Sleep, 15(5):461–469

Technical Note

Automatic Sleep/Wake Identification From Wrist Activity

*Roger J. Cole, *Daniel F. Kripke, †William Gruen, *Daniel J. Mullaney and *J. Christian Gillin

**Department of Psychiatry, University of California, San Diego; and San Diego Veterans Affairs Medical Center, San Diego, California 92093, U.S.A.; and †Ambulatory Monitoring, Inc., Ardsley, New York 10502, U.S.A.*

Summary: The purpose of this study was to develop and validate automatic scoring methods to distinguish sleep from wakefulness based on wrist activity. Forty-one subjects (18 normals and 23 with sleep or psychiatric disorders) wore a wrist actigraph during overnight polysomnography. In a randomly selected subsample of 20 subjects, candidate sleep/wake prediction algorithms were iteratively optimized against standard sleep/wake scores. The optimal algorithms obtained for various data collection epoch lengths were then prospectively tested on the remaining 21 subjects. The final algorithms correctly distinguished sleep from wakefulness approximately 88% of the time. Actigraphic sleep percentage and sleep latency estimates correlated 0.82 and 0.90, respectively, with corresponding parameters scored from the polysomnogram ($p < 0.0001$). Automatic scoring of wrist activity provides valuable information about sleep and wakefulness that could be useful in both clinical and research applications. **Key Words:** Activity—Sleep—Wrist actigraph—Automatic scoring.

Many research and clinical situations call for an inexpensive, unobtrusive method of obtaining human sleep/wake data outside of the laboratory. Mullaney et al. (1) showed that data obtained from a wrist-mounted movement detector could be manually scored to distinguish sleep from wakefulness with a high degree of accuracy, as compared to polysomnographic (PSG) scoring. However, the painstaking labor involved in manual scoring reduces the practicality of this method for everyday use. Webster et al. (2) developed an automatic method for scoring wrist activity data for sleep and wakefulness. The scoring algorithm they developed was optimized only for their experimental wrist actigraph, leaving it uncertain whether their method could be generalized to actigraphic instruments now commercially available.

The goal of this study was to replicate and expand upon Webster's work by developing a set of automatic scoring algorithms that could distinguish sleep from wakefulness in a wide variety of normal and sleep-disordered subjects, using wrist activity data obtained by a commercially available wrist actigraph.[1] We measured wrist activity during overnight polysomnography in 41 subjects and analyzed the results on a minute-by-minute basis. Activity scoring algorithms were developed for several data collection epoch lengths. These algorithms correctly distinguished sleep from wakefulness in approximately 88% of minutes scored.

METHODS

Subjects

Our sample of 41 subjects included 32 men and nine women. Mean age was 50.2 ± 14.7 years. Table 1 shows subject ages and polysomnographic sleep efficiencies in each diagnostic category.

The sample was selected to survey both normal sleep and a diversity of sleep disorders. It included 15 normal controls, three elderly normals, 12 psychiatric inpatients, four sleep apnea patients, three patients with disorders of maintaining sleep (DIMS), three bereaved widows, and one back pain patient. Primary diagnoses of the psychiatric patients included DSM IIIR (3) major depressive disorder (n = 7), bipolar disorder (n = 2), organic mood disorder (n = 1), personality disorder (n = 1) and schizophrenia (n = 1). Most of these patients also had comorbid diagnoses, including alcohol abuse (n = 6), substance abuse (n = 2) and personality disorder (n = 3). The sleep apnea patients showed an average of 33.7 ± 23.1 arterial oxygen desaturation events per hour, where a desaturation event was defined as a reduction in arterial oxygen saturation of at

[1] Motionlogger Actigraph, Ambulatory Monitoring, Inc., 731 Saw Mill River Road, Ardsley, New York 10502, U.S.A.

Accepted for publication June 1992.
Address correspondence and reprint requests to Roger J. Cole, Ph.D., 9500 Gilman Dr., La Jolla, California 92093-0667, U.S.A.

TABLE 1. *Sleep/wake identification accuracy[a] in various diagnostic categories*

	Training sample					Validation sample				
Diagnosis	n	Age mean ± SD	PSG sleep efficiency	Minutes compared	Agreement[b]	n	Age mean ± SD	PSG sleep efficiency	Minutes compared	Agreement
Normal	8	47.4 ± 16.7	89.6 ± 4.3%	3,802	91.87%	7	41.9 ± 12.2	86.9 ± 10.5%	3,426	87.97%
Elderly normal	1	68.0 ± —	83.6 ± —%	366	88.25%	2	73.0 ± 8.5	76.6 ± 9.5%	891	85.30%
Psychiatric	6	50.2 ± 7.6	86.8 ± 6.3%	2,804	89.80%	6	36.2 ± 7.4	85.9 ± 14.1%	2,804	89.23%
Sleep apnea	2	60.5 ± 6.4	59.1 ± 18.0%	931	84.00%	2	55.5 ± 17.7	81.6 ± 3.1%	838	83.89%
DIMS	1	75.0 ± —	81.5 ± —%	419	83.05%	2	68.5 ± 7.2	74.0 ± 17.9%	1,001	89.71%
Bereaved	1	59.0 ± —	48.0 ± —%	431	59.16%	2	50.0 ± 5.7	92.9 ± 7.2%	842	91.81%
Back pain	1	53.0 ± —	23.7 ± —%	215	77.67%	0	— ± —	— ± —	—	—
All diagnoses	20	52.8 ± 13.4	80.4 ± 18.7%	8,968	87.93%	21	47.8 ± 15.7	85.0 ± 11.6%	9,802	88.25%

[a] Sleep/wake identification was performed using the algorithm for the maximum 10-second overlapping epoch per minute. Rescoring was performed using the five Webster et al. rules described in the text.

[b] Values represent the percent of minutes in which the sleep/wake score derived from wrist activity agreed with the polysomnographic sleep/wake score.

least 4% lasting for at least 10 seconds. The DIMS patients were self-referred for "difficulty maintaining sleep", and showed an average polysomnographic sleep efficiency of 76.5 ± 13.4% on the night of study.

All subjects entered the laboratory for either clinical evaluation or participation in unrelated research protocols involving overnight polysomnographic recording. All gave informed consent to participate in actigraph research. Two of the 41 subjects (both controls) were recorded on two separate nights; all others were recorded for one night only, giving a total of 43 overnight recordings.

Procedure

Subjects arrived at the laboratory between 2100 hours and midnight. The actigraph was placed on the subject's nondominant wrist in approximately 80% of cases. Placement on the dominant wrist was used when subjects requested it or when nondominant placement would interfere with other monitoring equipment. Actigraphs were set to run in zero-crossing, nonevent mode with a data storage epoch length of 2 seconds. A standard polysomnographic montage [C_{3-4} to linked ears, two channels electrooculogram (EOG), mentalis electromyogram (EMG)] was applied, and the subject lay down in bed. In some cases, subjects remained awake in bed reading or engaged in other activity before lights out; in others, lights out and polysomnographic recording began immediately upon lying down. At lights out, PSG recording began and subjects were instructed to sleep if possible. The actigraph and polysomnographic electrodes were removed when subjects spontaneously arose in the morning (generally between 0500 and 0730 hours).

A time code was automatically written each minute on the polysomnogram. Each minute was scored as either sleep [stage 1, 2, 3, 4 or rapid eye movement (REM)] or awake (stage W plus movement time) according to the standard criteria of Rechtschaffen and Kales (4). Wrist activity data were collapsed from 2-second epochs into 1-minute epochs by methods described below. The actigraph's internal timer was synchronized to the time-code generator clock to allow each minute of the activity record to be matched with the corresponding minute of the polysomnogram.

During the period after donning the actigraph and before the start of PSG recording (e.g. during electrode placement or while reading in bed), subjects were observed either directly or by videocamera. Times during which it was clear that the subject was continuously awake were scored as "wakeful PSG" for purposes of comparison to wrist activity. Times in which sleep/wake status was uncertain before PSG recording began were not used in the data analyses.

The mean duration of overnight activity recordings was 443.7 ± 61.5 minutes. A total of 19,081 minutes of simultaneous PSG sleep/wake and wrist activity data were collected. The activity scoring algorithms (described below) did not allow the first 4 or the last 2 minutes of each record to be assigned a sleep/wake score. This, along with occasional electroencephalogram (EEG) artifact, reduced the total number of minutes of matched actigraph/PSG data to 18,770.

Subjects were stratified by diagnosis and randomly assigned to either a training sample (n = 20) or a validation sample (n = 21). Data from the training sample were used to derive sleep scoring algorithms. Data from the validation sample were used to prospectively assess the accuracy of the algorithms derived from the training sample.

RESULTS

Reducing activity scores

The following methods were compared for reducing the data in the training sample:

TABLE 2. *Sleep/wake identification accuracy[a] by regression analysis for various epoch lengths (training sample)*

			Nonoverlapping epochs			
Epoch length (seconds)	2	6	10	20	30	60
Agreement[b]	85.14%	86.02%	86.67%	86.01%	85.29%	83.86%
			Overlapping epochs			
Epoch length (seconds)	6	8	10	12	14	
Agreement	86.72%	85.96%	86.75%	86.74%	86.74%	

[a] Sleep/wake identification was performed using preliminary algorithms derived from multiple regression on training sample data, as described in the text.

[b] Values represent the percent of minutes in which the sleep/wake score derived from wrist activity agreed with the polysomnographic sleep/wake score.

1) Taking the mean of all 2-second activity scores per minute,
2) Taking the maximum 2, 6, 10, 20 or 30 seconds of activity per minute, considering nonoverlapping epochs (e.g. for an epoch length of 6 seconds, epoch one spans seconds 1–6, epoch two spans seconds 7–12, and so on), and,
3) Taking the maximum 6, 8, 10, 12 or 14 seconds of activity per minute, considering overlapping epochs (e.g. for an epoch length of 6 seconds, epoch one spans seconds 1–6, epoch two spans seconds 3–8, and so on).

Scores derived by each method were entered into a linear regression model, with the PSG sleep/wake score (coded 0 for sleep, 2 for wake) as the dependent variable. The independent variables (seven in all) were activity scores for the same minute, each of the preceding 4 minutes, and each of the following 2 minutes. These independent variables were chosen to match the form of our algorithm to that of Webster et al. (2). The regression analysis yielded an estimated value of the sleep/wake variable for each minute, based on a weighted sum of the seven activity scores. Each estimated value was recoded 0 (sleep) if it was <1, or 2 (wake) if it was ≥ 1. This yielded minute-by-minute sleep/wake predictions from wrist activity. Predicted sleep/wake was then compared to actual sleep/wake for each minute, and the percentage of correct and incorrect predictions was determined.

The lowest percentage of correct predictions was obtained when data were reduced to the mean activity value per minute (83.86%). Using the maximum 30-second nonoverlapping epoch per minute improved the accuracy rate to 85.29%. The highest percentage of correct sleep/wake predictions obtained from training sample data was 86.75%, using the maximum 10-second overlapping epoch per minute. The maximum 12-second and 14-second overlapping epochs and the maximum 10-second nonoverlapping epoch achieved virtually the same accuracies (86.74%, 86.74% and 86.67%, respectively). Accuracy rates from the regression analyses are shown in Table 2.

Optimizing the algorithm

Because regression analysis is based on certain assumptions (e.g. normality of the data) that might not hold true for our data set, a computer program was developed for iteratively optimizing any combination of parameters in a model of the form:

$$D = P(W_{-4}A_{-4} + W_{-3}A_{-3} + W_{-2}A_{-2} + W_{-1}A_{-1} + W_0A_0 + W_{+1}A_{+1} + W_{+2}A_{+2})$$

where $D < 1$ = sleep, $D \geq 1$ = wake, P = a scale factor for the entire equation, W_0, W_{-1}, W_{+1}, etc. = weighting factors for the present minute, the previous minute, the following minute, etc., and A_0, A_{i-1}, A_{i+1}, etc. = activity scores for the present minute, the previous minute, the following minute, etc.

Using the maximum 10-second overlapping epoch to represent each minute's score, we applied this equation to the training sample data, iteratively varying factors P and W_i to maximize the percent of correct sleep/wake identifications. We were able to increase our accuracy rate from 86.75% to 87.05% with the following equation:

$$D = 0.00001(404A_{-4} + 598A_{-3} + 326A_{-2} + 441A_{-1} + 1{,}408A_0 + 508A_{+1} + 350A_{+2})$$

Rescoring

The optimal algorithm misscored actual wake as sleep about 3.5 times as often as it misscored sleep as wake (Table 3A). This is in part because subjects falling asleep stop moving a few minutes before the PSG shows the onset of sleep stage 1. To mitigate this problem, five "rescoring rules" developed by Webster et al. (2) to maximize the accuracy of sleep/wake identification were tested. These rules were: (a) After at least 4 minutes scored as wake, the next 1 minute scored as sleep is rescored wake; (b) after at least 10 minutes scored as wake, the next 3 minutes scored as sleep are rescored wake; (c) after at least 15 minutes scored as wake, the

TABLE 3. *Sleep/wake identification from wrist activity by optimized algorithm*[a]

Polysomnograph	Actigraph: Scored as sleep Minutes	Scored as sleep Row%	Scored as wake Minutes	Scored as wake Row%	Row total (minutes)
A. Training sample, not rescored: overall agreement 87.05%					
PSG sleep	6,213	96.04%	256	3.96%	6,469
PSG wake	905	36.21%	1,594	63.79%	2,499
Total	7,118		1,850		8,968
B. Training sample, rescored[b]: overall agreement 87.93%					
PSG sleep	6,153	95.12%	316	4.88%	6,469
PSG wake	766	30.65%	1,733	69.35%	2,499
Total	6,919		2,049		8,968
C. Validation sample, not rescored: overall agreement 87.91%					
PSG sleep	7,283	96.09%	296	3.91%	7,579
PSG wake	889	39.99%	1,334	60.01%	2,223
Total	8,172		1,630		9,802
D. Validation sample, rescored[b]: overall agreement 88.25%					
PSG sleep	7,216	95.21%	363	4.79%	7,579
PSG wake	789	35.49%	1,434	64.51%	2,223
Total	8,005		1,797		9,802

[a] Sleep/wake identification was performed using the optimized algorithm for the maximum 10-second overlapping epoch per minute.
[b] Rescoring was performed using the five Webster et al. rules described in the text.

next 4 minutes scored as sleep are rescored wake; (d) 6 minutes or less scored as sleep surrounded by at least 10 minutes (before and after) scored as wake are rescored wake, and (e) 10 minutes or less scored as sleep surrounded by at least 20 minutes (before and after) scored as wake are rescored wake. In the training sample, we tested each of these rules separately and in all possible combinations.

Each rule, applied by itself, raised the rate of correct sleep/wake identification to some degree. The combined application of rules (a) through (e) produced the greatest improvement, increasing the accuracy rate from 87.05% to 87.93%, and reducing the ratio of "false sleeps" to "false wakes" from over 3.5:1 to less than 2.5:1 (Table 3B). Identical results were obtained when only rules (a) through (d) were applied. The addition of rule (e) had no effect because the few (n = 3) epochs that met the rule (e) criteria also met the rule (d) criteria. All epochs to which rule (e) could have been applied would have been rescored correctly. An example of an overnight recording with and without rescoring is shown in Fig. 1.

Prospective assessment

The results reported above validate the use of this algorithm to retrospectively identify sleep/wake in the sample from which it was derived (i.e. the training sample). The true test of an algorithm is to determine how well it prospectively identifies sleep/wake in an entirely new set of data. We performed this test by applying our optimal algorithm to the validation sam-

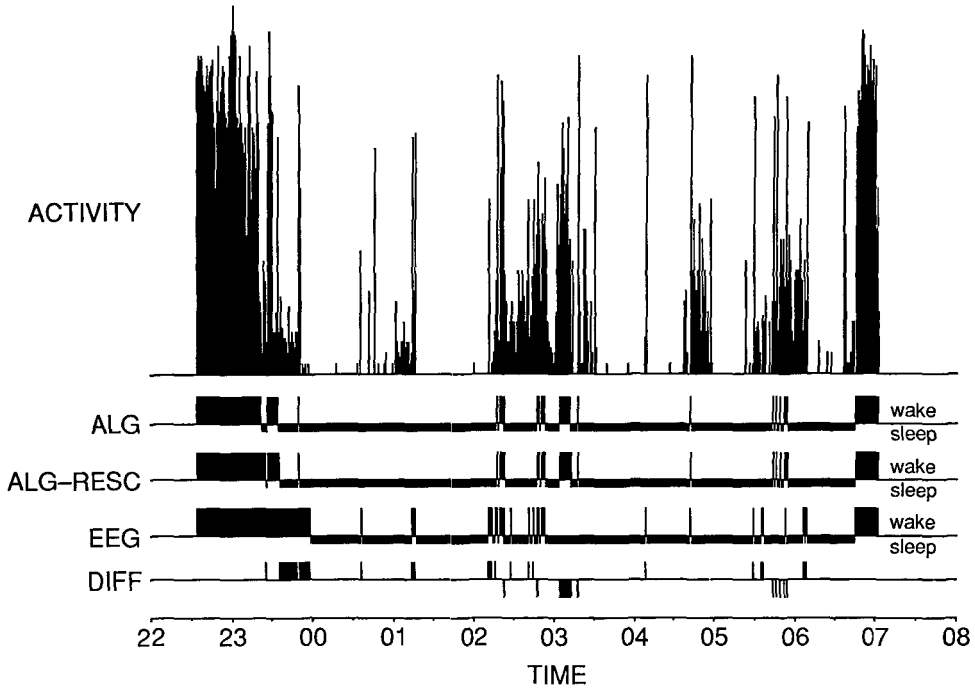

FIG. 1. Automatic sleep/wake scoring from wrist activity. Normal control subject, male, age 52. Actigraph algorithm agreed with polysomnogram 88.61% of the time (after rescoring). ACTIVITY: Wrist activity data compressed to one-minute epochs. ALG: Minute-by-minute sleep/wake scores produced by algorithm for 10-second overlapping epochs. ALG–RESC: Algorithm sleep/wake values after rescoring according to Webster rescoring rules. EEG: Minute-by-minute sleep/wake scores from polysomnogram. DIFF: Discrepancy between sleep scores from rescored algorithm and polysomnogram (marks above line show polysomnogram wake labeled sleep by algorithm, marks below line show polysomnogram sleep labeled wake).

TABLE 4. *Comparison of four sleep onset criteria in the training sample*[a]

Criterion[b]	PSG latency (minutes)[c]	Actigraph latency (minutes)[c]	r^d
1 minute	52.1 ± 46.8	23.1 ± 24.1	0.53
10 minutes	61.6 ± 55.4	37.5 ± 33.1	0.58
15 minutes	73.5 ± 82.7	57.1 ± 86.5	0.87
20 minutes	75.6 ± 89.7	70.9 ± 102.0	0.94

[a] Actigraph sleep/wake identification was performed using the algorithm for the maximum 10-second overlapping epoch per minute. Rescoring was performed using the five Webster et al. rules described in the text.

[b] All criteria take the form, "sleep onset is the beginning of the first interval containing at least n minutes scored as sleep stage 1 or greater with no more than 1 minute of wakefulness intervening". The same criteria were applied to both PSG and actigraph records.

[c] Mean ± S.D. Latencies are computed from the start of simultaneous actigraph/PSG recording, regardless of whether the subject was attempting to sleep.

[d] Pearson correlation between PSG and actigraph latency scores.

ple of 21 subjects. Before rescoring, the overall accuracy rate was 87.91%. After applying the five rescoring rules, the accuracy rate was 88.25% (Tables 3C and 3D). Thus, the algorithm actually performed slightly better at identifying sleep/wake in the new group of subjects than it did in the sample from which it was derived.

Assessment of whole-night sleep parameters

Another way of assessing the accuracy of sleep/wake scoring by wrist activity is to compare whole-night sleep parameter estimates by polysomnography vs. actigraphy. We compared estimates of total sleep time, percent sleep, sleep efficiency, sleep onset latency and wake time after sleep onset. In all analyses, activity data were scored with the optimal 10-second algorithm plus the five rescoring rules.

The criterion for sleep onset was chosen by testing four candidate criteria on the training sample. Each candidate took the form "sleep onset is the beginning of the first interval containing at least n minutes scored as sleep stage 1 or greater with no more than 1 minute of wakefulness intervening". The same criteria were applied to both the PSG and actigraph records. Values of n tested were 1, 10, 15 and 20. The 20-minute criterion showed the highest correlation between PSG and actigraph scores (Table 4), and so was used in all further analyses on the training and validation samples (Table 5). It is of interest to note, however, that the 20-minute actigraph criterion also correlated very highly with the 15-minute PSG criterion (training: $r = 0.91$, validation: $r = 0.90$). In addition, in both the training and validation samples, the mean sleep latency by the 20-minute actigraph criterion was about 2 minutes closer to the mean sleep latency by the 15-minute PSG criterion than to the mean sleep latency by the 20-minute PSG criterion.

Using the 20-minute criterion for both actigraph and PSG, actigraph sleep latencies were, on average, 4.7 minutes shorter than PSG latencies in the training sample and 9.1 minutes shorter in the validation sample. Neither of these differences was statistically significant (Table 5). This, along with the very high correlations between PSG and actigraph sleep latencies in both samples, shows that the actigraph provides a reliable estimate of sleep onset time.

Actigraphy overestimated total sleep time by an average of 21.5 minutes per night in the training sample and 19.4 minutes per night in the validation sample (Table 5). Although these mean differences were both statistically significant, the correlations between sleep percent (total sleep time × 100/total minutes scored) by actigraphy and sleep percent by polysomnography were quite high. This shows that, although actigraphy tended to overestimate sleep time, it did so consis-

TABLE 5. *Sleep parameters scored by polysomnograph (PSG) vs. actigraph*[a]

	Training sample			Validation sample		
	PSG[b]	Actigraph[b]	r[c]	PSG	Actigraph	r
Minutes scored	427.0 ± 74.7	427.0 ± 74.7		445.5 ± 43.1	445.5 ± 43.1	
Total sleep time (minutes)[d]	308.0 ± 93.5	329.5 ± 95.1*	0.91***	344.5 ± 52.5	363.9 ± 54.4*	0.77***
Percent sleep (%)[d,e]	71.4 ± 20.3	76.7 ± 20.1*	0.89***	77.9 ± 13.0	82.4 ± 13.5*	0.82***
Sleep efficiency (%)[d,e]	80.4 ± 18.7	84.4 ± 19.4	0.85**	85.0 ± 11.6	88.6 ± 11.4	0.71**
Sleep latency (minutes)[f]	75.6 ± 89.7	70.9 ± 102.2	0.94***	59.2 ± 46.1	50.1 ± 50.7	0.90***
Wake time after sleep onset (minutes)[f]	51.3 ± 45.6	40.2 ± 32.7	0.49*	49.9 ± 37.2	36.6 ± 30.6*	0.63**

[a] Actigraph sleep/wake identification was performed using the algorithm for the maximum 10-second overlapping epoch per minute. Rescoring was performed using the five Webster et al. rules described in the text.

[b] Means ± SD. Asterisks indicate significant differences between PSG and actigraph by paired t test.

[c] Pearson correlation between PSG and actigraph scores. Asterisks indicate significant correlations.

[d] Includes all minutes scored as sleep, including those which occurred before 20-minute criterion for sleep onset was met.

[e] Percent sleep is based on the entire record. Sleep efficiency is based on time in bed.

[f] Both PSG and actigraph sleep onset defined as beginning of first interval containing 20 minutes scored as sleep with no more than one minute of wakefulness intervening. Latencies are computed from the start of simultaneous actigraph/PSG recording, regardless of whether the subject was attempting to sleep.

* $p < 0.05$, ** $p < 0.002$, *** $p < 0.0001$.

tently, so that longer sleep time estimates were reliably associated with more actual sleep and shorter sleep time estimates were associated with less actual sleep.

Sleep efficiency (total sleep time × 100/total minutes in bed with lights out) means estimated from actigraphy did not differ significantly from sleep efficiency means obtained by polysomnography (Table 5). The correlations between actigraphic and polysomnographic sleep efficiencies were similar to those for sleep percent, but slightly lower.

The actigraphic estimate of wake time after sleep onset was, on average, 11.1 minutes per night shorter than the PSG score in the training sample and 13.3 minutes shorter in the validation sample (Table 5). Only the validation sample difference was statistically significant. The correlations between the actigraph and PSG scores on this parameter were significant, but were lower than those for other sleep parameters. This was to be expected because calculations of wake time after sleep onset compounded errors due to minute-by-minute actigraph/PSG differences in sleep/wake scoring with errors due to actigraph/PSG discrepancies in estimated sleep onset time.

Application of Webster's algorithm

Wrist activity transducers change as technology improves. Therefore, it is of interest to know whether a sleep/wake identification algorithm optimized for one transducer can be adapted for use on another. Webster et al. (2) developed an algorithm similar to ours using data collected with a different transducer. Their algorithm was:

$$D = 0.025(0.15A_{-4} + 0.15A_{-3} + 0.15A_{-2} + 0.08A_{-1} + 0.21A_0 + 0.12A_{+1} + 0.13A_{+2})$$

We applied Webster's algorithm to our data (10-second overlapping epochs), holding the weights (W) of the various minutes constant and varying only the scaling factor (P). By this method, we found that with the scaling factor optimized for the training sample at 0.04146, we obtained an accuracy rate of 86.44% before rescoring and 87.24% after applying the five rescoring rules. In the validation sample, these parameters produced an accuracy rate of 87.54% before rescoring and 87.73% after rescoring. This is not far from the accuracy rate produced by our best algorithm, suggesting that the minute-by-minute parameter weights for scoring wrist activity remain useful (if not quite optimal) from one transducer to the next.

Algorithms for other epoch lengths

Wrist activity data were collected in 2-second epochs to allow us to collapse the data into a variety of longer epoch lengths for algorithm development. Actigraphs in actual use in the field are often programmed to save activity data only once every 30 seconds or once per minute in order to conserve memory and thereby monitor for longer periods of time. Data collected in 30-second or 1-minute epochs cannot yield a maximum 10-second activity score per minute for use in our optimal algorithm. Even data collected in 10-second nonoverlapping epochs cannot be scored optimally using our algorithm for 10-second overlapping epochs. Therefore, we used the iterative method to optimize separate algorithms for scoring sleep/wake from the mean 2 seconds of activity per minute, maximum 30 seconds of activity per minute (nonoverlapping epochs), and from the maximum 10 seconds of activity per minute (nonoverlapping epochs).

The optimal parameters for the mean activity per minute were:

$$D = 0.001(106A_{-4} + 54A_{-3} + 58A_{-2} + 76A_{-1} + 230A_0 + 74A_{+1} + 67A_{+2})$$

The optimal parameters for the maximum 30-second nonoverlapping epoch of activity per minute were:

$$D = 0.0001(50A_{-4} + 30A_{-3} + 14A_{-2} + 28A_{-1} + 121A_0 + 8A_{+1} + 50A_{+2})$$

The optimal parameters for the maximum 10-second nonoverlapping epoch of activity per minute were:

$$D = 0.00001(550A_{-4} + 378A_{-3} + 413A_{-2} + 699A_{-1} + 1{,}736A_0 + 287A_{+1} + 309A_{+2})$$

Results for these algorithms are shown in Table 6. The mean-1-minute algorithm always produced accuracy rates below those obtained with the other two algorithms. In the training sample, the percentage of correct sleep/wake identifications produced with the maximum-30-second algorithm exceeded that produced with the 10-second overlapping algorithm, but in the validation sample the 30-second algorithm performed more poorly than the 10-second overlapping algorithm. The 10-second nonoverlapping algorithm performed almost as well as the 10-second overlapping algorithm in both the training sample and validation sample.

Subject subgroups

Our 41 subjects included 15 young or middle-aged normal controls and 26 others. All of the others (in-

TABLE 6. *Sleep/wake identification accuracy[a] by optimized algorithms at various epoch lengths*

Algorithm[b]	Training sample		Validation sample	
	Before rescoring[c]	After rescoring	Before rescoring	After rescoring
Max. 10-second overlapping epoch	87.05%	87.93%	87.91%	88.25%
Max. 10-second nonoverlapping epoch	86.80%	87.62%	87.72%	87.97%
Max. 30-second nonoverlapping epoch	87.34%	88.07%	86.83%	86.93%
Mean activity per minute	86.67%	87.39%	86.23%	86.00%

[a] Values in table represent the percent of minutes in which the sleep/wake score derived from wrist activity agreed with the polysomnographic sleep/wake score.
[b] Scoring algorithms are described in the text.
[c] Rescoring was performed using the five Webster et al. rules described in the text.

cluding healthy elderly) had conditions that might be expected to impair sleep. Therefore, it is of interest to know how well our optimal algorithm identified sleep/wake in various subcategories of subjects. These results are presented in Table 1.

Note that our scoring method performed approximately as well on psychiatric patients as on normals. Its performance on apnea patients, DIMS patients and elderly normals was only slightly poorer. Differences between patient subgroups should be interpreted with caution, however, because the number of patients in each group was small. Among normal control subjects, the correlation between sleep percent by actigraphy and sleep percent by polysomnography was 0.94 in the training sample ($p < 0.0005$) and 0.72 in the validation sample ($p < 0.07$).

Pooling all subjects, our optimal scoring method produced accuracy rates above 90% on 19 of 43 recording nights. Accuracies below 80% were obtained on only six nights. Only one record's score fell below 73%. The highest accuracy for any individual was 97.98% for a 54-year-old bereaved widow. Oddly, the lowest accuracy (59.16%) was also for a bereaved widow (age 59), who remained awake but motionless for much of the night.

DISCUSSION

The automatic scoring algorithm for wrist activity data prospectively distinguished sleep from wakefulness with over 88% accuracy in a mixed sample including normal controls, elderly individuals, sleep disorders patients, psychiatric patients and others. This demonstrates that the wrist actigraph can be a useful tool for sleep studies in a wide variety of subjects.

Actigraphic estimates of sleep latency did not differ significantly from actual PSG sleep latency scores, indicating that the actigraph is an excellent tool for identifying sleep onset times. It should be noted, however, that the actigraph was most reliable for identifying the onset of PSG sleep periods at least 15 minutes in duration. Thus, the actigraph is best at identifying latency to persistent sleep.

Actigraphic estimates of total sleep time and sleep efficiency correlated highly with PSG scores, demonstrating that the actigraph provides reliable information about these parameters. Actigraphic estimates of wake time after sleep onset were somewhat less reliable, but still provided much useful information. The actigraph's systematic tendency to overestimate sleep should be taken into account when comparing actigraph to PSG total sleep time, sleep efficiency and wake-after-sleep-onset scores. It might prove worthwhile in some situations to modify the scoring algorithm to automatically reduce actigraph estimates of these parameters by an amount proportional to the duration of recording. Caution must be used in applying such methods, however, because it has been reported that actigraphy may underestimate sleep in at least one subject subgroup (DIMS without objective findings) (5).

Our study replicated the finding of Webster et al. (2) that taking the maximum activity value within a minute provides a better indicator of sleep/wake than taking the average activity per minute. Simple rescaling of the scoring algorithm developed by Webster et al. produced a high rate of correct sleep/wake identification in our sample, suggesting that the rules for scoring sleep/wake from wrist activity are quite robust and may remain fairly constant across different activity transducers and subject samples.

Our results are remarkably consistent with those of other investigators who used very different algorithms to score sleep/wake from wrist activity. Sadeh et al. (6) compared wrist activity to polysomnography using the same Motionlogger actigraph model employed in the present study. They used a discriminant function analysis on overnight data from nine healthy normal subjects to obtain an automatic scoring algorithm. Their algorithm differed qualitatively from ours in that they considered not only the raw activity scores surrounding the minute under consideration, but also the standard deviation and minimum value of these scores. Furthermore, they considered activity that occurred up to 9 minutes before the minute being scored, whereas we considered only 4 minutes of prior activity. Despite

the differences between our methods, their results were strikingly similar to ours. Their accuracy rates were 91.8% in their normal training sample (compared to 91.87% in our training normals), 86.2% in four validation normals (compared to 87.97% in our validation normals), 85.7% in 25 sleep apnea patients (compared to 83.89% in our apnea validation subjects), 78.2% in 16 insomniacs (compared to 89.71% in our DIMS patients) and 89.9% in 13 children with various sleep problems (no comparison group in our sample). Dunham et al. (7) applied another algorithm, based on the natural logarithm of activity scores from two consecutive 32-second epochs, to Motionlogger actigraph data in 11 young normal subjects. They reported 96.9% correct sleep/wake identification on repeat records from the same subjects. Levine et al. (8) applied a simple threshold algorithm to 5-minute average wrist activity data collected by an experimental actigraph that used a bank of mercury tilt switches as its transducer (as compared to the Motionlogger's piezoelectric transducer). They reported 92.6% correct sleep/wake identification during a single overnight sleep in seven young normal subjects. The similarity between our results and those of other investigators provides further evidence that automated scoring of sleep from wrist activity is relatively insensitive to details of the algorithm or to the nature of the transducer, although the excellent results of Dunham et al. suggest that algorithms based on the logarithm of activity deserve further scrutiny. This strengthens the conclusion that the technique can produce reliable results.

Our average accuracy for scoring sleep/wake from wrist activity (about 88%) was not as high as that reported by Webster et al. (94–96%) (2), despite similar algorithms. Several differences between our studies may have contributed to this discrepancy: 1) Webster et al. used 24-hour recordings, which contained a great deal of unambiguous daytime wakefulness. Much of the wakefulness in our overnight data occurred near wake/sleep or sleep/wake transitions, making epochs difficult to characterize clearly, even on the polysomnogram. Misscoring of some of this ambiguous wakefulness as sleep by our algorithm reduced our rate of correct sleep/wake identification. 2) Webster et al. studied primarily young, normal controls who presumably slept well. Our sample was mostly middle aged, and included many individuals suffering from conditions that disrupt sleep. 3) Although we made every effort to synchronize our actigraph data collection epochs with our polysomnograph sleep scoring epochs, we did not use a direct electronic linkage between our all-on-the-wrist instruments and the polygraph, as Webster et al. did. Therefore, there was an error of up to 2–4 seconds in epoch synchronization in some subjects. This probably reduced scoring accuracy somewhat.

In spite of these limitations, the wrist actigraph compared favorably to other automated methods of sleep/wake determination. Using an activity transducer mounted on a headband plus a movement transducer taped to the eyelid, Mamelak and Hobson (9) reported 85.57% correct identification of wake, nonrapid eye movement (NREM) sleep and REM sleep in four healthy young subjects. Kubicki et al. (10) reported that automatic scoring of polysomnograms with the Oxford Sleep Stager correctly distinguished wake from sleep 94.19% of the time in 10 healthy men ages 20–40. The actigraph method is considerably less obtrusive and easier to use than either of these systems, yet it outperformed the headband system and was not far below the performance of the Oxford system.

Within a given individual, night-to-night variability in sleep parameters is very high. For example, Moses et al. (11) found that even in young, healthy subjects habituated to the sleep laboratory, the correlation between sleep efficiency (manually scored from the polysomnogram) from one night to the next was only 0.304. Furthermore, Hauri (12) reported that insomniacs' sleep is far more variable at home than in the laboratory, apparently because the perceived social demands of the laboratory setting make patients reluctant to spend either very short or very long times in bed, as they often do at home. Therefore, a truly accurate assessment of the quantity and quality of an individual's sleep requires that recordings be made over several nights, preferably at home.

It is much more convenient and less expensive to perform multi-night sleep recordings with a wrist actigraph than with a polygraph. Given the high night-to-night variability in human sleep, it is possible that, despite the actigraph's lower within-night scoring accuracy, a wrist activity record many nights in duration may yield a more accurate picture of a subject's typical sleep and wakefulness than the usual one to three nights of clinical polysomnography. This may be especially true when actigraphy is performed in the home environment and polysomnography is performed at the clinic or laboratory. Of course, the polygraph can measure physiological functions that the actigraph cannot. To realize the best of both worlds, a combination of actigraphy and polysomnography may be particularly advantageous. Still, in many research and clinical applications where simple sleep/wake estimation is the goal, the wrist actigraph by itself may be the instrument of choice. In fact, the actigraph has already proven useful for documenting sleep/wake changes in studies of drug effects (12–16,19), insomnia (4,11,12), jet lag (13,14), shift work (16), sleep apnea (17), sleep disorders of infancy and childhood (18), psychiatric disorders (13,19) and narcolepsy (13).

Acknowledgments: This work was supported by NIMH grants MH39880, MH30914, MH44943, MH00117, MH18398 and by the Department of Veterans Affairs.

REFERENCES

1. Mullaney DJ, Kripke DF, Messin S. Wrist-actigraphic estimation of sleep time. *Sleep* 1980;3:83–92.
2. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. *Sleep* 1982;5:389–99.
3. Spitzer AL. *Diagnostic and statistical manual of mental disorders DSMIIIR.* Washington DC: American Psychiatric Assoc., 1987.
4. Rechtschaffen A, Kales A. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Bethesda, Maryland: U.S. Dept. of Health, Education, and Welfare, 1968.
5. Hauri PJ. Wrist actigraphy in insomniacs. *Sleep Research* 1989; 18:239.
6. 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.
7. Dunham DW, Hoffmann RF, Broughton RJ. Wrist actigraphy and sleep/wake estimation revisited. *Sleep Research* 1991;20: 491.
8. Levine B, Moyles T, Roehrs T, Fortier J, Roth G. Actigraphic monitoring and polygraphic recording in determination of sleep and wake. *Sleep Research* 1986;15:247.
9. Mamelak A, Hobson JA. Nightcap: home-based sleep monitoring system. *Sleep* 1989;12:157–66.
10. Kubicki St, Höller L, Berg I, Pastelak-Price C, Dorow R. Sleep EEG evaluation: a comparison of results obtained by visual scoring and automatic analysis with the Oxford sleep stager. *Sleep* 1989;12:140–9.
11. Moses J, Lubin A, Naitoh P, Johnson LC. Reliability of sleep measures. *Psychophysiology* 1972;9:78–82.
12. Hauri PJ. Laboratory and home sleep in insomniacs. *Sleep Research* 1989;18:238.
13. Borbély AA, Loepfe M, Mattmann P, Tobler I. Midazolam and triazolam: hypnotic action and residual effects after a single bedtime dose. *Arzneimittelforschung* 1983;33:1500–2.
14. Borbély AA. New techniques for the analysis of the human sleep-wake cycle. *Brain Dev* 1986;8:482–8.
15. Lavie P. Effects of midazolam on sleep disturbances associated with westward and eastward flights: evidence for directional effects. *Psychopharmacology* 1990;101:250–4.
16. Lavie P, Lorber M, Tzischinsky O, Epstein R, Sharf Y. Wrist actigraphic measurements in patients with rheumatoid arthritis: a novel method to assess drug efficacy. *Drug Invest* 1990;2(Suppl 3):15–21.
17. Walsh JK, Schweitzer PK, Anch AM, Muehlbach MJ, Jenkins NA, Dickins QS. Sleepiness/alertness on a simulated night shift following sleep at home with triazolam. *Sleep* 1991;14:140–6.
18. Aubert-Tulkens G, Culeé C, Harmant-van Rijckevorsel K, Rodenstein, DO. Ambulatory evaluation of sleep disturbance and therapeutic effects in sleep apnea syndrome by wrist activity monitoring. *Am Rev Respir Dis* 1987;136:851–6.
19. Sadeh A, Lavie P, Scher A, Tirosh E, Epstein R. Actigraphic home-monitoring sleep-disturbed and control infants and young children: a new method for pediatric assessment of sleep-wake patterns. *Pediatrics* 1991;87:494–9.
20. Klein E, Mairaz R, Pascal M, Hefez A, Lavie P. Discontinuation of lithium treatment in remitted bipolar patients: relationship between clinical outcome and changes in sleep–wake cycles. *J Nerv Ment Dis* 1991;179:498–500.