Wake Detection Capacity of Actigraphy During Sleep

Jean Paquet, PhD[1]; Anna Kawinska, MSc[1]; Julie Carrier, PhD[1,2]

[1]Centre d'étude du sommeil et des rythmes biologiques, Hôpital du Sacré-Cœur de Montréal, Canada; [2]Department of Psychology, Université de Montréal, Canada

Study Objectives: To evaluate the ability of actigraphy compared to polysomnography (PSG) to detect wakefulness in subjects submitted to 3 sleep conditions with different amounts of wakefulness: a nocturnal sleep episode and 2 daytime recovery sleep episodes, one with placebo and one with caffeine. A second objective was to compare the ability of 4 different scoring algorithms (2 threshold algorithms and 2 regression analysis algorithms) to detect wake in the 3 sleep conditions.
Design: Three nights of simultaneous actigraphy (Actiwatch-L, Mini-Mitter/Respironics) and PSG recordings in a within-subject design.
Setting: Chronobiology laboratory.
Participants: Fifteen healthy subjects aged between 20 and 60 years (7M, 8F).
Interventions: 200 mg of caffeine and daytime recovery sleep.
Results: An epoch-by-epoch comparison between actigraphy and PSG showed a significant decrease in actigraphy accuracy with increased wakefulness in sleep conditions due to the low sleep specificity of actig-

raphy (generally <50%). Actigraphy overestimated total sleep time and sleep efficiency more strongly in conditions involving more wakefulness. Compared to the 2 regression algorithms, the 2 threshold algorithms were less able to detect wake when the sleep episode involved more wakefulness, and they tended to alternate more between wake and sleep in the scoring of long periods of wakefulness resulting in an overestimation of the number of awakenings.
Conclusion: The very low ability of actigraphy to detect wakefulness casts doubt on its validity to measure sleep quality in clinical populations with fragmented sleep or in situations where the sleep-wake cycle is challenged, such as jet lag and shift work.
Keywords: Actigraphy, sleep, polysomnography, validation, wakefulness, sensitivity
Citation: Paquet J; Kawinska A; Carrier J. Wake detection capacity of actigraphy during sleep. *SLEEP* 2007;30(10):1362-1369

INTRODUCTION

POLYSOMNOGRAPHY (PSG) IS THE GOLD STANDARD FOR MEASURING SLEEP. RECENTLY, WRIST ACTIGRAPHY HAS EMERGED AS THE MOST POPULAR ALTERNATIVE to PSG, due to its low invasiveness and cost and the ease of monitoring sleep-wake cycles in ecological environments. The actigraph is a small, wrist-worn device that contains an accelerometer to monitor the number of wrist movements per epoch (e.g., 30 or 60 sec). Scoring algorithms are used to identify sleep or wake states from activity counts and to determine sleep parameters such as sleep onset latency (SOL), total sleep time (TST), number and duration of awakenings, and sleep efficiency (SE = ratio of TST to total time in bed *100). Several types of devices that monitor activity have been used for clinical and research purposes (e.g., Actillume, Actiwatch, Gaëhwiler, MotionLogger). Each of these devices records activity in different ways and has unique algorithms for estimating sleep/wake variables from activity counts. The raw activity counts and sleep/wake measures from these systems may or may not be comparable.

A number of studies have evaluated the ability of actigraphy to discriminate between sleep and wake as defined by PSG criteria,

although it must be noted that the majority of these were done using the Actillume device. In an extensive literature review of the role of actigraphy in sleep research, Ancoli-Israel et al[1] reported that actigraphy and PSG show overall minute-by-minute concordance rates of 91%-93% in adult populations. In a more recent review, Acebo et al[2] reported high epoch-by-epoch agreement rates (>85%) between actigraphy and PSG in healthy subjects of different age groups. However, these high overall concordance rates often mask the very low capacity of actigraphy to detect wake. Several studies using different types of devices have estimated actigraphy wake detection at around 35%-50%, a level equivalent to chance (35% for Blood et al[3]; 34%-44% for de Souza et al[4]; 48% for Kushida et al[5]; 34%-56% for Signal et al[6]; 36% for Sivertsen et al[7]). Since the sleep episodes of healthy subjects usually comprise more than 90% of sleep, total agreement rates would be high despite the large discrepancies in wake detection.[4,7-9] In general, actigraphy tends to overestimate TST and SE,[4,10] while underestimating sleep latency[11] compared with PSG. These biases are not surprising, since actigraphy has difficulty detecting wake when the subject lies immobile in bed in a nonsleeping state.[2]

Actigraphy in general appears to be less accurate in populations showing fragmented sleep compared to healthy subjects.[1,2,7,8,12,13] This points to a very important weakness for the use of actigraphy in clinical populations or in situations where the sleep-wake cycle is challenged, such as jet lag and shift work. The majority of studies on insomniac populations have shown that actigraphy overestimates TST (from 14 to 60 minutes) and SE compared to PSG,[5,7,14-17] with only a few reporting an underestimation of TST with actigraphy in similar populations.[18,19] Although actigraphy measurements occasionally report false negatives (actigraphy scores wake when the subject is sleeping), in periodic leg movement disorder for example,[15] false positives (actigraphy scores sleep when the subject is awake) remain a major concern for ac-

---

Disclosure Statement

This was not an industry supported study. The authors have reported no financial conflicts of interest.

tigraphy.[1,2] Since large intersubject differences exist in individual movement during the sleep episode, it is important to evaluate the direct impact of different amounts of sleep fragmentation on actigraphy-determined sleep parameters in the same subject to control for intersubject variability.

Various algorithms have been proposed to score sleep/wake states using activity counts recorded by actigraph.[10,20-22] Some use activity count thresholds to determine sleep/wake states.[10,20] Others use statistical analyses to determine the actigraphy variables that predict sleep/wake in PSG, then use these variables to build regression equations to predict sleep/wake states.[21-22] Only one study compared the performances of different actigraphy scoring algorithms.[4] A comparison of algorithms developed by Cole[10] and Sadeh[22] showed that both exhibit good sensitivity to detect sleep (>97%) but very low specificity (ability to detect wake), especially Cole's (34%). Using a commercial algorithm with different activity counts thresholds, Kushida et al[5] and Signal et al[6] showed that low thresholds are more specific (better able to detect wakefulness) while high-threshold algorithms are more sensitive (better able to detect sleep).

The aim of this study was to compare the ability of actigraphy (as measured by the Actiwatch device) and PSG to detect wakefulness in subjects submitted to 3 sleep conditions with different amounts of wakefulness: a nocturnal sleep episode and 2 daytime recovery sleep episodes, one with placebo and one with caffeine. Even after a night of sleep deprivation, daytime recovery sleep is more fragmented than nocturnal sleep, because recovery sleep is initiated when the biological clock normally promotes wakefulness.[23] Moreover, our previous work demonstrated that, compared to placebo, caffeine administered before daytime recovery sleep increases total wake time and decreases SE.[24] A second objective of the present study was to compare the wake detection capacity of 4 scoring algorithms (2 based on activity count thresholds and 2 based on regression analysis) in the 3 sleep conditions. We predicted that actigraphy wake detection would decrease with increased amount of wakefulness across the 3 sleep conditions.

METHODS

Subjects

Twenty-three moderate caffeine consumers (equivalent to one to 3 cups of coffee per day) were selected from a study that evaluated the effects of caffeine on daytime recovery sleep.[24] Three subjects were excluded due to missing actigraphic data on at least one PSG sleep recording. All subjects were in good physical and mental health. They were all nonsmokers, and none consumed any drugs or medications that could affect sleep. Participants were excluded for the presence of sleep disturbances such as sleep complaints, sleep apneas and hypopneas (index per hour >10), and periodic leg movements (index per hour >10). Inclusion/exclusion criteria are reported in detail in Carrier et al.[24] The research project was approved by the hospital's Ethics Committee. All subjects signed a consent form that informed them of the nature and risks of the study, and subjects received financial compensation for participating.

Procedures

Subjects participated in both caffeine and placebo conditions in a double-blind cross-over design. They attended 2 sessions (Caffeine and Placebo) at the chronobiology laboratory, separated in time by one month. Each session included one nocturnal sleep and one night of total sleep deprivation followed by a daytime recovery sleep. All subjects wore an actigraph and underwent PSG recording on 3 sleep episodes that were specifically chosen because of their increasing levels of awakenings: a baseline nocturnal sleep episode (NS), a daytime recovery sleep episode after the administration of a placebo (lactose; DRS), and a daytime recovery sleep episode after the administration of 200 mg of caffeine (CDRS). To ensure optimal adaptation to the lab environment, the NS from the second visit was used in the present analyses. Following departure from the lab in the morning, subjects performed their regular activities until the end of the afternoon, at which point they returned to the lab. Subjects then remained awake in bed until the next morning. A research assistant was constantly present to make sure the subjects did not fall asleep. A morning recuperative sleep episode was initiated 1 h after the subjects' habitual wake time (following 25 h of wakefulness). Subjects received 1 capsule containing either caffeine (100 mg) or placebo 3 h prior to their bedtime and the remaining dose of caffeine (100 mg) or placebo 1 h before bedtime. Subjects were asked to stay in bed for their habitual sleep duration. The experimental design is described in detail in Carrier et al.[24]

Measures

Actigraphy

Nondominant wrist activity was recorded using an Actiwatch-L (Mini Mitter, Respironics Inc, Bend, OR). This small, watch-like device contains an accelerometer that senses and records physical motion in all directions and a photodiode to monitor light intensity. Motion is converted to an electric signal and digitally integrated to derive an activity count. Sensitivity is 0.05 g, with a bandwidth between 3 Hz and 11 Hz and a sampling frequency of 32 Hz. The 3-11 Hz bandwidth used in the Actiwatch monitor is higher and larger than in other commercially available actigraphs (0.25-3 Hz) but that bandwidth has shown sensitivity to movement detection.[25] Data were averaged by the monitor into one-minute epochs. After monitoring, the data were downloaded onto a computer to an Actireader via a wireless link set up by Actiware 5.0. All sleep episodes were visually inspected before analysis to screen for artifacts and malfunctioning. Seven Actiwatch-L activity monitors were used in this study. Monitor calibration showed only slight differences in recorded activity counts (mean coefficient of variation was 13% for the 7 monitors).

PSG Sleep Recordings

EEG electrodes were placed according to the international 10-20 system, using a referential montage with linked ears, chin EMG, and left and right EOG. A Grass Model 15 Neurodata system with 15A54 amplifiers (gain of 10,000, 0.3-100 Hz bandpass, -6dB) was used, with signals digitized at a sampling rate of 256 Hz using a commercial software (Harmonie 5.1, Stellates Systems, Montreal, Canada). Sleep stages were scored visually on a computer screen (LUNA, Stellates Systems, Montreal, Canada) using standard criteria,[26] modified for 20-s epochs. The 20-s epochs were rescored to one-minute epochs by way of majority (ex: 2 wake or sleep epochs out of 3 were rescored wake or sleep, respectively) to match the one-minute actigraphic epoch.

Data Analysis

Perfect synchronization between the PSG and actigraph is required to evaluate epoch-by-epoch concordance. Prior to each sleep recording, the PSG computer clock and the computer clock that signals the actigraph were precisely synchronized with the main server. A visual inspection of the PSG tracings and recorded actigraphy activity bursts was then performed to detect any temporal gaps between the 2 measures. Five subjects were excluded from the analysis due to dysynchronization between the 2 measures. Data from the remaining 15 subjects (7 men and 8 women), 20 to 60 years old (mean=39.3; SD=15.1) were analyzed.

Two sets of analyses were performed to determine PSG and actigraphy agreement: an epoch-by-epoch agreement analysis and a sleep parameters concordance analysis. The epoch-by-epoch agreement analysis provided sensitivity, specificity, and accuracy parameters. Sensitivity was defined as the proportion of all epochs scored as sleep by the PSG that were also scored as sleep by actigraphy. Specificity was the proportion of all epochs scored as wake by the PSG that were also scored as wake by actigraphy. Accuracy was the proportion of all epochs correctly identified by actigraphy. The second set of analyses involved comparisons between sleep parameters estimated with PSG and with actigraphy.

Four methods of scoring the actigraphy-derived sleep/wake activity counts were applied. The first 2 were threshold-based method algorithms, provided by Actiwatch-L manufacturers (Mini Mitter,Respironics, Inc. Bend, OR). Actiware uses a weighting algorithm with 3 different thresholds: low (20), medium (40), and large (80). These algorithms,[5] validated on sleep disordered patients, score original activity counts by a weighting scheme that reflects the temporal distance relative to the scored epoch. For example, the 1-minute epoch is rescored as follows:

$$A = 0.04E_{-2} + 0.2E_{-1} + 1E_0 + 0.2E_{+1} + 0.04E_{+2}$$

where A = sum of activity counts for the 1-minute scored epoch and the surrounding epochs; E_n = activity counts of the previous, successive, or scored epoch. If the summed activity count exceeds the defined threshold, the epoch is scored as wake; otherwise it is scored as sleep. The 40 (ACT40) and 20 (ACT20) activity count thresholds were used in the present study because of their superior sensitivity/specificity ratio.[5]

The two other algorithms used here are derived from Lötjönen et al,[21] who applied to the Actiwatch device a method developed by Sadeh et al,[22] referred to here as regression analysis based methods. First, the raw activity count of each epoch is converted to create 4 variables: mean activity in a window of 7 epochs around the scored epoch; standard deviation of the activity in a window of 8 epochs around the scored epoch; number of activity counts above 10 in a window of 11 epochs around the scored epoch; and the natural logarithm of the activity in the scored epoch. A logistic regression is then performed on all subject data with the 4 converted variables and the activity value of the scored epoch as the independent variables and PSG sleep/wake classification as the dependent variable. The result of the logistic function gives a sleep score that is positive for a sleep epoch and negative for a wake epoch. The exact logistic regression equation with the regression coefficient used by Lötjönen et al[21] is as follows:

$$SS = 1.687 + 0.003*[s] - 0.034*[mean] -0.419*[nat] + 0.007*[sd] - 0.127*[ln]$$

where SS = sleep score; s = activity value of the scored epoch; mean = mean activity in a window of 7 epochs around the scored epoch; nat = number of activity counts above 10 in a window of 11 epochs around the scored epoch; sd = standard deviation of the activity in a window of 8 epochs around the scored epoch; and ln = natural logarithm of the scored epoch. Our third method (LötEq) consisted of directly applying the function to our data to determine its applicability to other datasets.

The fourth method (LötMt) was to apply the method described by Lötjönen et al[21] to calculate coefficients derived from our actigraphic data. The same 4 variables were created and a classification function was derived from the logistic regression:

$$SS = 2.457 - 0.004*[s] - 0.689*[nat] - 0.007*[sd] - 0.108*[ln]$$

The variable mean activity in a window of 7 epochs around the scored epoch was not a significant variable in the logistic regression, and was omitted from our equation.

Four sleep parameters were calculated with the same definition for PSG and from the four scoring algorithms derived from the actigraphy data. The sleep parameters derived from the threshold-based method algorithms were calculated with the Actiware program provided with the Actiwatch and the sleep parameters derived from PSG and from regression analysis based methods were calculated using a homemade visual C++ program. Sleep latency is the number of 1-minute epochs from the time of lights off to the first 10 successive sleep epochs (the default criterion for the Actiware program). TST is the total number of 1-minute epochs scored as sleep from lights off to lights on. SE is TST/ total recording time * 100. Number of awakenings is the number of continuous blocks of 1-minute epochs or more from the end of sleep latency to lights on.

Statistical Analysis

Two-way repeated measures ANOVAs with sleep condition (NS, DRS, and CDRS) and algorithm (ACT40, ACT20, LötEq, and LötMt) factors were performed on sensitivity (ability to detect sleep), specificity (ability to detect wake), and accuracy (sleep and wake). Two-way repeated measures ANOVAs with sleep condition (NS, DRS, and CDRS) and scoring method (PSG, ACT40, ACT20, LötEq, and LötMt) factors were performed on sleep parameters. In view of their abnormal distribution, sleep latency and number of awakenings were log-transformed before analysis. Simple effect analyses were performed when significant interactions were found. The post hoc Tukey HSD test was used for multiple comparisons of mean on significant main effect. When repeated measures with more than 2 levels were used, the Huynh-Feldt correction for sphericity was applied, and epsilon values and original degrees of freedom were reported.

To assess PSG and actigraphy agreement, sleep parameters obtained from PSG and the 4 algorithms were compared by Bland-Altman plotting.[27] For each subject, average of (on X axes) and difference between (on Y axes) actigraphic and PSG estimates were plotted on a graph for each sleep parameter in each sleep condition. Mean differences and standard deviations of the differences were calculated. Mean difference (bias) represents the difference between the 2 measures, positive bias indicates an overestimation of actigraphy, and negative bias indicates an underestimation of actigraphy. Standard deviation of the bias (SD) provides an estimate of the variation of mean difference between the two measures. Statistical analyses were conducted using SPSS version 13 (SPSS Inc, Chicago,IL). Significance level was set at 0.05.

Table 1—Mean (± SD) Sleep Sensitivity, Specificity, and Accuracy of Epoch-by-Epoch Comparison with PSG of 4 Actigraphy Scoring Algorithms in 3 Sleep Conditions (N=15)

Statistical parameters	Sleep conditions[1]	Scoring algorithms[2]			
		Act40	Act20	LötEq	LötMt
Sensitivity (%)	NS	95.3 (2.6)	91.4 (3.8)	94.6 (3.5)	94.8 (3.1)
	DRS	96.0 (2.2)	92.3 (3.4)	96.0 (2.6)	95.8 (2.9)
	CDRS	96.0 (2.6)	93.3 (3.5)	95.3 (2.5)	95.0 (3.3)
Specificity (%)	NS	54.3 (21.6)	65.3 (18.2)	47.3 (18.2)	52.6 (20.8)
	DRS	45.1 (20.2)	54.9 (21.7)	47.1 (24.5)	49.8 (23.6)
	CDRS	37.3 (15.7)	47.8 (16.5)	48.5 (20.3)	49.4 (19.2)
Accuracy (%)	NS	90.7 (4.7)	88.2 (4.0)	90.3 (3.6)	90.6 (3.6)
	DRS	84.0 (9.2)	83.3 (6.6)	85.6 (6.3)	86.2 (5.8)
	CDRS	71.7 (15.4)	74.2 (12.4)	76.8 (12.5)	77.5 (11.7)

[1] NS, night sleep; DRS, day recovery sleep; CDRS, caffeine day recovery sleep
[2] Act40, Actiware medium threshold algorithm; Act20, Actiware low threshold algorithm; LötEq, Lötjönen et al's equation algorithm; LötMt, Lötjönen et al's method algorithm

RESULTS

Epoch-by-epoch Agreement

Table 1 shows sensitivity, specificity, and accuracy values (means and SD) derived from epoch-by-epoch comparisons between each actigraphy scoring algorithm and PSG for the 3 conditions. In general, sensitivity was high for all algorithms and conditions (around 95%), specificity was low for all algorithms and conditions (around 50%), and accuracy varied from 90% in NS to 85% in DRS and 75% in CDRS. Two-way repeated measures ANOVA performed on sensitivity showed a main effect of algorithm only ($F_{3,42}$ = 42.0; P <0.001; ε = 0.96), with the Act20 algo-

rithm showing the lowest sensitivity (Tukey HSD test <0.001). The two-way ANOVA on specificity showed an interaction between algorithms and sleep conditions ($F_{6,84}$ = 6.1; P = 0.007; ε = 0.32). Sleep condition comparisons revealed a significant decrease of specificity from NS to CDRS sleep conditions for Act40 and Act20 algorithms only ($F_{6,84}$ >5.5; P <0.009). There was also an interaction effect between algorithms and sleep conditions for the accuracy parameter ($F_{6,84}$ = 5.8; P = 0.01; ε = 0.26). Sleep condition comparisons showed a systematic decrease of accuracy from NS to CDRS sleep conditions for all algorithms ($F_{2,28}$ >14.8; P <0.001), although the effect was more pronounced for the Act40 algorithm (with a decrease of 19% between the NS and CDRS sleep conditions compared to 14% for the other algorithms).

Table 2—Sleep Parameters (Mean ± SD) Scored with PSG and Estimated by 4 Actigraphy Scoring Algorithms in 3 Sleep Conditions (N=15)

Sleep parameters / Sleep conditions[1]	PSG	Scoring algorithms[2]			
		Act40	Act20	LötEq	LötMt
Sleep latency (min)					
NS	21.2 (33.6)	7.3 (7.6)	12.3 (9.4)	8.0 (9.7)	9.6 (10.6)
DRS	5.5 (6.5)	4.1 (8.2)	4.5 (8.4)	3.1 (6.3)	3.4 (6.4)
CDRS	11.7 (17.1)	3.5 (4.0)	4.3 (4.0)	3.4 (4.3)	3.7 (4.5)
Total sleep time (min)					
NS	434.7 (56.1)	438.3 (47.5)	416.3 (49.1)	434.6 (53.2)	432.6 (52.3)
DRS	366.2 (64.8)	416.9 (46.4)	392.8 (52.0)	407.7 (60.1)	403.1 (62.5)
CDRS	281.9 (103.9)	387.7 (45.7)	357.8 (52.1)	358.5 (69.1)	355.4 (71.7)
Sleep efficiency (%)					
NS	90.7 (8.1)	91.4 (4.1)	86.8 (4.8)	90.6 (5.5)	90.1 (5.0)
DRS	78.3 (16.8)	88.2 (8.6)	83.1 (10.0)	86.4 (12.1)	85.4 (12.6)
CDRS	61.8 (25.2)	83.1 (10.3)	76.8 (12.4)	77.2 (16.1)	76.6 (16.7)
No. of awakenings					
NS	12.7 (4.7)	20.0 (8.2)	25.6 (8.9)	11.4 (5.7)	13.6 (5.7)
DRS	13.9 (4.9)	22.9 (10.6)	26.0 (9.7)	10.8 (5.6)	13.5 (6.7)
CDRS	17.0 (10.8)	29.3 (14.2)	30.9 (11.0)	15.2 (7.0)	16.9 (7.7)

[1] NS, night sleep; DRS, day recovery sleep; CDRS, caffeine day recovery sleep
[2] Act40, Actiware medium threshold algorithm; Act20, Actiware low threshold algorithm; LötEq, Lötjönen et al's equation algorithm; LötMt, Lötjönen et al's method algorithm
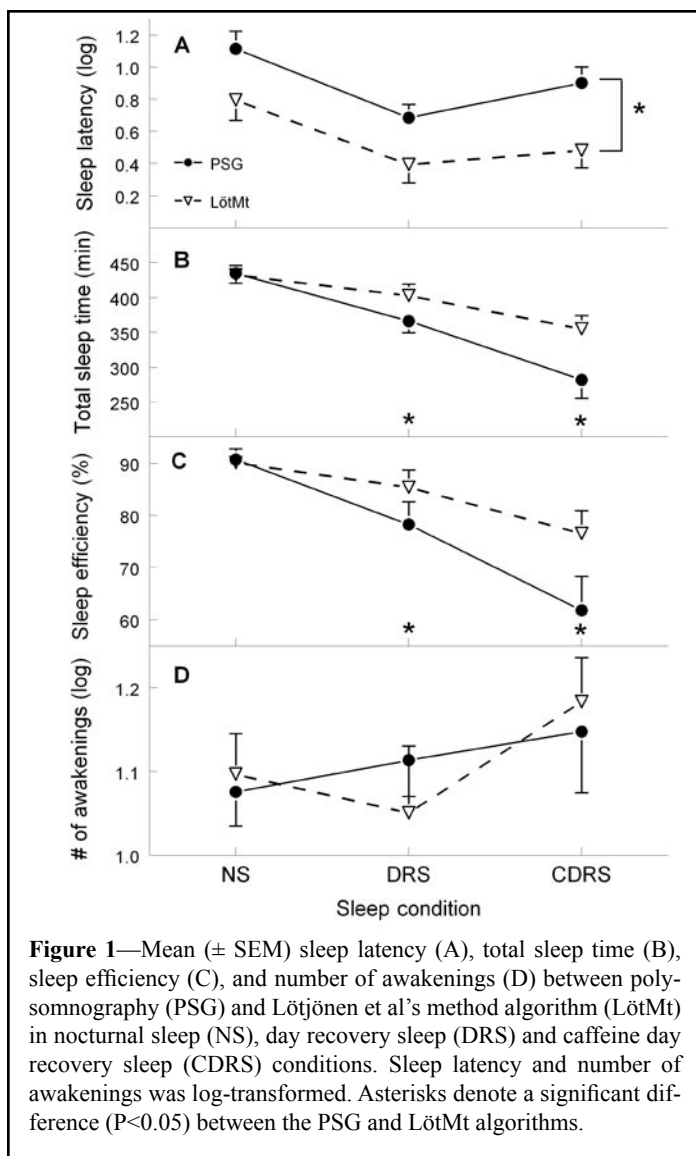
Figure 1—Mean (± SEM) sleep latency (A), total sleep time (B), sleep efficiency (C), and number of awakenings (D) between polysomnography (PSG) and Lötjönen et al's method algorithm (LötMt) in nocturnal sleep (NS), day recovery sleep (DRS) and caffeine day recovery sleep (CDRS) conditions. Sleep latency and number of awakenings was log-transformed. Asterisks denote a significant difference (P<0.05) between the PSG and LötMt algorithms.

Sleep Parameters Concordance

Table 2 presents sleep parameters calculated from PSG and estimated from the 4 actigraphy scoring algorithms for the 3 sleep conditions. As expected, PSG-derived TST and SE decreased gradually from NS to CDRS condition, while number of awakenings increased along these sleep conditions. Sleep deprivation induced a small PSG sleep latency in the DRS condition, but caffeine increased sleep latency in the CDRS condition.

Two-way repeated measures ANOVA performed on sleep latency showed a main effect of algorithm ($F_{4,56}$ = 12.3; P <0.001; ε = 0.62) and a main effect of condition ($F_{2,28}$ = 9.5; P <0.001; ε = 1.0). Post hoc Tukey HSD revealed that sleep latency was higher for PSG compared to the 4 algorithms and that the NS condition had longer sleep latency than the 2 other conditions. Significant interactions between algorithm and sleep condition were found for TST and SE ($F_{8,112}$ >12.6; P <0.001; ε >0.22). Sleep condition comparisons revealed a significant decrease of TST and SE from NS to CDRS sleep condition for all algorithms (P <0.001), but with a steeper decrease for the PSG. Compared to PSG, the 4 algorithms overestimated TST by 39 minutes and 83 minutes on average in the DRS and CDRS conditions, respectively, while SE was overestimated by 7.5% and 16.6%. The ANOVA performed

on number of awakenings showed a main effect of algorithm ($F_{4,56}$ = 36.5; P <0.001; ε = 0.35) and a main effect of condition ($F_{2,28}$ = 4.8; P = 0.01; ε = 1.0). Post hoc Tukey HSD tests revealed that both the Act40 and Act20 algorithms overestimated the number of awakenings compared to PSG, and that the CDRS condition had a higher number of awakenings than the 2 other sleep conditions. Figure 1 illustrates sleep parameters as estimated with PSG and one of the algorithms (LötMt).

Bland and Altman's method was used to compare the sleep parameters estimated via actigraphy scoring algorithms with the PSG-scored sleep parameters. Figure 2 illustrates an example of the method with TST in the caffeine daytime recovery sleep condition. Mean bias was relatively high, with SD of the bias even higher (SD = 83.1), suggesting a large discrepancy between the 2 measures.

Bland and Altman's mean bias and SD between the 4 scoring algorithms and PSG in the 3 sleep conditions are reported in Figure 3. Large mean biases and standard deviations were observed between each algorithm and PSG for sleep latency in the NS and CDRS conditions. For TST and SE, a gradual increase in mean bias and SD was noted as PSG wakefulness increased between NS and CDRS for the 4 algorithms. For the number of awakenings, the Act20 algorithm showed higher mean bias and SD from the NS to CDRS condition and the Act40 algorithm showed only an increase in SD of the bias from the NS to CDRS condition. The LötEq and LötMt algorithms showed very small mean bias, but showed increasing SD of bias from the NS to CDRS condition.

DISCUSSION

In the present study, the ability of the Actiwatch device to detect sleep/wake with 4 scoring algorithms was evaluated in 3 conditions of increasing wake propensity using a within-subject design. Results clearly showed that the accuracy of the Actiwatch to identify wake and sleep as defined by PSG criteria decreases significantly with increasing amount of wake during the sleep episode, which can be accounted for by the low specificity of actigraphy. These results tend to support the recent proposition that actigraphy may not be a sensitive measure in clinical populations with fragmented sleep.[1,2,6-8,12,13] Our results show that the 2 algorithms based on threshold methods were more affected in their ability to detect wake and tended to alternate more between wake and sleep in the scoring of long periods of wakefulness resulting in an overestimation of the number of awakenings compared to the 2 regression algorithms.

Ability to Detect Wakefulness

The low specificity (around 50%) observed in our data in all sleep conditions is similar to that of previous reports,[3-8] highlighting the difficulty of actigraphy to detect wake episodes even in healthy subjects submitted to sleep challenges. When individuals are asleep (as confirmed by PSG), they are immobile most of the time, in which case actigraphy has no difficulty detecting sleep. This explains why the Actiwatch sensitivity is high in our 3 conditions (>90%). The superior ability of actigraphy to detect sleep may explain the high concordance between actigraphy and PSG in healthy subjects, because their nights are composed almost entirely of sleep epochs. When individuals are awake (as confirmed by PSG), they are not necessarily moving. The 50% specificity ob-
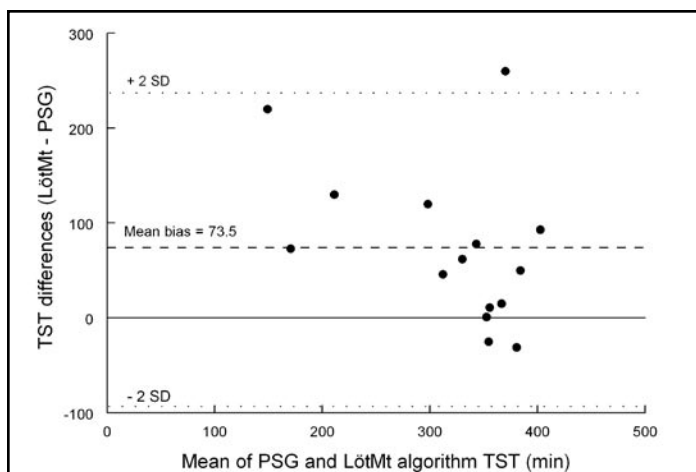
Figure 2—Bland-Altman plot for each subject of the difference between polysomnography (PSG) and Lötjönen et al's method algorithm (LötMt) for total sleep time in the caffeine daytime recovery sleep condition. Mean bias (dashed line) and 2 standard deviations of bias (dotted lines) are presented.
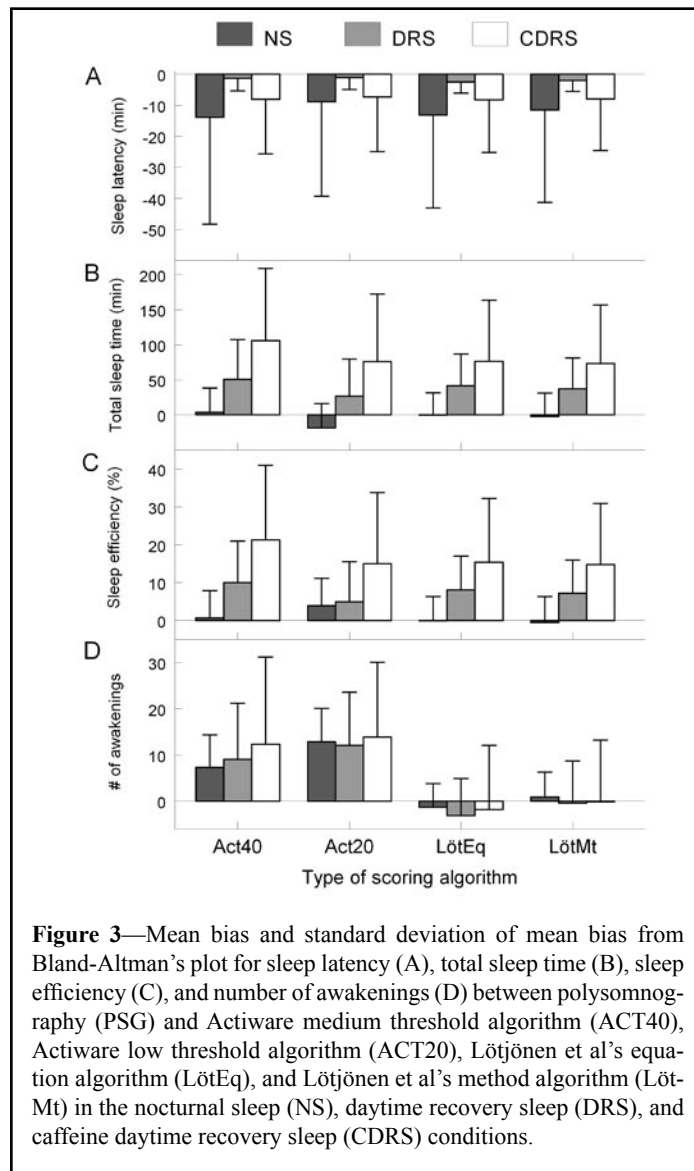

Figure 3—Mean bias and standard deviation of mean bias from Bland-Altman's plot for sleep latency (A), total sleep time (B), sleep efficiency (C), and number of awakenings (D) between polysomnography (PSG) and Actiware medium threshold algorithm (ACT40), Actiware low threshold algorithm (ACT20), Lötjönen et al's equation algorithm (LötEq), and Lötjönen et al's method algorithm (LötMt) in the nocturnal sleep (NS), daytime recovery sleep (DRS), and caffeine daytime recovery sleep (CDRS) conditions.

served in our data therefore suggests that the subjects in our study were immobile about half of the time when awake. For validation purposes, specificity is as important as sensitivity. When subjects have relatively good SE during nights when they are evaluated at home with actigraphy, how can we tell whether they actually sleep well or whether there is a large proportion of quiet wakefulness? This issue merits investigation, since actigraphy has been used to evaluate sleep quality in populations with restricted mobility, e.g., hospitalized burn patients[28] and intensive care patients.[29]

Estimation of Sleep Parameters

In accordance with the results obtained in insomnia studies, the Actiwatch overestimated TST and SE in the DRS and CDRS conditions.[5,7,14-16] We found this overestimation even higher in conditions with more wakefulness, as estimated with PSG. The higher mean bias and SD obtained with Bland and Altman's method in sleep conditions with more wakefulness support this notion. In the CDRS condition, actigraphic algorithms overestimated TST by 1 to 1.5 hours on average. This level of bias could yield an important impact when using actigraphy to assess specific sleep disorders or to monitor treatment effects.

In the NS condition, sleep parameters estimated by actigraphy algorithms and PSG were relatively similar, although specificity was at around chance rate. This indicates that when sleep percentage is high during the night, estimated sleep parameters are less affected by low specificity, since fewer wake epochs have to be detected. This similarity in parameters could also be explained by the fact that actigraphy and PSG score sleep and wake in similar proportions but the epochs scored as sleep and wake are not necessarily the same.[6] This supports the argument that concordance between sleep parameters is not the best way to validate actigraphy.

Consistent with previous studies, the Actiwatch, like other devices, tends to underestimate sleep latency.[4,7,19,30] The difficulty of detecting wake by actigraphy, especially in the transition between wake and sleep, is undoubtedly the main factor to explain this bias.[2] However, in his 5-phase sleep onset spectrum theory, Tryon[11] argued that sleep onset is not a discrete event. Rather, immobility is a necessary stage that precedes falling asleep, which could explain why actigraphy evaluates sleep onset before PSG does. In our study, the large mean bias and SD observed in the Bland and Altman graphs indicate that the process of trying to fall asleep must include a substantial amount of quiet wakefulness. Another difficulty in evaluating sleep latency at home is the relative inability to precisely evaluate the time during which the subject is trying to go to sleep. This problem adds more measurement error in sleep latency and casts doubt on the validity of that actigraphy-estimated sleep parameter.

Estimation of Treatment Effect

Actiwatch-estimated TST and SE showed a significant effect of sleep condition, thus validating actigraphy as a measure of treatment effect. However, the sleep condition effect is reduced considerably compared to PSG—from a difference between NS and CDRS of 153 min in TST (or 29% of SE) for PSG to a difference of 77 min in TST (or 14% of SE) for the best-case actigraphy scenario (LötMt). This reduction by half of the treatment effect when using actigraphy in place of PSG is important. In a comparison between PSG and actigraphy in the treatment of older insomniacs, Sivertsen et al[7] also found a lower treatment effect with actigraphy. They showed that, compared to PSG, actigraphy

detected only 60% of the changes (from before to after treatment) in SE and failed to identify a significant treatment effect for that variable.

Impact of Actigraphic Algorithms

The 2 Actiware algorithms (Act20 and Act40) appear to differ from the 2 regression-based algorithms (LötEq and LötMt) in 2 ways. First, the Actiware threshold algorithms were more affected in their ability to detect wake when the sleep episode involved more wakefulness. Second, while the Actiware algorithms underestimated the minutes of wakefulness, they overestimated the number of awakenings. This could be explained by the fact that the regression method appeared to be a smoother approach for lengthy wakefulness periods, with fewer transitions between sleep and wake epochs. Similar to the observations of Kushida et al,[5] the low threshold (Act20) algorithm was more specific (wake detection) while the medium threshold (Act40) algorithm was more sensitive to sleep, resulting in a 30-min difference for TST in the CDRS condition. The trade-off between sensitivity and specificity between thresholds had a low impact on the significantly decreasing accuracy across sleep conditions of increasing wake propensity. The 2 scoring algorithms based on the approach of Lötjonen et al[21] obtained very close estimates. Our results corroborate that the equation algorithm of Lötjonen et al[21] may be generalized to different individuals evaluated in different sleep conditions and with different Actiwatch monitors. It seems that the Lötjonen et al's equation algorithm is a good choice for estimating sleep parameters, but it is not integrated into a commercial program as are the 2 Actiware algorithms that come with the Actiwatch device.

Limitations and Future Studies

The small sample size, limited to young and middle-aged adults, restricts the generalization of the present study and advocates that the study be reproduced with a larger sample size, with other age groups and with other actigraph devices than the Actiwatch. The selection of a shorter actigraphy duration epoch (e.g., 30 sec) could increase the precision of the results. The DRS and CDRS procedures whereby subjects remained in bed for their habitual sleep duration may have artificially increased the possibility of immobile wakefulness in these conditions. However, there was no difference in specificity between the 3 sleep conditions with the regression algorithms, suggesting that the wake type was similar across all 3 conditions. Future studies using actigraphy as the main sleep evaluation measure should consider the limitations of the device for wake sensitivity, especially in a population with disturbed sleep. Further research is needed to explore the problem of actigraphy specificity. Epoch-by-epoch agreement between actigraphy and PSG night with more wakefulness or in jet-lag and shift-work populations should be investigated. In addition, actigraphy scoring algorithms should be refined to better evaluate the transition between sleep and wake and vice-versa, and methods should be sought to reduce the impact of quiet wakefulness on wake evaluation by actigraphy.

REFERENCES

1. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollack CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003;26:342-92.
2. Acebo C, LeBourgeois MK. Actigraphy. Respir Care Clin 2006;12:23-30.
3. Blood ML, Sack RL, Percy DC, Pen JC. A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. Sleep 1997;20:388-95.
4. de Souza L, Benedito-Silva AA, Nogueira Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. Sleep 2003;26:81-5.
5. Kushida CA, Chang A, Gadkary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. Sleep Med 2001;2:389-96.
6. Signal TL, Gale J, Gander PH. Sleep measurement in flight crew: comparing actigraphic and subjective estimates to polysomnography. Aviat Space Environ Med 2005;76:1058-63.
7. Sivertsen B, Omvik S, Havik OE, et al. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. Sleep 2006;29:1053-8.
8. Gale J, Signal TL, Gander PH. Statistical artifact in the validation of actigraphy. Sleep 2005;28:1017-8.
9. Tryon WW. Issues of validity in actigraphic sleep assessment. Sleep 2004;27:158-65.
10. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992;15:461-69.
11. Tryon WW. Nocturnal activity and sleep assessment. Clin Psychol Rev 1996;16:197-213.
12. Littner M, Kushida CA, McDowell Anderson W, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. Sleep 2003;26:337-41.
13. Sadeh A, Acebo C. The role of actigraphy in sleep medicine. Sleep Med Rev 2002;6:113-24.
14. Edinger JD, Means MK, Stechuchak KM, Olsen MK. A pilot study of inexpensive sleep-assessment devices. Behav Sleep Med 2004;2:41-9.
15. Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. Sleep 1992;15:293-301.
16. Verbeek I, Klip E, Declerck A. The use of actigraphy revised: The value for clinical practice in insomnia. Percept Mot Skills 2001;92:852-6.
17. Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. Sleep 2006;29:232-9.
18. Brooks JO, Friedman L, Bliwise DL, Yesavage JA. Use of the wrist actigraphy to study insomnia in older adults. Sleep 1993;16:151-5.
19. Vallières A, Morin CM. Actigraphy in the assessment of insomnia. Sleep 2003;26:902-6.
20. Jean-Louis G, von Gizycki H, Zizi F, et al. Determination of sleep and wakefulness with the actigraph data analysis software (ADAS). Sleep 1996;19:739-43.
21. Lötjönen J, Korhonen I, Hirvonen K, Eskelinen S, Myllymäki M, Partinen M. Automatic sleep-wake and nap analysis with a new wrist worn online activity monitoring device Vivago Wristcare. Sleep 2003;26:86-90.
22. Sadeh A, Sharkey M, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994;17:201-7.

23. Dijk DJ, Czeisler CA. Paradoxical timing of the circadian rhythm of sleep propensity serves to consolidate sleep and wakefulness in humans. Neurosci Lett 1994;166:63-8.

24. Carrier J, Fernandez-Bolanos M, Robillard R, et al. Effects of caffeine are more marked on daytime recovery sleep than on nocturnal sleep. Neuropsychopharmacology 2007;32:964-72.

25. Van Someren EJW, Lazeron RHC, Vonk BFM, Mirmiran M, Swaab DF. Gravitational artefact in frequency spectra of movement acceleration: implications for actigraphy in young and elderly subjects. J Neurosci Methods 1996;65:55-62.

26. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Los Angeles: UCLA Brain Information Service/Brain Research Institute, 1968.

27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.

28. Raymond I, Ancoli-Israel S, Choinière M. Sleep disturbances, pain and analgesia in adults hospitalized for burn injuries. Sleep Med 2004;5:551-9.

29. Shilo L, Dagan Y, Smorjik Y et al. Effect of melatonin on sleep quality of COPD intensive care patients: a pilot study. Chronobiol Int 2000;17:71-6.

30. Hauri P. Evaluation of a sleep switch device. Sleep 1999;22:1110-7.