doi: 10.1111/j.1365-2869.2010.00858.x

Actigraphic assessment of a polysomnographic-recorded nap: a validation study

JENNIFER C. KANADY[1,3,5], SEAN P. A. DRUMMOND[1,2,4] and SARA C. MEDNICK[1,3]

[1]University of California, San Diego – Department of Psychiatry, [2]Veterans Affairs, San Diego Healthcare System – Psychology Service, [3]Veterans Affairs, San Diego Healthcare System – Research Service, [4]SDSU/UCSD Joint Doctoral Program in Clinical Psychology and [5]University of California, Berkeley – Department of Clinical Science, Berkeley, CA, USA

Accepted in revised form 24 April 2010; received 05 November 2009

SUMMARY This study aimed to determine if actigraphy could differentiate sleep and wake during a daytime nap and no-nap rest period. Fifty-seven subjects participated in the study; 30 subjects were in the nap group and the remaining 27 in the no-nap comparison group. All subjects wore actigraphs while simultaneously undergoing polysomnography (PSG). Three actigraphic sensitivity levels (high, medium, low) and two interval duration minimums (15 and 40 min) were used to score the nap and no-nap data. The variables examined included total sleep time (TST), sleep latency (SL), wake after sleep onset (WASO) and sleep efficiency (SE). The Bland–Altman technique was used to determine concordance. Epoch-by-epoch analysis examined actigraphic accuracy, sensitivity and specificity. For the naps, all actigraph settings except low-40 showed significant correlations with TST. The high and medium settings predicted SE significantly and the high settings predicted SL significantly. Bland–Altman analyses demonstrated high settings overestimated TST while high and medium settings overestimated SE. Overall, for the nap condition accuracy for the actigraph was 82–86%, sensitivity was 92–96% and specificity was 40–67%. In the no-nap condition, accuracy for the actigraph was 60–84%, sensitivity was 47–78% and specificity was 60–86%. Medium-40 and low-40 were the only settings that did not misidentify sleep in the no-nap condition. These results suggest that actigraphy can predict TST, SE and SL reliably, depending upon parameter settings, and actigraphy is a highly sensitive but not specific measure for daytime naps. Different actigraphy settings may be optimal depending upon the variables of interest. Discrimination of sleep and wake during periods of waking quiescence is not as robust as during periods of mainly daytime sleep.

KEYWORDS actigraphy, napping, polysomnography, validation

INTRODUCTION

Polysomnography (PSG), currently considered the gold standard for sleep measurement, is none the less relatively cumbersome and expensive for ambulatory monitoring. Therefore, PSG is not ideal for longitudinal and naturalistic examination. Actigraphy, on the other hand, is a portable device usually worn on the wrist or ankle and may serve as a suitable substitute for PSG, particularly when monitoring sleep over extended periods of time. Although actigraphy does not have the same capabilities as PSG, there is a vast amount of literature demonstrating the ability of the actigraph to accurately distinguish sleep versus wake when compared to PSG (Acebo and LeBourgeois, 2006; Ancoli-Israel *et al.*, 2003; Morgenthaler *et al.*, 2007). Numerous validation studies have supported the use of actigraphs in lieu of PSG in several different populations including: healthy adolescents (Johnson

Correspondence: Sean P. A. Drummond PhD, Laboratory of Sleep and Behavioral Neuroscience, UCSD and VA San Diego Healthcare System, 3350 La Jolla Village Dr. MC 151B, Building 13, 3rd floor, San Diego, CA 92161, USA. Tel: 858 642 1274; fax: 858 458 4201; e-mail: Drummond@ucsd.edu

et al., 2007; Sadeh *et al*., 1994), younger (Blood *et al*., 1997; Cole *et al*., 1992; De Souza *et al*., 2003; Jean-Louis *et al*., 1996; Jean-Louis *et al*., 1997; Monk *et al*., 1999; Paquet *et al*., 2007) and older adults (Blackwell *et al*., 2008). Actigraphy has also been validated in clinical populations such as insomnia (Friedman *et al*., 2000; Jean-Louis *et al*., 1997, 1999; Lichstein *et al*., 2006), major depressive disorder (Jean-Louis *et al*., 2000), dementia (Ancoli-Israel *et al*., 1997) and sleep-disordered breathing (Hyde *et al*., 2007; Johnson *et al*., 2007; Kushida *et al*., 2001), although it is often not as reliable in clinical samples as in healthy adults. These compelling results suggest actigraphy may be an economic and efficient technique for monitoring sleep in various ambulatory settings. However, these validation studies focus primarily upon longer, nocturnal sleep periods. The use of actigraphy for shorter periods of daytime sleep, or daytime naps, is not as well studied.

The ability of actigraphy to detect daytime sleep would be useful in both healthy and clinical populations. The National Sleep Foundation's 2008 *Sleep in America Poll* (2009; http://sleepfoundation.org/article/sleep-america-polls/2009-health-and-safety) recently found 46% of the population reports napping on a regular basis (two or more times per month). While many studies utilize actigraphy to screen subjects and/or verify compliance with a specified sleep schedule, they rarely report information on daytime sleep. This is due, in part, to a lack of validation studies establishing the ability of actigraphy to detect daytime sleep reliably. Thus, it is possible these studies lose important information related to daytime sleep in their subjects. Actigraphic nap detection would also be beneficial in clinical and operational settings where daytime sleepiness is common [e.g. insomnia (Lichstein *et al*., 1994; Moul *et al*., 2002 and sleep apnea (John, 1993; Roehrs *et al*., 1989)] and/or poses significant risk. For example, actigraphy may be able to detect operational environments, work schedules and/or individuals who are more prone to falling asleep when it would be unsafe to do so. Additionally, an accurate measure of daytime sleep may help better assess treatment compliance, e.g. in insomnia when recommending sleep restriction or no naps. Studies examining the accuracy of self-report data have shown sleep diaries are often unreliable (Bradshaw *et al*., 2007; Carney *et al*., 2004). Thus, an accurate, objective measure of daytime sleep would be valuable for both clinical and research endeavors.

Prior validation studies have found that the most effective determination of actigraphy and PSG concordance employs multiple approaches to the analysis, as each can convey different and complementary information. For example, correlation coefficients alone can be insufficient and much more information can be attained when the principles of sensitivity, specificity and accuracy are applied (De Souza *et al*., 2003). Sensitivity reflects the ability of actigraphy to detect sleep; specificity is the ability of actigraphy to detect wakefulness; and accuracy is the ability of actigraphy to detect both sleep and wakefulness when compared to PSG (Tilmanne *et al*., 2009). The Bland–Altman technique is another useful technique that plots the difference between actigraphy and PSG against their means and is often used to determine concordance and/or the direction of discordance (Bland and Altman, 1995). Additionally, Gale *et al*. demonstrated that any of these measurements alone might be misleading without the kappa statistic, which is useful in determining the amount of agreement that might be expected by chance (Gale *et al*., 2005). The utilization of all these techniques together would provide a great deal of information as to the capability and limitations of actigraphy to detect short daytime sleep episodes.

The present study aims to determine if actigraphy can detect accurately sleep in healthy, young adults during a 90-min mid-afternoon nap opportunity when compared to PSG. We used the automatic minor rest interval (AMRI) of the Respironics Actiware program (Bend, OR, USA) to determine if the actigraph and scoring algorithm can predict sleep reliably during a nap. The sleep variables of interest were total sleep time (TST), wake after sleep onset (WASO), sleep efficiency (SE) and sleep latency (SL). We also wished to establish the ideal parameters for detecting a nap and used several different AMRI settings including high, medium and low sensitivity for 15- and 40-min interval duration minimums. After establishing significant AMRI settings, we used the Bland–Altman technique to determine if the actigraph over- or underestimates each significant sleep variable when compared to PSG. Next, epoch-by-epoch analysis was performed on all feasible actigraphic records in order to access accuracy, sensitivity and specificity. Finally, the kappa statistic was used to determine the amount of agreement that could be attributed to chance. In a further attempt to test the AMRI settings, we also examined whether the actigraph can discriminate wake from sleep during a no-nap comparison group, which consists of a period of quiet rest (no sleep).

METHODS

Subjects

Participants were between the ages of 18 and 35 years and had to be both physically and psychologically healthy as assessed by an extensive phone screen. Exclusion criteria included self-reported personal or familial diagnosis of a DSM-IV Axis I disorder, lifetime use of psychotropic medications, serious neurological and medical disorders, a loss of consciousness exceeding 2 min, learning disabilities and mental retardation, sleep disorders such as insomnia and sleep apnea, the consumption of more than 300 mg of caffeine per day, current drug use, a body mass index (BMI) exceeding 30 and individuals exhibiting an erratic sleep–wake schedule (e.g. shift workers). All eligible subjects gave their written informed consent as outlined by the University of California, San Diego Human Research Protections Program. Fifty-seven subjects completed the study. Thirty subjects (21 female, average age: 20.4 ± 2.8 years, education: 14.4 ± 2.3 years) were in the nap group and the remaining 27 (25 female, average age: 20.1 ± 1.8 years, education: 14.7 ± 2.5 years) were in the no-nap group (period of rest, no sleep). An epoch-by-epoch analysis

was performed on 19 (12 female, age: 19.7 ± 1.5 years, education: 14.3 ± 2.3 years) of the 30 subjects in the nap group and 19 (12 female, age: 19.9 ± 1.6 years, education: 14.7 ± 2.3 years) of the 27 subjects in the no-nap group.

Protocol

Participants kept a regular sleep–wake schedule the week prior to the study (> 6.5 h/night) and were asked to refrain from caffeine and alcohol 24 h before the study and throughout the experimental day. On the experimental day, all subjects wore an actigraph while attached simultaneously to electrodes using standard polysomnographic recording procedures. All subjects were in bed, ready to nap by 13:30 h. Participants in the nap group were allowed to sleep a maximum of 90 min (average nap: 66.7 ± 21.2 min), but were given no more than 120 min in bed. Participants in the no-nap group relaxed in a comfortable chair while listening to classical music and were instructed not to sleep. Subjects were asked to press the event marker on the actigraph exactly at 'lights out' and again at 'lights on' in order to synchronize the PSG and actigraph records. Trained PSG technologists monitored the recordings continuously and tracked total sleep time online to limit sleep to a maximum of 90 min and to prevent sleep in the no-nap group. If a technologist noticed a no-nap subject begin to enter Stage 1 sleep, the subject's name was called over an intercom to initiate wake and the subject was instructed to keep his/her eyes open.

Collecting and scoring actigraphy

We used the Actiwatch-64 actigraph (Respironics) and all watches were configured to collect data in 1-min epochs. Actigraphy data were scored using AMRI detection of the Respironics Actiware 5.52.0003 program. The AMRI is a feature of the scoring program that automatically creates 'rest intervals' during periods when the participant appears to be napping (sleep periods shorter than 3 h). The sleep variables examined were TST, WASO, SE and SL. We examined three AMRI sensitivity levels (high, medium, low) and two interval duration minimums (15 and 40 min). For the Respironics scoring program examined here, 'sensitivity' refers to the sensitivity to detect minor rest intervals (i.e. not the sensitivity to detect sleep, as the term is used typically in actigraphy). A high-sensitivity setting is more sensitive to immobility and would detect more minor rest intervals and a low-sensitivity setting, being less sensitive to immobility, would detect fewer minor rest intervals. The interval duration minimums indicate the shortest allowable rest interval to be detected. Therefore a 15-min interval duration minimum would search for sleep only within any minor rest interval greater than or equal to 15 min and a 40-min interval duration minimum would only score any minor rest interval greater than or equal to 40 min. We did not manipulate the wake threshold of the algorithm, and once the rest intervals were established we used the default setting of medium 40 to distinguish sleep versus wake within the various rest intervals.

For all records with event markers, a separate minor rest interval was created manually to start and end at the event markers. The manual rest interval was created in order to conduct an epoch-by-epoch (minute-by-minute) comparison to PSG to access actigraphic accuracy, sensitivity, and specificity as well as the kappa statistic. High-, medium- and low-sensitivity settings were examined. No interval duration minimums were needed due to the fact that event markers determined the beginning and end of the rest interval.

Collecting and scoring polysomnography

All PSG data were collected using Astro-Med Grass Heritage Model 15 amplifiers with Grass Gamma software. Scalp electroencephalogram (EEG) and electro-oculogram (EOG) electrodes were referenced to unlinked opposite mastoids and submental muscle tone EMGs were attached under the chin. The low-frequency filters were set at 0.3 Hz and the high-frequency filters at 100 Hz for all EEGs and EOGs. A 60-Hz notch filter was also utilized to eliminate potential background noise. At the beginning of each recording, an internal 50-μV calibration signal was generated followed by impedance checks and biocalibrations. EEG data were digitized at a sampling rate of 256 Hz and were imported to Pass Plus waveform analysis software (Delta Software, St Louis, MO, USA; http://www.deltapass.com) and scored visually in 30-s epochs according to Rechtschaffen and Kales sleep staging criteria (Rechtscaffen and Kales, 1968). The head technologist of the UCSD General Clinical Research Center (GCRC)'s Christian Gillin Laboratory for Sleep and Chronobiology scored all records; she is a Registered Polysomnographic Technologist (RPSGT) and maintains the gold standard for training of other technologists in the GCRC and has an intrascorer reliability of 0.90. The variables examined from these PSG data were TST, WASO, SL and SE. Additionally, 1-min epochs were scored visually to determine sleep versus wake in order to conduct an epoch-by-epoch comparison to actigraphy. If an epoch consisted of both sleep and wake, we scored sleep or wake depending on which comprised the majority of the epoch.

Statistical analyses

Nap group

In the nap group, in order to determine the relationship between AMRI settings and PSG, linear regressions with Bonferroni corrections were performed for each sleep variable (i.e. six regressions per sleep variable, with significance set at $P \leq 0.008$). After determining significant AMRI settings, concordance was then examined using the Bland–Altman technique for all significant AMRI settings.

Bland–Altman technique

The Bland–Altman technique plots the difference between actigraphy and PSG (actigraphy minus PSG) against the

average of actigraphy and PSG for each sleep variable to determine whether there is a bias in actigraphy (Bland and Altman, 1995). The actigraphy bias is represented as the mean difference between actigraphy and PSG with a negative mean difference representing an underestimation and a positive mean difference representing an overestimation. The upper and lower limits based on 95% confidence intervals were used to determine the significance of the mean difference.

Accuracy, sensitivity and specificity

An epoch-by-epoch analysis was conducted for 19 of the 30 nap records in order to determine accuracy, sensitivity and specificity. We were only able to analyze 19 of the subjects because the remaining 11 failed to use the event marker as requested or the PSG technologist forgot to instruct the participant to use the event marker. We used the definitions set forth by Tilmanne *et al*. to calculate accuracy, sensitivity and specificity (Tilmanne *et al*., 2009). When compared to PSG, a true positive (TP) indicates that the actigraph identifies sleep correctly, a true negative (TN) indicates that the actigraph correctly identifies wake, a false negative (FN) indicates that the actigraph misidentifies wake and a false positive (FP) indicates that the actigraph misidentifies sleep. Accuracy is then defined as (TP + TN)/(TP + TN + FN + FP) and represents the agreement rate between PSG and actigraphy; sensitivity is defined as TP/(TP + FN) and represents the percentage of epochs identified correctly as sleep; and specificity is defined as TN/(TN + FP) and represents the percentage of epochs identified correctly as wake. Accuracy, sensitivity and specificity were examined for high-, medium- and low-sensitivity settings.

Kappa statistic

The epoch-by-epoch analysis was also used to calculate the kappa statistic. The kappa statistic accounts for the amount of agreement expected by chance (Cohen, 1960). To calculate kappa, a ratio is formed between the chance-corrected observed agreement and the chance-corrected perfect agreement (Feinstein and Cicchetti, 1990). The kappa statistic ranges from 1, which demonstrates a perfect agreement, to 0, which demonstrates agreement based on chance alone, to –1, which demonstrates complete disagreement. In this study, we calculated kappa for high-, medium- and low-sensitivity settings.

No-nap comparison group

In the no-nap group, we examined the percentage of rest intervals that the AMRI misidentified as sleep (rest intervals with epochs of actigraphic-scored sleep/total number of rest intervals). Any no-nap rest intervals with PSG-identified sleep were excluded from this analysis. A total of 27 no-nap periods were examined. Additionally, an epoch-by-epoch analysis was conducted on 19 of the 27 no-nap subjects in order to access actigraphic accuracy, sensitivity and specificity for the no-nap period of quiet rest. Again, we were only able to examine 19 of the records because of event marker non-compliance. The same definitions and calculations of accuracy, sensitivity and specificity used in the nap group were also used in the no-nap group. In these 19 subjects, the percentage of epochs misidentified as sleep out of the total number of epochs was also examined (epochs misidentified as sleep/total number of epochs).

RESULTS

Nap group

Table 1 provides the sleep variables of interest as scored by the AMRI settings and PSG. Table 2 provides the correlations between actigraphy and PSG scored sleep variables. After performing Bonferroni corrections, all AMRI settings except for low-40 significantly predicted TST when compared to PSG, with the high-sensitivity setting having the strongest correlations. Both high- and medium-sensitivity settings predicted SE significantly, and high-sensitivity settings predicted SL significantly. WASO was not significant for any of the AMRI settings.

Bland–Altma technique

Results are presented in Table 3. Although the high-sensitivity setting had the strongest correlation with PSG, the Bland–Altman technique demonstrates that the high-sensitivity settings overestimated TST significantly. Based on confidence intervals, the setting of high-15 overestimated TST by 2.8–10.3 min, and the high-40 setting overestimated TST by 2.3–10.0 min. The high-sensitivity setting also overestimated SE significantly: high-15 overestimated SE 3.6–17.0% and high-40

Table 1 Sleep variable averages

	High-15	*High-40*	*Med-15*	*Med-40*	*Low-15*	*Low-40*	*PSG*
TST (min)	72.0 ± 22.8	73.8 ± 21.2	66.9 ± 21.9	65.9 ± 19.5	62.2 ± 19.5	58.5 ± 17.0	67.6 ± 21.0
WASO (min)	10.5 ± 8.8	10.3 ± 8.9	12.3 ± 18.7	3.3 ± 3.5	6.0 ± 6.7	1.7 ± 1.7	7.7 ± 5.3
SE (%)	79.2 ± 19.2	80.0 ± 19.0	82.0 ± 18.7	92.1 ± 10.7	90.7 ± 8.3	96.2 ± 4.5	77.7 ± 13.6
SL (min)	7.6 ± 12.1	7.7 ± 12.3	2.5 ± 4.6	2.0 ± 4.8	0.4 ± 1.2	0.5 ± 1.4	11.9 ± 9.6

PSG: polysomnography; SE: sleep efficiency; SL: sleep latency; TST: total sleep time; WASO: wake after sleep onset.

Table 2 Correlation between ACT and PSG in the Nap Group

AMRI	*TST*	*WASO*	*SE*	*SL*
High-15	***R* = 0.870, *P* ≤0.001**	*R* = −0.137, *P* = 0.495	***R* = 0.658, *P* ≤0.001**	***R* = 0.726, *P* ≤0.001**
High-40	***R* = 0.834, *P* ≤0.001**	*R* = −0.138, *P* = 0.491	***R* = 0.658, *P* ≤0.001**	***R* = 0.726, *P* ≤0.001**
Med-15	***R* = 0.778, *P* ≤0.001**	*R* = 0.230, *P* = 0.248	***R* = 0.588, *P* ≤0.001**	*R* = 0.319, *P* = 0.104
Med-40	***R* = 0.661, *P* ≤0.001**	*R* = 0.475, *P* = 0.022	***R* = 0.809, *P* ≤0.001**	*R* = 0.421, *P* = 0.045
Low-15	***R* = 0.632, *P* ≤0.001**	*R* = 0.076, *P* = 0.708	*R* = 0.183, *P* = 0.360	*R* = 0.185, *P* = 0.356
Low-40	*R* = 0.336, *P* = 0.113	*R* = –0.096, *P* = 0.695	*R* = 0.193, *P* = 0.429	*R* = 0.396, *P* = 0.094

AMRI: automatic minor rest interval; SE: sleep efficiency; SL: sleep latency; TST: total sleep time; WASO: wake after sleep onset.

Table 3 Bland–Altman statistics for significant automatic minor rest interval (AMRI) settings (grouped by sensitivity)

AMRI Setting	*High-15 TST*	*High-15 SE*	*High-15 SL*	*High-40 TST*	*High-40 SE*	*High-40 SL*
Mean difference	**6.55**	**10.28**	−3.13	**6.16**	**8.71**	−3.38
Standard deviation	**10.57**	**18.76**	12.64	**10.53**	**17.07**	12.79
Upper limit	**10.33**	**16.99**	1.39	**9.99**	**14.92**	1.27
Lowerl limit	**2.77**	**3.57**	−7.65	**2.33**	**2.50**	−8.03
Over- or Underestimation	**Over**	**Over**		**Over**	**Over**	

AMRI setting	*Med-15 TST*	*Med-15 SE*	*Med-40 TST*	*Med-40 SE*
Mean difference	1.22	**13.55**	−5.27	**17.00**
Standard deviation	13.87	**18.37**	15.27	**16.45**
Upper limit	6.18	**20.12**	0.74	**23.58**
Lower limit	−3.74	**6.98**	−11.28	**10.42**
Over- or underestimation		**Over**		**Over**

AMRI setting	*Low-15 TST*
Mean difference	**−7.84**
Standard deviation	**16.55**
Upper limit	**−1.71**
Lower limit	**−11.28**
Over- or underestimation	**Under**

The mean difference (estimated bias of actigraphy), standard deviation (fluctuation around mean) and upper and lower limits (95% confidence interval) for all significant AMRI settings. Bold type indicates significance and the last row states whether the AMRI setting significantly over- or underestimates the particular sleep variable. Here, a non-significant result is important, because it means that actigraphy does not systematically mis-score sleep/wake in a particular direction.

overestimated SE 2.5–15.0%. The high-sensitivity setting did not over- or underestimate SL significantly.

The medium-sensitivity settings correspond better with PSG than do the high settings for TST, neither significantly over- nor underestimating sleep. However, the medium-sensitivity settings significantly overestimated SE: the medium-15 setting overestimated SE from 7.0 to 20.1% and the medium-40 setting overestimated SE from 10.4 to 23.6%.

Table 4 Accuracy, sensitivity and specificity percentages for automatic minor rest interval (AMRI) sensitivity settings (high, medium, low) in the nap group

	High	*Med*	*Low*
Accuracy (%)	82.74	85.05	86.20
Sensitivity (%)	96.48	94.61	92.48
Specificity (%)	40.05	55.34	66.66

Accuracy, sensitivity and specificity

Results are presented in Table 4. Accuracy, sensitivity and specificity values for the high, medium and low settings are comparable to those found in nocturnal validation studies (Table 1). Accuracy values are high, ranging from 82.7 to 86.2%. Sensitivity values, also extremely high, range from 92.5 to 96.5%. Specificity values are lower, ranging from 40.1 to–66.7%. The low-sensitivity setting is the most accurate and specific of the settings, and the high-sensitivity setting is most sensitive.

Kappa statistic

The kappa statistic demonstrates, as might be expected, that some of the agreement between actigraphy and PSG may be attributed to chance. The high sensitivity setting had the lowest kappa (κ = 0.42), followed by the medium setting (κ = 0.54), and lastly the low setting (κ = 0.65).

No-nap comparison group

Despite careful monitoring by PSG technologists, six no-nap records contained brief periods of PSG-defined Stage 1 sleep; therefore only 27 records were examined. These brief periods of sleep comprised between 5 and 8% of each sleep record. Of the remaining 27 no-nap records, the AMRI detected between 0 and 22 rest intervals depending on the AMRI setting (a single record can have multiple rest intervals should periods of quiescence be interrupted by periods of greater movement). Four of the six AMRI settings misidentified at least one epoch of sleep within at least one of these rest intervals, with the high-sensitivity setting misidentifying sleep most frequently. Results are presented in Table 5. Additionally, actigraphic accuracy, sensitivity and specificity were examined in 19 of the no-nap records to determine whether the actigraph could distinguish sleep versus wake accurately during a period of quiet restfulness that contains little to no sleep. Results are presented in Table 6. Overall, accuracy and sensitivity values are lower and specificity values are higher when compared to the values of the nap group. The percentage of epochs misidentified as sleep out of the total number of epochs for high-, medium- and low-sensitivity settings were also examined. The high-sensitivity setting misidentified sleep in 39.0% of the epochs, medium misidentified sleep in 22.1% of the epochs, and low misidentified sleep in 13.8% of the epochs.

DISCUSSION

This study examined the ability of a specific automatic actigraphy-scoring algorithm to distinguish sleep and wake during daytime naps and no-naps rest periods. The data demonstrate that the automatic minor rest interval is a useful tool for detecting daytime sleep in healthy, young adults. Depending upon the parameter settings, actigraphy was able to predict TST, SL and SE significantly during a nap when compared to PSG. Actigraphy tended to overestimate sleep during a nap; however, accuracy, sensitivity and specificity values are high, with specificity values surpassing those found in nocturnal validation studies (Blood *et al*., 1997; De Souza *et al.,* 2003; Hyde *et al*., 2007; Kushida *et al*., 2001; Lichstein *et al*., 2006). Furthermore, actigraphy was fairly proficient in distinguishing the difference between a nap and a no-nap period of quiet rest. Although actigraphy overestimated sleep during the no-nap condition, the accuracy values remained reasonably high, particularly with the low-sensitivity setting (84%). The results of this study also indicate that different actigraphic sensitivity settings should be used depending upon the sleep parameters of interest. This study demonstrates that high-sensitivity settings are best for determining sleep and low-sensitivity settings are best for determining wake during both nap and periods of quiet rest.

Although the correlation coefficients between PSG and actigraphy for daytime naps are significant, they are not consistent with the effect sizes of previous validation studies examining nocturnal sleep (Table 1). Prior literature has demonstrated TST correlation coefficients between 0.91 and 0.98 when comparing actigraphy and PSG in healthy, younger adults (Gale *et al*., 2005; Jean-Louis *et al*., 1996; Kripke *et al*., 1978). However, as pointed out by De Souza *et al*., correlation coefficients have limitations and are not sufficient for determining concordance (De Souza *et al*., 2003). The added value of concordance measures is demonstrated in our results. Although high-sensitivity settings had the strongest correlation with PSG, the Bland–Altman technique demonstrated the high-sensitivity settings overestimated significantly both TST and SE. Conversely, the medium-sensitivity setting, which did not correlate as strongly with PSG, did not over- or underestimate TST significantly. Medium-sensitivity, however, overestimated SE to approximately the same extent as the high settings, and unlike the high settings did not correlate significantly with SL.

The examination of accuracy, sensitivity and specificity demonstrate that although a highly sensitive measure, actigraphy is not as specific and thus may underscore wake during relatively fragmented daytime sleep. This point is strengthened by the fact that actigraphy did not correlate significantly with PSG-defined WASO and overestimated TST and SE consistently during the nap. These results are consistent with other validation studies that have found actigraphy to be a highly sensitive, but not specific measure during periods of mainly sleep (Blood *et al*., 1997; De Souza *et al*., 2003; Hyde *et al*., 2007; Kushida *et al.,* 2001; Lichstein *et al*., 2006) and highlight one potential weakness of relying solely on sensitivity/specificity measures (De Souza *et al*., 2003). Paquet *et al*. found a significant decrease in actigraphic accuracy with increased wakefulness and that actigraphy overestimated TST and SE more strongly in conditions involving more wake (Paquet

Table 5 Percentage of intervals scored as sleep in the no-nap comparison group

AMRI	*% Scored as sleep*
High-15	14.80%
High-40	11.10%
Med-15	7.40%
Med-40	0.00%
Low-15	3.70%
Low-40	0.00%

In the no-nap group, the percentage of rest intervals with no polysomnography-defined sleep that each automatic minor rest interval setting scored as having at least one epoch of sleep.

Table 6 Accuracy, sensitivity and specificity percentages for automatic minor rest interval (AMRI) sensitivity settings (high, medium, low) in the no-nap comparison group

	High	*Med*	*Low*
Accuracy (%)	60.20	76.58	84.37
Sensitivity (%)	77.78	62.22	46.66
Specificity (%)	59.58	77.09	85.69

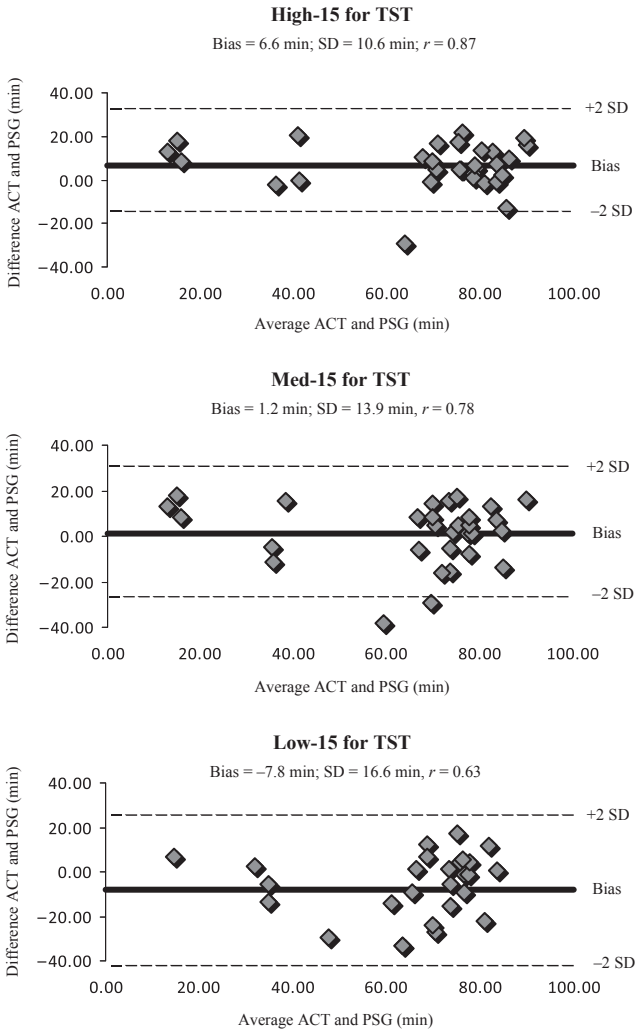

Figure 1. Bland-Altman Graphs for TST. Plotted difference against average of ACT and PSG for TST, with 95% limits of agreement. One representative graph is displayed for each sensitivity setting. Mean bias, standard deviation, and correlation between the two measures are presented. A positive mean bias represents an actigraphic overestimation and a negative mean bias an underestimation. The high sensitivity setting overestimates TST; the low sensitivity setting underestimates TST; and the medium sensitivity setting neither over or underestimates TST.

et al., 2007). Additionally, in a comparison study looking at actigraphy and diaries, Kawada found sleep diaries more accurate than actigraphy in determining daytime wake episodes (Kawada, 2008). This is due probably to the fact that periods of increased immobility, such as quietly reading a book or watching television, will be misidentified as sleep by the actigraph. Our results, taken together with others in the literature, indicate that actigraphy is a useful measure for detecting sleep during a nap. Indeed, actigraphic specificity found here during a nap was higher than studies examining nocturnal sleep (our data suggests a specificity as high as 66%). This may be due to the fact that there are fewer absolute minutes of wake (WASO = 7.7 ± 5.3) during these naps than longer periods of nocturnal sleep.

The kappa statistics found here were rather low (0.42, 0.54, 0.60), but still higher than those found in nocturnal validation studies ($\kappa = 0.399$) (Gale *et al.*, 2005), suggesting that there is less random overlap between actigraphy and PSG for naps than there is for nocturnal sleep. These low kappa values are not surprising, however, given that these intervals contain mainly sleep and the probability of overlap is thus naturally high. Additionally, as noted by Gale *et al.*, the kappa statistic is intended primarily for comparison of symmetrical measures, and in this case PSG is regarded as definitive and actigraphy an approximation (Feinstein and Cicchetti, 1990; Gale *et al.*, 2005). Therefore, it is not unusual to find high agreement rates but low kappa values in validation studies.

Prior validation studies examining nocturnal sleep have found that low-sensitivity settings are ideal for determining sleep and medium and high settings are best for detecting wake (Hyde *et al.*, 2007). However, for an afternoon nap we found the opposite to be true. This is due probably to different uses of the term 'sensitivity'. In many actigraphy studies, sensitivity refers to the likelihood of detecting sleep, while sensitivity refers here to the likelihood of detecting a minor rest interval (within which sleep will be scored). Thus, it makes sense here that sensitivity was greatest in the high settings and specificity was greatest in the low settings. This suggests that investigators may want to choose settings for the AMRI scoring algorithm based on the outcome variables of interest. For example, in addition to having the highest sensitivity, the high settings had the strongest correlations with TST and SE, were the only settings to correlate with SL and did not systematically over- or underestimate SL. On the other hand, they systematically overestimated TST (by ~3–10 min) and misidentified wake as sleep during the no-nap rest periods. Thus, investigators may wish to utilize the high settings when they are most interested in capturing any possible sleep or estimating SL in an ambulatory setting. Investigators concerned with quantifying TST more accurately and not misidentifying wake as sleep may wish to choose the medium settings. Finally, the low setting showed the greatest specificity. This may be because the only periods identified as sleep by the low settings were those with the least amount of movement and thus most likely to correspond to PSG defined sleep. These data suggest that the low setting may be useful when the main concern is not misidentifying wake as sleep. Alternatively, the Youden index provides investigators with a way to balance sensitivity and specificity if both are of equal concern (Youden, 1950). Finally, it should be noted that these suggestions are not meant to imply that investigators should use different algorithm settings for sleep versus wake periods within a single analysis. Rather, one setting should be selected for an entire analysis, but the selection of the setting may depend on the goal of the analysis (e.g. identify any possible sleep versus not overscore sleep).

The findings from the no-nap group appear more equivocal but are an informative adjunct to the nap group data. On one hand, the AMRI algorithm misidentified sleep in as many as 14.8% of the rest intervals containing no PSG-defined sleep.

Additionally, when the no-nap periods were scored manually, actigraphic accuracy and sensitivity were comparatively lower in the no-nap group than the nap group and actigraphy misidentified sleep in 13.8–39.0% of the wake epochs. On the other hand, specificity was relatively higher during these rest periods than during naps. The differences in accuracy, sensitivity and specificity can be attributed to the fact that these no-nap periods contained little to no sleep. Participants in the no-nap group were asked to relax and remained relatively immobile for the 90-min period. Therefore, given that there was little activity and these rest periods consisted of mainly wake, if the actigraph were to err it would err on the side of over scoring sleep. Over-scoring sleep is perhaps even more likely in the specific no-nap condition utilized here than what might be typical, as subjects were prohibited explicitly from reading, watching TV and other activities that may promote more movement. Thus, this no-nap condition represents one of the hardest tests of the ability for the algorithm to discriminate sleep from wake.

Similar to the nap data, there are changes in accuracy, sensitivity and specificity values when applying high, medium and low settings for the no-nap data. These changes mirror those found in the nap data and provide additional information for optimal actigraphic settings. As found in the nap data, sensitivity is highest in the high setting, but accuracy and specificity values are low and the high setting misidentified sleep in as many as 39% of the wake epochs. Sensitivity is lowest when using the low setting; however, accuracy and specificity values are high and the low setting misidentifies sleep in only 13.8% of the wake epochs. Therefore, it appears as though the high setting is optimal when an investigator is interested in identifying all sleep epochs; however, there is a high risk of over-scoring sleep. The low setting is most accurate and should be used when trying to correctly identify wake. It is important to note that these data are derived from the no-nap periods that were scored manually. When using the AMRI to score no-nap periods without manual input, investigators should take into account that high settings misidentified sleep in as many as 14.8% of the no-nap rest intervals. The medium-40 and low-40 settings of the AMRI did not misidentify sleep at all; however, one risks losing the identification of actual sleep epochs as sensitivity drops as one moves from the high to low settings.

Although the use of event markers is useful for determination of bedtime, wake time and time in bed, this is not always feasible. Often, subjects are not compliant and forget to press the event marker as requested (as is evident in a proportion of our subjects). Therefore, it is more beneficial to assess the ability of an algorithm to score daytime sleep automatically. Overall, this study suggests that actigraphy, specifically the automatic minor rest interval of the Respironics scoring program, can be a useful tool for measuring short daytime sleep episodes. At least in healthy young adults, investigators may thus be able to use actigraphy to identify daytime sleep with as much confidence as they currently use actigraphy for nocturnal sleep. The caveat is that some sleep may be misidentified during periods of quiescence that do not actually include sleep. It is important to note, however, that the current study employed one specific brand of actigraph with a proprietary scoring algorithm. These results should not be generalized to other models of actigraphs. Additionally, the participants investigated were healthy, young adults and results may not translate to other populations. Future studies may wish to extend this validation work to daytime sleep episodes with varying time-in-bed periods and *ad libitum* sleep, as well as a variety of other subject populations, including different age ranges and clinical populations.

CONFLICTS OF INTEREST

Sean Drummond has received grant support from Cephalon Inc. and Actelion Inc. Respironics supplied the watches for the current project. Sara Mednick has received research support from Jazz Pharmaceuticals. Jennifer Kanady has nothing to declare.

REFERENCES

Acebo, C. and LeBourgeois, M. K. Actigraphy. *Respir. Care Clin.*, 2006, 12: 23–30.

Ancoli-Israel, S., Clopton, P., Klauber, M. R., Fell, R. and Mason, W. J. Use of wrist activity for monitoring sleep/wake in demented nursing home patients. *Sleep*, 1997, 20: 24–27.

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

Blackwell, T., Redline, S., Ancoli-Israel, S., *et al*. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. *Sleep*, 2008, 31: 283–291.

Bland, J. M. and Altman, D. G. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet*, 1995, 346: 1085–1087.

Blood, M. L., Sack, R. L., Percy, D. C. and Pen, J. C. A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. *Sleep*, 1997, 20: 388–395.

Bradshaw, D. A., Yanagi, M. A., Pak, E. S., Peery, T. S. and Ruff, G. A. Nightly sleep duration in the 2-week period preceding multiple sleep latency testing. *J. Clin. Sleep Med.*, 2007, 3: 613–619.

Carney, C. E., Lajos, L. E. and Waters, W. F. Wrist actigraph versus self-report in normal sleepers: sleep schedule adherence and self-report validity. *Behav. Sleep Med.*, 2004, 2: 134–143.

Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, 1960, 20: 37–46.

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

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

Feinstein, A. R. and Cicchetti, D. V. High agreement but low kappa: I. The problems of two paradoxes. *J. Clin. Epidemiol.*, 1990, 43: 543–549.

Friedman, L., Benson, K., Noda, A., *et al*. An actigraphic comparison of sleep restriction and sleep hygiene treatments for insomnia in older adults. *J. Geriatr. Psychiatry Neurol.*, 2000, 13: 17–27.

Gale, J., Signal, T. L. and Gander, P. H. Statistical artifact in the validation of actigraphy. *Sleep*, 2005, 28: 1017–1018.

Hyde, M., O'Driscoll, D. M., Binette, S., *et al*. Validation of actigraphy for determining sleep and wake in children with sleep disordered breathing. *Sleep*, 2007, 16: 213–216.

Jean-Louis, G., Mendlowicz, M. V., Gillin, J. C., *et al*. Sleep estimation from wrist activity in patients with major depression. *Physiol. Behav*., 2000, 70: 49–53.

Jean-Louis, G., von Gizycki, H., Zizi, F., *et al*. Determination of sleep and wakefulness with the actigraph data analysis software. *Sleep*, 1996, 19: 739–743.

Jean-Louis, G., von Gizycki, H., Zizi, F., Spielman, A., Hauri, P. and Taub, H. The actigraph data analysis software: I. A novel approach to scoring and interpreting sleep–wake activity. *Percept. Mot. Skills*, 1997, 85: 207–216.

Jean-Louis, G., von Gizycki, H., Zizi, F., Spielman, A., Hauri, P. and Taub, H. The actigraph data analysis software: II. A novel approach to scoring and interpreting sleep–wake activity. *Percept. Mot. Skills*, 1997, 85: 219–226.

Jean-Louis, G., Zizi, F., von Gizycki, H. and Hauri, P. Actigraphic assessment of sleep in insomnia: application of the Actigraph Data Analysis Software (ADAS). *Physiol. Behav*., 1999, 65: 559–563.

John, M. W. Daytime sleepiness, snoring, and obstructive sleep apnea. *Chest*, 1993, 103: 30–36.

Johnson, N. L., Kirchner, H. L., Rosen, C. L., *et al*. Sleep estimation using wrist actigraphy in adolescents with and without sleep disordered breathing: a comparison of three data modes. *Sleep*, 2007, 30: 899–905.

Kawada, T. Agreement rates for sleep/wake judgments obtained via accelerometer and sleep diary: A comparison. *Beh. Res. Meth*., 2008, 40: 1026–1029.

Kripke, D. F., Mullaney, D. J., Messin, S. and Wyborney, V. G. Wrist actigraphic measures of sleep and rhythms. *Electroencephalogr. Clin. Neurophysiol*., 1978, 44: 674–676.

Kushida, C. A., Chang, A., Gadkary, C., Guilleminault, C., Carillo, O. and Dement, W. C. Comparison of actigraphic, polysomnographic and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med*., 2001, 2: 389–396.

Lichstein, K. L., Stone, K. C., Donaldson, J., *et al*. Actigraphy validation with insomnia. *Sleep*, 2006, 29: 232–239.

Lichstein, K. L., Wilson, N. M., Noe, S. L., Aguillard, R. N. and Bellur, S. N. Daytime sleepiness in insomnia: behavioral, biological and subjective indices. *Sleep*, 1994, 17: 693–702.

Monk, T., Buysse, D. and Rose, L. R. Wrist actigraphic measures of sleep in space. *Sleep*, 1999, 22: 948–954.

Morgenthaler, T., Alessi, C., Friedman, L., *et al*. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. *Sleep*, 2007, 30: 519–529.

Moul, D. E., Nofzinger, E. A., Pilkonis, P. A., Houck, P. R., Miewald, J. M. and Buysse, D. J. Symptom reports in severe chronic insomnia. *Sleep*, 2002, 25: 553–563.

National Sleep Foundation. *2008 Sleep in America Poll: Summary of Findings. [Internet]*. Washington: National Sleep Foundation. 2009 March. Available at: http://sleepfoundation.org/article/sleep-america-polls/2009-health-and-safety (accessed 15 April 2009).

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

Rechtschaffen A, Kales A. *A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects*. US Government Printing Office, Washington DC, 1968.

Roehrs, T., Zorick, F., Wittig, R., Conway, W. and Roth, T. Predictors of objective level of daytime sleepiness in patients with sleep-related breathing disorders. *Chest*, 1989, 95: 1202–1206.

Sadeh, A., Sharkey, K. M. and Carskadon, M. A. Activity-based sleep–wake identification: an empirical test of methodological issues. *Sleep*, 1994, 17: 201–207.

Tilmanne, J., Urbain, J., Kothare, M. V., Wouwer, A. V. and Kothare, S. V. Algorithms for sleep–wake detection using actigraphy: a comparison study and new results. *J. Sleep Res*., 2009, 18: 85–98.

Youden W. J. Index for rating diagnostic tests. *Cancer*, 1950, 3: 32–35.