 33, 406–7.
- Sadeh, A. (2011). The role and validity of actigraphy in sleep medicine: an update. *Sleep Med Rev*, 15, 259–67.
- Sadeh, A, Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Med Rev*, 6, 113–24.
- Sadeh, A, Alster, J, Urbach, D, Lavie, P. (1989). Actigraphically based automatic bedtime sleep-wake scoring: validity and clinical applications. *J Ambulat Monit*, 2, 209–16.
- Sadeh, A, Sharkey, K, Carskadon, M. (1994). Activity-based sleep-wake identification: an empirical test of methodological issues. *Sleep*, 17, 201–7.
- Sadeh, A, Hauri, PJ, Kripke, DF, Lavie, P. (1995). The role of actigraphy in the evaluation of sleep disorders. *Sleep*, 18, 288–302.
- Tonetti, L, Pasquini, F, Fabbri, M, et al. (2008). Comparison of two different actigraphs with polysomnography in healthy young participants. *Chronobiol Int*, 25, 145–53.
- Trost, SG, McIver, KL, Pate, RR. (2005). Conducting accelerometer-based activity assessments in field-based research. *Med Sci Sports Exerc*, 37, S531–43.
- Tucker, MA, Hirota, Y, Wamsley, EJ, et al. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiol Learn Mem*, 86, 241–7.
- Weiss, A, Johnson, N, Berger, N, Redline, S. (2010). Validity of activity-based devices to estimate sleep. *J Clin Sleep Med*, 6, 336–42.