Sleep, 17(3):201–207

Fundamental Research

Activity-Based Sleep–Wake Identification: An Empirical Test of Methodological Issues

*Avi Sadeh, Katherine M. Sharkey and Mary A. Carskadon

E. P. Bradley Hospital/Brown University, East Providence, Rhode Island, U.S.A.

Summary: The effects of actigraph placement and device sensitivity on actigraphic automatic sleep–wake scoring were assessed using concomitant polysomnographic and wrist actigraphic data from dominant and nondominant hands of 20 adults and 16 adolescents during 1 laboratory night. Although activity levels differed between dominant and nondominant wrists during periods of sleep ($F = 4.57$; $p < 0.05$) and wake ($F = 15.5$; $p < 0.0005$), resulting sleep–wake scoring algorithms were essentially the same and were equally explanatory ($R^2 = 0.64$; $p < 0.0001$). When the sleep–wake scoring algorithm derived from the nondominant hand was used to score the nondominant data for sleep–wake, overall agreement rates with polysomnography scoring ranged between 91 and 93% for the calibration and validation samples. Results obtained with the same algorithm for the dominant-wrist data were within the same range. Agreement for sleep scoring was consistently higher than for wake scoring. Statistical manipulation of activity levels before applying the scoring algorithm indicated that this algorithm is quite robust toward moderate changes in activity level. Use of "twin-wrist actigraphy" enables identification of artifacts that may result from breathing-related motions. **Key Words:** Actigraphy—Polysomnography—Sleep—Wake—Automatic scoring.

The use of activity-based sleep assessment has gained considerable attention and prominence among sleep researchers and clinicians. Cost-effective actigraphs provide the opportunity to conduct longitudinal, naturalistic studies of the sleep–wake system. The reliability and clinical validity of the method have been evaluated in a number of studies [see Cole et al. (1) for a recent review]. These studies have demonstrated that the minute-by-minute agreement between activity-based sleep–wake scoring and traditional electroencephalogram- (EEG) based scoring ranges between 85 and 95% for most normal and clinical samples.

Despite the growing number of validation and clinical studies, several methodological issues have not been addressed. A single paper by Alster and Sadeh (2) examined a number of the potential artifacts in wrist actigraphy. These artifacts were related to device placement, breathing movements, device-related artifacts, device sensitivity and wrist positioning during sleep. Furthermore, actigraphs traditionally have been attached to the nondominant wrist, but this common practice has never been systematically examined and has not been followed consistently. The sensitivity of actigraphs presumably is calibrated and monitored by the manufacturers, but may vary considerably in practice for a variety of technical reasons.

The present study was aimed at developing a new sleep–wake scoring algorithm for miniature actigraphs and examining three central methodological issues in wrist actigraphy: a) effect of the specific placement of the actigraph (i.e. dominant vs. nondominant wrist), b) tolerance of the automatic sleep–wake scoring algorithms to device sensitivity level and/or subject's overall activity level and c) variability in device sensitivity.

STUDY 1

Methods

Subjects. Thirty-six normal adults, children and adolescents participated in this study. The adult sample included 11 women and nine men, ranging in age between 20 and 25 years (mean age = 22.6; SD = 1.7).

Accepted for publication January 1994.
Address correspondence and reprint requests to Avi Sadeh, D.Sc., Department of Psychology, Tel-Aviv University, Ramat Aviv, Tel Aviv 69978, Israel.

* Current address: Department of Psychology, Tel-Aviv University, Israel.

TABLE 1. *Stepwise discriminant analysis of input measures derived from actigraphic raw activity. Criterion is EEG sleep–wake score*

Wrist	Step	Variable[a]	Partial R^2	F	R^2
Nondominant	1	Mean-W-5-min	0.55	6,588*	0.55
	2	NAT	0.13	826*	0.61
	3	SD-last 6 min	0.05	330*	0.63
	4	LOG-Act	0.04	233*	0.64
Dominant	1	Mean-W-5-min	0.56	6,953*	0.56
	2	SD-last 6 min	0.11	689*	0.61
	3	NAT	0.04	280*	0.63
	4	LOG-Act	0.03	191*	0.64

[a] Mean-W-5-min is the average number of activity counts during the scored epoch and the window of five epochs preceding and following it; SD-last 6 min is the standard deviation of the activity counts during the scored epoch and the five epochs preceding it; NAT is the number of epochs with activity level equal to or higher than 50 but lower than 100 activity counts in a window of 11 minutes that includes the scored epoch and the five epochs preceding and following it; and LOG-Act is the natural logarithm of the number of activity counts during the scored epoch + 1.

* $p < 0.0001$.

TABLE 2. *Agreement rates between activity- and PSG-based sleep–wake scoring. The nondominant wrist algorithm was tested with data from dominant and nondominant wrists[a]*

Sample	n	Sleep	Wake	Total
Nondominant wrist data				
Adults	10	4,185	881	5,066
Calibration		96.43%	78.59%	92.77%
Adults	10	4,322	734	5,056
Validation		97.92%	74.29%	92.58%
Adolescents	16	7,528	1,344	8,872
Validation		94.95%	74.50%	91.16%
Dominant wrist data				
Adults	10	4,157	894	5,051
Calibration		95.78%	79.75%	92.49%
Adults	10	4,254	745	4,999
Validation		96.38%	75.40%	92.54%
Adolescents	16	7,512	1,380	8,892
Validation		94.75%	76.50%	91.37%

[a] Minutes mutually scored by the algorithms and PSG scoring as sleep or wake and total number of consensus sleep and wake minutes (the percentages are percent from PSG scoring).

Eight girls and eight boys aged 10–16 years (mean age = 13.8; SD = 1.9) participated in the children and adolescent sample. All participants were volunteers recruited to participate in a study to assess the effects of nasal occlusion on sleep-related breathing patterns in normal subjects.

Procedure. Subjects spent 2 nights in the sleep laboratory, 1 baseline night and a 2nd night with nasal occlusion. During both nights each subject was monitored with traditional polysomnography (PSG); in addition, a miniature wrist actigraph (AMA-32, Ambulatory Monitoring, Inc., Ardsley, NY) was attached to both wrists. Actigraphs were initialized for zero crossing, mode 18 (internal device code), with a 1-minute epoch interval. Specific actigraphs were not randomized between the two wrist positions, thus most actigraphs were consistently used for either dominant or nondominant wrist.

The present report is based on data collected from the baseline night. Actigraphic data were collected in 1-minute epochs. Traditional PSG-based hand scoring was performed using a 30-second epoch (3). Data were then transformed to 1-minute epoch scoring by preferring the "Wake" score whenever both sleep and wake states were present in the first and second 30-second epoch of each minute. Actigraphic and PSG scoring were matched on a minute-by-minute basis.

Results

The first stage of the statistical analysis was to develop a new sleep–wake scoring algorithm for the miniature actigraph. A new algorithm was considered necessary because the new miniature devices differ from their previous generation in size and weight as well as in sensitivity. Discriminant analysis techniques were used to reach the algorithm for both dominant and nondominant wrists (analyzed separately), based on the method described by Sadeh et al. (4). The algorithms were developed using data collected from the baseline night of the first 10 randomly chosen adult subjects (calibration sample) and then validated on another 10 adult subjects (validation sample) and 16 adolescents. Sixty-two activity variables were calculated for each 1-minute epoch (and its surrounding 10-minute window). The use of such a large number of exploratory measures in this type of analysis is possible because each data point (i.e. each minute of each nocturnal record) represents an observation (a total of 5,066 data points in the calibration sample). These variables included the activity level of each minute (A), the natural logarithm and various combinations of standard deviation, mean value, and minimum value for windows around the scored epoch. Stepwise discriminant analysis (5) identified four activity-derived variables that explained nearly 64% of the PSG sleep–wake scoring (see Table 1). These results indicate that the explanatory power of the dominant and nondominant wrist activity data is very similar and point to similar predictive variables for either wrist.

Discriminant analysis (5) was performed to develop the scoring algorithm based on the four most predictive measures. The scoring algorithm was (see Table 1 for variable definitions):

$$PS = 7.601 - 0.065 \cdot \text{Mean-W-5-min} - 1.08 \cdot \text{NAT} - 0.056 \cdot \text{SD-last 6 min} - 0.703 \cdot \text{LOG-Act}$$

If PS (probability of sleep) is zero or greater, then the specific epoch is scored as sleep; if PS is less than

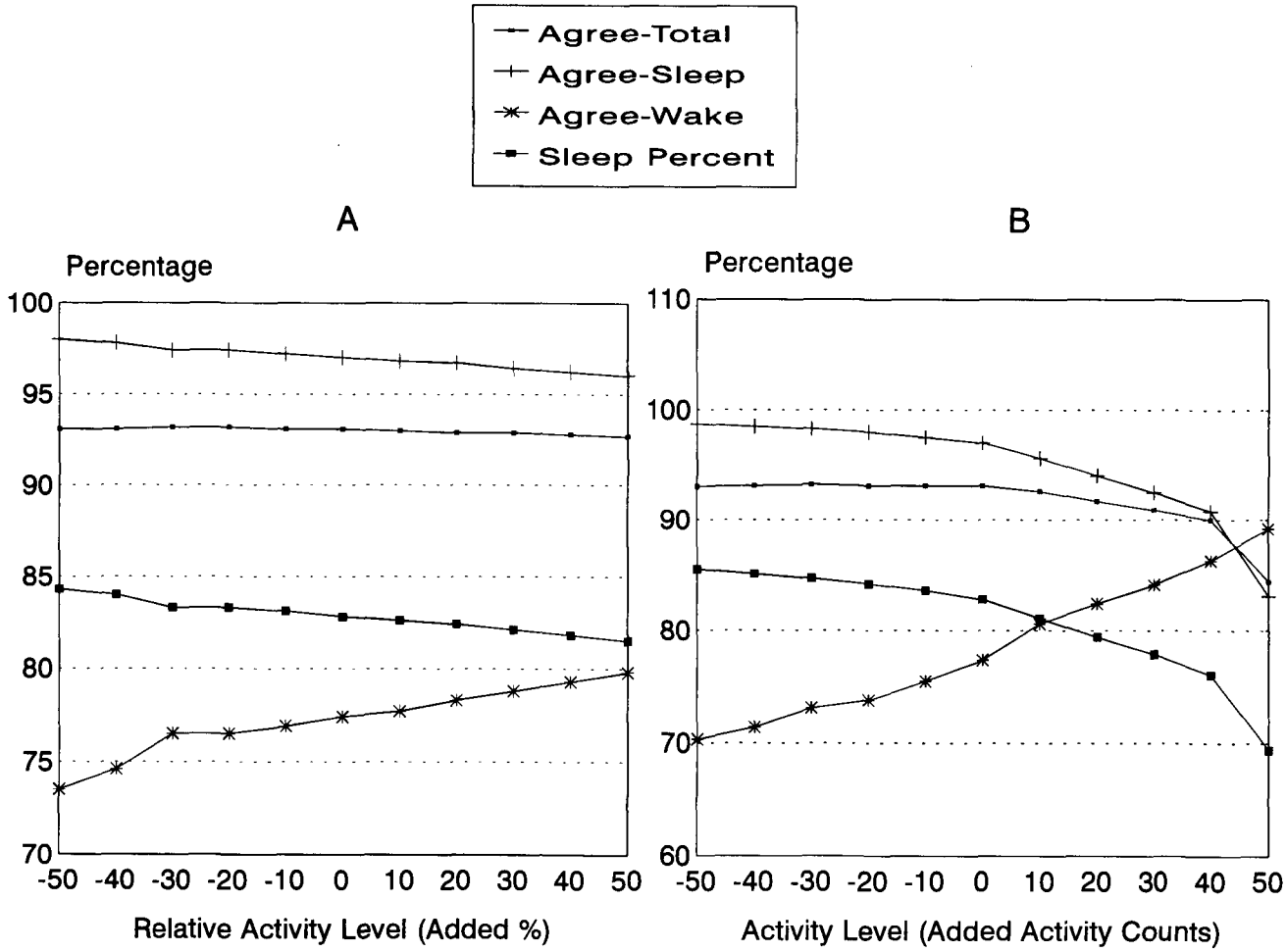

FIG. 1. A: Polysomnographic and actigraphic minute-by-minute scoring agreement rates and actigraphic sleep percent as a function of relative changes in activity level (Agree Total = agreement for both wake and sleep minutes; Agree Sleep = agreement for sleep minutes; Agree Wake = agreement for wake minutes). B: Polysomnographic and actigraphic minute-by-minute scoring agreement rates and actigraphic sleep percent as a function of constant changes in activity level (Agree Total = agreement for both wake and sleep minutes; Agree Sleep = agreement for sleep minutes; Agree Wake = agreement for wake minutes).

zero it is scored as wake. The algorithm was then applied to the raw data of all 36 subjects and compared on a minute-by-minute basis to the EEG-based sleep–wake scoring. These results are summarized in Table 2 and show only minor differences in the agreement rates obtained by applying the algorithm derived from the nondominant wrist to data collected from either dominant or nondominant hand. Similar results were also obtained when an algorithm based on dominant-wrist data was applied to dominant- and nondominant-wrist data.

With regard to wrist placement, minute-by-minute activity levels obtained from the dominant and nondominant wrist during the concomitant PSG–actigraphy monitoring period were significantly different (ANOVA for repeated measures). The mean activity level of the dominant wrist was significantly higher than that of the nondominant wrist during PSG-determined sleep (6.84 vs. 6.16; $F = 4.57$, $p < 0.05$), as well as during wakefulness (25.8 vs. 22.3; $F = 15.5$; $p < 0.0005$).

To evaluate the issue of algorithm robustness to device sensitivity variation, we conducted two tests based on data manipulation and algorithm assessment. In the first analysis, each minute's activity value in every subject was transformed using a constant ratio: $A = A \cdot R$, where A is the original value and R is a ratio ranging from 0.5 to 1.5 (50–150% of the original value, in 10% steps). This type of manipulation does not affect minutes of zero activity. The sleep–wake scoring algorithm was then tested versus PSG again with the new values for each subject. The effects of these changes in activity counts on the accuracy of the automatic actigraphic sleep–wake scoring are illustrated in Fig. 1A. Increasing or decreasing the activity level by 50% resulted in a maximum of 0.4% change in total PSG versus actigraphy agreement rates and in up to 1.5% change in actigraphically derived sleep percent.

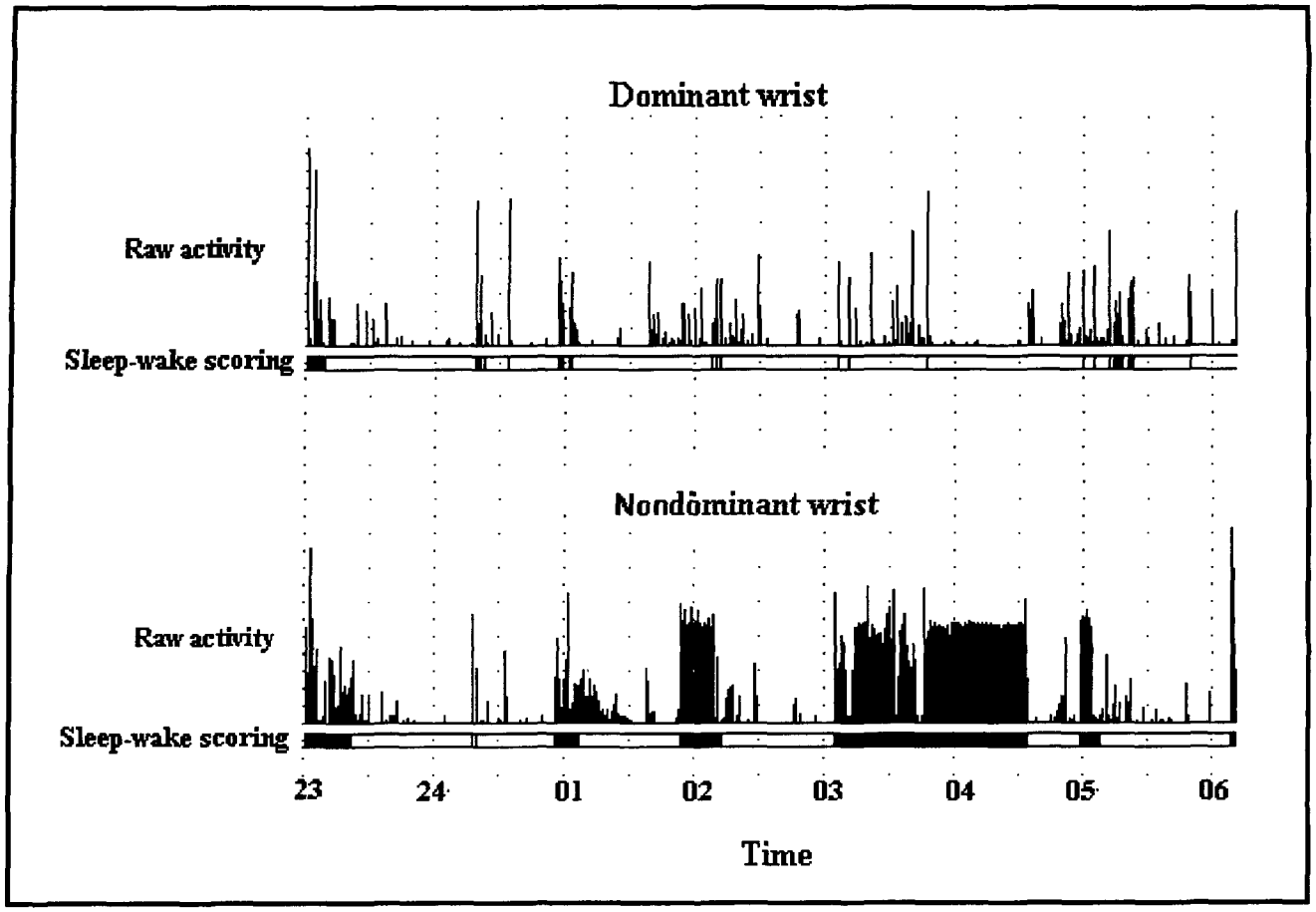

FIG. 2. Raw activity data and sleep–wake scoring from concomitant dominant and nondominant wrist actigraphy.

In the second test, we added (or subtracted) a constant value from each minute's activity count in each subject, A = A + C (C varied from −50 to +50 in 10-count steps; negative values were transformed to zero), and again examined the effects of these changes on the sleep–wake scoring by comparing to PSG. These effects are depicted in Fig. 1B. An increase or reduction of activity level by 20 counts per minute (approximately the average level of activity during waking periods in bed) resulted in as much as 1.4% change in total PSG versus actigraphy agreement rates and 3.6% in actigraphically determined sleep percent.

During the course of our work with twin-wrist actigraphy (simultaneous actigraphic recording from both dominant and nondominant wrists) on this and other projects, a few examples have demonstrated the risks of certain artifacts that may arise with wrist actigraphy. Figure 2 illustrates a striking example, showing the raw activity data registered by two actigraphs, one attached to the dominant wrist and the other to the nondominant wrist. The discrepancy between these two data sets is striking. Familiarity with the pattern of data seen during the period between 3:30 a.m. and 5:00 a.m. leads us to believe that this is a "breathing artifact" that resulted from placing the wrist on the chest or stomach or from lying on the wrist. When the record from the nondominant wrist was subjected to sleep–wake scoring, this artifact resulted in a greater than 28% decrease in sleep percent.

Other artifacts resulting from externally induced movement, such as the case of sleep in a moving vehicle, are illustrated in Fig. 3, which shows actigraphic raw data of one adult subject sleeping in a train, another sleeping in a moving automobile, and an infant sleeping (his first 2 hours of sleep) with an automatic device rocking his bed. Actigraphic sleep data under such circumstances are indistinguishable from wake data.

STUDY 2

Methods

Subjects. Eight normal healthy subjects participated in this study (five women and three men; aged 12–46 years). No screening criteria were used.

Procedure. Each subject was monitored by two actigraphs attached to the same wrist. Subjects wore the units continuously for 42–48 hours. Each actigraph was

used only once, thus 16 devices were tested during this study. After the monitoring period, the entire sample of activity data was scored for sleep–wake scoring using the scoring algorithm developed in study 1. The data from each actigraph in each pair were matched on a minute-by-minute basis and then subjected to interdevice comparisons.

Results

To assess the interdevice variability in sensitivity, Pearson correlations were conducted on the minute-by-minute activity counts for each subject. In addition, minute-by-minute agreement rates were calculated based on the sleep–wake scoring of each actigraph's data for each subject. Each analysis was performed twice: a) for the entire recording period (approximately 48 hours) and b) for the nocturnal sleep periods. The results of these analyses are summarized in Table 3. For the entire recording period (using both daytime and nocturnal activity data), the interdevice correlations of activity counts (number of zero crossings in each minute) ranged between 0.80 and 0.96. The agreement rates between actigraphic sleep–wake scoring obtained from the two devices ranged between 93 and 99%. The differences in mean activity levels recorded by each pair of devices ranged between 1 and 29 counts, which represents a variance of 1–32% from the averaged activity count. Separate analysis of the nocturnal sleep periods resulted in larger proportional differences in mean activity levels, as demonstrated by more than double counts of activity in two sets of devices. Interdevice minute-by-minute activity correlations had a lower range between 0.60 and 0.96. The interdevice sleep–wake scoring agreement rates remained high (ranging between 93 and 99%).

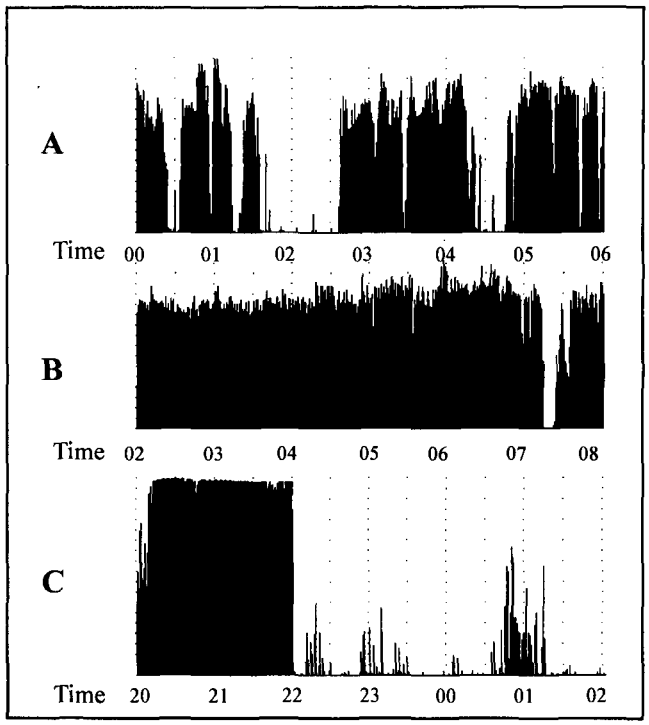

FIG. 3. Raw activity data of three subjects sleeping in motion. A: An adult reportedly sleeping through the night on a night train; B: an adult reportedly sleeping through the night during a nocturnal car ride; C: a 6-month-old infant sleeping his first 2 hours with an automatic device rocking his crib.

DISCUSSION

The results of the present study indicate once more that automatic actigraphic sleep–wake scoring in normal subjects can reach high levels (above 90%) of agreement with traditional PSG for nocturnal sleep periods. Although this was the first validation study conducted with a miniature actigraph (AMA-32), similar results have been reported for other devices (1,4). Cole et al. (1) have pointed out the similarity of the results obtained by a number of groups using different automatic scoring algorithms.

Although significant differences were found between dominant and nondominant wrist activity levels in the present study, the interpretation of these differences is not clear because of the sensitivity issue demonstrated in the second study. Because the actigraphs were not randomly assigned in this study (i.e. most actigraphs were used consistently on either dominant or nondominant wrist), it is unclear whether these differences could be attributed to physiological phenomena or to the variable sensitivity of the actigraphs used in the study. Nevertheless, these wrist-dependent differences in activity level had virtually no detectable effect on the accuracy of the sleep–wake scoring.

The minimal effects of activity-count data manipulation on sleep–wake scoring also suggest that the actigraphic sleep–wake scoring procedure is robust and relatively unresponsive to systematic variations in input activity levels. Although these tests do not represent an exact replica of an actigraphic sensitivity problem (because of the nonlinear, frequency-dependent nature of this measurement method), we used them as an additional component in our assessment of algorithm robustness. The results of these tests indicate that significant changes in activity level, within some limits, produce only modest changes in the accuracy of the automatic actigraphic sleep–wake scoring as manifested in the actigraph–EEG agreement rates. It should be emphasized, however, that these results were obtained in a controlled laboratory situation during a nocturnal period. Naturalistic actigraphic studies may face additional reliability problems, resulting from uncontrolled actigraph removal or other artifacts affecting body movements.

TABLE 3. *Comparison of two actigraphs worn on the same wrist: activity level, Pearson correlations, sleep percents, agreement rates and correlations between sleep percents*

Subject	Period[a]	No. of minutes scored	Mean activity, device 1	Mean activity, device 2	Activity correlation	Sleep percent, device 1	Sleep percent, device 2	Sleep–wake scoring agreement (%)
1	48 hours	2,880	107 ± 104	115 ± 108	0.96	35.9	32.8	98
	SP	809	11.7 ± 28	12.7 ± 28	0.84	88.0	87.6	97
2	48 hours	2,880	103 ± 106	104 ± 107	0.80	39.2	37.9	95
	SP	1,044	9.5 ± 25	9.6 ± 26	0.71	93.6	91.3	95
3	48 hours	2,880	126 ± 110	134 ± 116	0.95	37.6	37.1	99
	SP	1,129	8.2 ± 23	9.4 ± 26	0.83	94.5	93.1	98
4	48 hours	2,524	91 ± 91	101 ± 97	0.86	38.7	37.0	97
	SP	1,005	10.1 ± 26	11.8 ± 28	0.60	92.6	91.0	96
5	48 hours	2,880	131 ± 104	111 ± 96	0.83	30.9	32.1	98
	SP	901	11.3 ± 36	9.9 ± 33	0.96	93.3	94.1	99
6	48 hours	2,566	119 ± 109	115 ± 107	0.93	38.9	39.1	99
	SP	911	5.3 ± 16	4.6 ± 14	0.88	98.8	98.9	99
7	48 hours	2,656	77 ± 96	106 ± 108	0.82	48.1	42.4	93
	SP	1,179	6.8 ± 16	13.5 ± 27	0.60	98.6	92.0	93
8	48 hours	2,880	119 ± 104	114 ± 102	0.95	26.6	27.2	98
	SP	655	8.0 ± 24	17.1 ± 49	0.66	95.6	95.9	98

[a] Analysis conducted for two separate periods: entire recording period (48 hours) and nocturnal sleep periods only (SP).

In the second study, significant interdevice variations in sensitivity were demonstrated by large differences in averaged activity counts measured on the same wrist over a prolonged period of time (42–48 hours) and during nocturnal sleep periods. However, these differences resulted in relatively minor changes in the results of the sleep–wake scoring. The results of the second study support our initial impression that variability in device sensitivity is indeed an important methodological issue in wrist actigraphy and should be properly addressed.

Two methodological implications or guidelines may be drawn from the present findings: a) using sleep–wake interpretation of actigraphic data over a designated sleep period is a robust procedure that is unaffected by placement (dominant vs. nondominant wrist) or device sensitivity (within reasonable limits), and b) researchers using activity data proper (i.e. unprocessed activity counts rather than derived sleep measures) should be extremely cautious regarding device placement and device sensitivity. With regard to actigraph sensitivity, we also suggest several controlled approaches that may be profitably applied in designing actigraphic studies: within-subjects designs (repeated measures) should assign the same actigraph unit for the same individual to minimize error variance resulting from variability in device sensitivity. Furthermore, between-group designs should include a controlled, balanced assignment of actigraphs across groups.

Recent studies have begun to address other methodological issues involved in wrist actigraphy. Hilten et al. (6), for example, studied the issues of intrasubject and internight variability and concluded that although wrist actigraphy seems to involve no "first-night effect", considerable internight and intrasubject variability does occur. Ongoing studies in our laboratories are assessing the effects of "co-sleeping" on wrist actigraphy, and open questions remain as to the effects of sleeping surface (e.g. hard surface vs. waterbed).

We raise a final caution attendant to the development of computerized automatic scoring procedures for actigraphic data. Application of such systems by naive investigators may lead to a mechanistic approach to actigraphic data and preclude development of familiarity with the unique patterns of activity that may occur. A similar risk also arises with EEG automatic sleep scoring procedures. The use of twin-wrist actigraphy is occasionally valuable in highlighting some artifacts and may enable their elimination (see Fig. 3). Nevertheless, a close scrutiny of raw data based on familiarity with unique activity patterns is still necessary for proper use of actigraphy in sleep medicine and sleep research.

Acknowledgements: We thank Clayton Bennett, Jr., James Wyatt, Max Stone, Cecilia Vieira, David Jacobs, Elizabeth Huntsman, William Gruen and Thomas Kazlausky for their assistance. We also thank Christine Acebo and Ron Seifer for their suggestions and comments on the manuscript. This research was supported by NIH grants HL44138 and MH46757.

REFERENCES

1. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist actigraphy. *Sleep* 1992; 15:461–9.

2. Alster J, Sadeh A. Artifact and pattern recognition in wrist actigraphy. *J Polysom Technol* 1990;Spring/Summer:27–30.
3. Rechtschaffen A, Kales A. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Bethesda, Maryland: U.S. Department of Health, Education and Welfare, 1968.
4. Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically based automatic bedtime sleep–wake scoring: validity and clinical applications. *J Ambul Monitoring* 1989;2:209–16.
5. SAS Institute Inc. *SAS/STAT user's guide,* version 6, 4th ed. Cary, NC: SAS Institute Inc., 1989.
6. Hilten JJ van, Braat EAM, Velde EA van der, Middelkoop HAM, Kerkhof GA, Kamphuisen HAC. Ambulatory activity monitoring during sleep: an evaluation of internight and intrasubject variability in healthy persons aged 50–98. *Sleep* 1993;16:146–50.