How Accurately Does Wrist Actigraphy Identify the States of Sleep and Wakefulness?

Charles P. Pollak,[1] Warren W. Tryon,[2] Haikady Nagaraja,[3] and Roger Dzwonczyk[4]

[1]Division of Sleep Medicine, Department of Neurology, The Ohio State University; [2]Department of Psychology, Fordham University; [3]General Clinical Research Center, The Ohio State University; [4]Department of Anesthesiology, The Ohio State University

Study Objectives: Because sleep and wakefulness differ from each other by the amount of body movement, it has been claimed that the two states can be accurately distinguished by wrist actigraphy. Our objective was to test this claim in lengthy polysomnographic (*psg*) and actigraphic (*act*) samples that included night and day components.

Design: Fourteen healthy young (21—35 years) and old (70—72 years) men and women lived in a laboratory without temporal cues for 7 days. Each subject continuously wore sleep-recording electrodes as well as 2 wrist-movement recorders. *Act* measurements were converted to predictions of sleep and wakefulness by simple-threshold and multiple-regression methods. *Psg* served as the gold standard for calculation of predictive values (PV, the probability that an act prediction is correct by *psg* criteria).

Setting: N/A

Participants: N/A

Interventions: N/A

Measurements and Results: The 7-day *act* recordings showed clear circadian cycles of high and low activity that respectively corresponded to subjective days, when subjects were wakeful, and subjective nights when they slept. Lower *act* levels corresponded to deeper states of psg sleep. Logistic regression on a 20-minute moving average of *act* gave the highest overall PV's. Nevertheless, the mean PV for sleep (PVS) was only 62.2% in complete, day + night samples. PVS was 86.6% in night samples. *Act* successfully predicted wakefulness during subjective nights (PVW = 89.6) and accurately measured circadian period length and the extent of sleep-wake consolidation, but it overestimated sleep rate and sleep efficiency. *Act* systematically decreased before sleep onset and increased before awakening, but reliable transitions among joint *psg/act* states (the Markov-1 property) were not demonstrated.

Conclusions: Low PV's and overestimation of sleep currently disqualify actigraphy as an accurate sleep-wake indicator. Actigraphy may, however, by useful for measuring circadian period and sleep-wake consolidation and has face validity as a measure of rest/activity.

Key words: Actigraphy; motor activity; rest/activity; sleep; wakefulness; sleep disorders; circadian; predictive value; Markov property

INTRODUCTION

AS AN ANIMAL FALLS ASLEEP, IT NEARLY STOPS MOVING. Immobility at intervals of about one solar day is a universal feature of sleep but, in animals whose behavioral states are generated by a complex nervous system, "sleep" is preferably defined neuroelectrically. In humans, a small set of joint EEG, EOG, and EMG patterns has been codified[1] and, for all practical purposes, constitutes the definition of sleep and wakefulness.

Actigraphy (*act*) is the continuous recording of body (often wrist) movement by means of a body-worn device that detects movement (usually acceleration) and stores the information for days, weeks, or months, along with the times it was measured.[2] The measurement of rest-activity cycles rank among the earliest and most widely used chronobiological techniques in birds and mammals, including humans.[3-6]

The parallelism between the sleep-wake and rest-activity rhythms has suggested that actigraphic levels of activity can be substituted for the standard, polysomnographic (*psg*) techniques whenever sleep-wake (not sleep-stage) information is required.[7-8] It has also been claimed that actigraphic measures can be used to estimate sleep amounts and sleep continuity in patients with sleep disorders.[9] If actigraphy could indeed emulate standard sleep-recording methods, it would liberate research subjects from the confines and labor-intensive techniques of the sleep laboratory. Sleep recordings could then be greatly extended in time, a development that would be valuable, for example, for clinical investigations of insomnia. Unfortunately, the evidence upon which such claims are based has been limited or flawed in several ways.

Equal Priority of *Act* and *Psg* Measures

Previous efforts to relate actigraphic data to electrographic data have often used the agreement rate (proportion of observations for which actigraphic and electrographic ratings agree) as a criterion of success.[7-8] Motor activity and electrographic sleep-wake state are thereby treated as if they were alternative and equally valid measures of neurobehavioral state. As we have seen, however, EEG/EOG/EMG constitute the established standard, and most sleep experts would agree that these *psg* measures are a richer and more sensitive measure of state than body movements. This argues for making *psg* ratings the standard by which *act* ratings are judged. To assess the value of actigraphy, the prospective user then requires the predictive value of actigraphy—the fraction of actigraphic determinations of sleep and wakefulness that are correct by the *psg* standard—and not the agreement rate.

Limitation of Samples to Sleep

Previous validation studies have used actigraphic samples collected largely during electrographic sleep.[7-8] As will be seen, however, predictive values from such samples approximate the true or base rates of electrographic sleep and wakefulness and cannot be applied to samples obtained during wakefulness. Similar considerations apply to the detection of wakefulness during sleep periods. In the present study, complete circadian sleep-wake cycles—consisting of approximately 25% sleep and 75% wakefulness—were sampled, and predictive values for sleep (PVS) and wakefulness (PVW) were calculated for the complete cycles as well as separately for the night and day components of the cycles.

Autocorrelation of *Act* and *Psg* Data

Like many longitudinal biological observations, actigraphic and electrographic sleep-wake date are autocorrelated—large observations (those greater than the mean) are likely to be immediately followed by large observations, small ones by small ones. The assumption that successive actigraphic and electrographic observations are independent, on which estimates of the variances of predictive values and other measures are based, are therefore not met.

Missing *Act* Data

Missing *act* data may not always be apparent, because unworn and therefore motionless activity recorders continue to store "data" values of zero. Such noncompliance can be an important problem in studies of demented or otherwise impaired subjects,[10] and it occurred even in the normal subjects and controlled conditions of the present experiment.

One of our objectives was to meet these methodological challenges. The lengthy periods of observation used in this investigation also made it necessary to solve two additional methodological problems.

Nonstationarity

The statistical properties of a nonstationary time series change over time. During circadian waking periods, for example, both the mean and the variability of motor activity are larger than the levels recorded during sleep. The solution adopted here was to segment the data into circadian wake and sleep periods, each of which is more stationary than the intact series.[11]

Nonsynchronicity of *Act* and *Psg* Data

Valid comparison of electrographic and actigraphic data requires that they be synchronous. This is not a major problem for comparisons of global measures such as sleep efficiency. Depending on the relative rates of the clocks (or paper speed) of the polygraphs and activity recorders that were employed, it may also be of minor importance for short (one-night) samples. In the present, seven-day study, however, even small differences in clock rates could result in cumulative timing errors that had to be measured and corrected.

The aim of this study was to measure the accuracy with which actigraphic motility patterns predict *psg* sleep and wakefulness

and thereby assess the increasingly common practice of substituting *act* for the standard laboratory technique. We also examined the alternate possibility, that actigraphic movement patterns contain unique information that supplements that of the *psg*. Our results confirmed the accuracy of actigraphic measurements of circadian sleep-wake cycles, but the predictive power of actigraphy circa 2001 remains too low for it to be accepted as a surrogate for the standard sleep-laboratory techniques. We did obtain preliminary evidence that actigraphic data may usefully supplement standard *psg* data.

METHODS

Subjects

Act and *psg* data were continuously collected for seven solar days and nights from men and women residing in the Chronobiology Laboratory of The Ohio State University. Subjects were recruited by advertisement for a study of sleep need as it may change with age. Inclusion criteria included age (either 20—35 years or 70—85 years), good health, and absence of complaints regarding either sleep or daytime alertness. The state of good health was confirmed by physical examination, routine laboratory tests, urine drug screening, and absence of significant sleep disorders during the first sleep period in the laboratory, when a full clinical sleep recording was performed. The study was approved by the Institutional Review Board of The Ohio State University, and all subjects provided written consent.

Procedure

Throughout their stay in the laboratory, subjects were not given access to information regarding the time of day or to any means of measuring the passage of time (temporal isolation). They were allowed to initiate and terminate bedrest periods, eat meals, and engage in other activities whenever they chose ("free-running"). During periods of bedrest, room lights were extinguished by the investigators (0–5 lux).

Sleep-recording electrodes were worn continuously except for brief periods in the shower. EEG, EOG, chin EMG, and ECG signals were led by wire to a digital polygraph (4150 Sleep WorkStation, Sensormedics Corp) located outside the time-free environment. During the first or second night, respiratory airflow, thoracoabdominal respiratory movements, oxygen saturation and anterior tibial EMG were also recorded to detect any sleep disorders. Subjects with medically significant or symptomatic sleep disorders were excluded. Sleep recordings were manually analyzed in 30-second epochs using standard criteria.[1] The resulting states were: wakefulness, nonREM sleep stages 1—4 and REM sleep. The Sensormedics data-acquisition software made it necessary to start a new file about once every 24 hours. This was usually done while electrodes were removed for showering. The scored *psg* data were then concatenated to form a continuous time series.

Actigraphic data were collected at 30-second intervals by two recorders worn next to each other on the nondominant wrist: the CSA (Model 7164 Activity Monitor, Computer Science and Applications, Inc., Shalimar, Florida) and the IM (ActiTrac, IM Systems, Inc., Baltimore, Maryland). The CSA instrument uses an accelerometer that responds to both intensity and frequency of

movement.[12] Its signal was sampled at 10 Hz, amplified, pass-band filtered (0.1 – 3.6 Hz) and led to an 8-bit A/D converter that discriminated 256 activity levels. Its effective threshold was 0.033 G. The IM instrument sampled accelerometer output at 40 Hz, passed data between 0.25 Hz and 10 Hz and also used an 8-bit A/D converter. When its gain was set to 0.3125 mG, its 250 (of 256 possible) activity levels encompassed 250x.3125 = 78 mG, and its effective sensitivity was better than 0.024 G. PC's running proprietary software were used to initialize the recorders and download data. Both recorders had sufficient memory to make it unnecessay to download data during the seven-day experiment. Both recorders were small (22.4, 29.1 cm³) and light (35, 42.6 g) and were worn continuously by the subjects, except in the shower. Their positions were reversed after 3—4 days to ensure that on average they were subjected to equivalent kinetic forces.

Because subjects were observed by CCTV, it is unlikely that they removed the recorders at unauthorized times. This possibility nevertheless had to be considered, because a recorder lying unworn continues to record zero activity, which may be mistaken for the relative physiological immobility of sleep. Periods of noncompliance have indeed been inferred from runs of zero activity that outlasted those thought to be of biological origin.[10] The availability in this study of a simultaneous electrographic measure of sleep-wake state made it possible to measure the maximum duration of zero activity during periods of verified sleep, when we could be confident that the recorders had not been removed. Runs of zero-activity that outlasted the sleep maximum were assumed to represent noncompliance and were replaced by the means of activity levels recorded at the same times on other days.

To ensure that *psg* and *act* data spaced at nominal 0.5 minute intervals remained synchronous, the rates of the Sensormedics polygraph and activity-recorder clocks were compared, and binary *psg* data were replaced by the values that were present at the start of the activity-data intervals during which they fell. Thus, "sleep" was recorded if the subject was last observed to be asleep; otherwise, "wake" was recorded. In a preliminary series of seven-day trials, the clock rates of four recorders relative to the polygraph were measured. As a further check, predictive values (described below) based on the first two days of data were compared to those of the last two days, by which time any cumulative timing errors would have been maximal.

Data Analysis

Each *act* and *psg* time series typically consisted of 19—20,000 observations (7days x 1440minutes/day x 2observations/minute). Rather than using the proprietary software supplied by the manufacturers of the movement recorders, the *act* data (as well as the *psg* data) were analyzed on a UNIX-workstation using functions written in the S language by the first author.[13]

For all analyses, sleep data were reduced to binary form (0=wakefulness, 1=any stage of sleep). Activity data were likewise reduced to binary form by one of two methods. The simpler method was to replace each measurement of activity with a 0 if it exceeded a threshold level and a 1 otherwise. The threshold was determined by the observed probability of sleep and wakefulness as functions of activity level. The second method was logistic

regression.   Logistic regression models were developed in a training sample of four randomly selected subjects. The models ranged from simple (1—10 terms selected for interest or by step-wise addition and subtraction of terms) to complex (up to 41 terms, including terms progressively incremented and decremented by 0.5-minute lags). The terms of the logistic regression models consisted of either single or multiple activity values or the mean or standard deviation of a series of activity values. The activity values were consecutive and ranged from 10 minutes before to 10 minutes after the corresponding *psg* epochs. The motor activity levels predicted by the models were distributed in bimodal fashion, and a cutting value cleanly divided them into predictions of sleep (≤0.2 counts/minute) and wakefulness (>0.2 counts/minute).

The *psg* and *act* time series were related to each other in four ways: 1) Ordering of *psg* sleep stages by the mean level of *act* with which each was synchronous, 2) relative magnitudes of predictive values (PV's), 3) relative magnitudes of global sleep-wake measures, and 4) inferences from joint, *psg/act*, state transition probabilities.

Analyses were carried out in complete data sets ("circadian" samples), subsets of data pooled within subjects from periods spent in bed with lights off ("nights"), and subsets pooled from periods spent out of bed with lights on ("days").

For the calculation of PV's, the binary *psg* and *act* data were cross-tabulated, as shown in the example in Table 1, and the following ratios were calculated: predictive value for sleep (PVS, percent of actigraphic sleep predictions that accorded with *psg* sleep), predictive value for wakefulness (PVW, percent of wake predictions that accorded with wakefulness), and agreement rate (Ag, percent of observations for which *psg* ratings accorded with *act* ratings). The logistic regression models were ranked according to their ability to yield high PVS and PVW in the training sample. Because they are correlated, the PV's were entered into a MANOVA with two factors, condition (day + night, night, day) and prediction method (simple threshold, logistic regression and their interaction). The design was unbalanced, because it excluded the four members of the training sample. The MANOVA was also re-run using the full sample. The contrasts cited in Table 2 were based on paired or unpaired t-tests and tested for statistical significance ($\alpha < 0.05$) by α adjusted for multiple tests by the Bonferroni procedure.[14]

The global sleep-wake measures were circadian period length, sleep rate, consolidation and sleep efficiency. Each was calculated from binary *psg* data and from *act* data converted to binary form by both the simple-threshold and logistic-regression methods. The period of each subject's circadian activity and sleep-wake rhythm was obtained by the following, iterative procedure. A chi-square periodogram[15] was calculated over the circadian range, 26±10 hours. Its peak served as the center of a second periodogram spanning a narrower range (peak±4 hours). The desired peak was the peak of the second periodogram. Sleep rate and consolidation were derived from a linear regression model fitted to a cumulative, seven-day sleep function (cumulative minutes of sleep vs. cumulative minutes of wakefulness) after confirming that the cumulative function was linear. Sleep rate is the rate at which sleep accumulated and is derived from the slope of the linear model. Consolidation is a measure of the deviations of the cumulative sleep function from the rate of accumulation of sleep predicted by the model. It represents the degree to which

Table 1—*PSG* and predicted sleep-wake states for one subject

	PSG		
	Sleep	**Wakefulness**	**Total**
Actigraphic prediction of sleep	6281	4244	10525
Acitgraphic prediction of wakefulness	555	6636	7191
Total	6836	10880	17716
Base rate	38.6%	61.4%	

Table entries are frequencies of 30-second epochs.
Data collected during subjective nights and days
Predictive value for sleep (PVS)=100x6281/10525=59.7%
Predicitve value for wakefulness (PVW)=100x6636/7191=92.3%
Sensitivity of activity reorder to sleep=100x6281/6836=91.9%
Specificity of recorder for sleep=100x6636/10880=61.0%
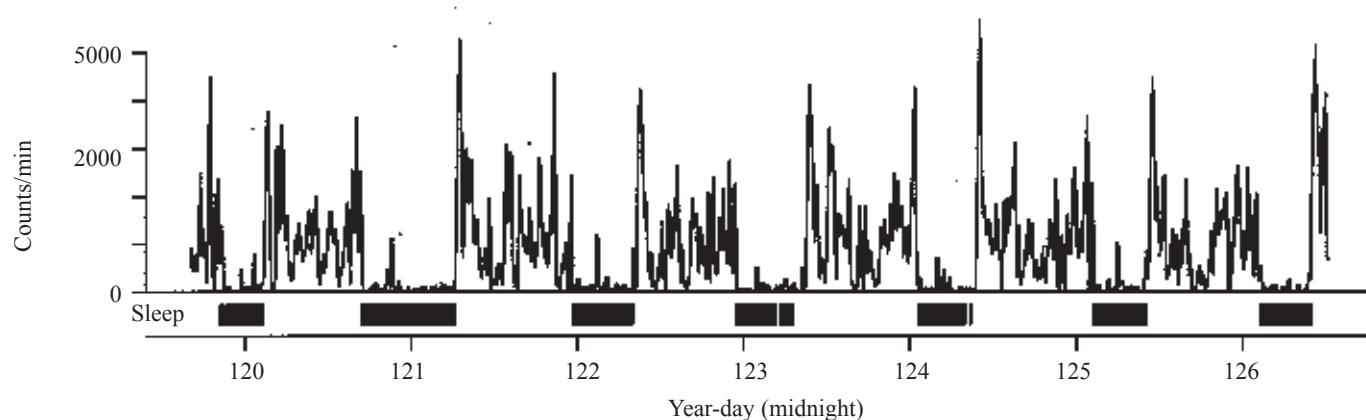Agreement rate=100x(6281+6636)/17716=72.9%

Figure 1—Representative plots of motor activity (upper panel) and sleep (bars, lower panel) of subject "oy15." Seven subjective night and about 7.5 subjective days of data are shown. Activity data were collected by the "CSA" recorder and were smoothed. As a result, the full range of activity levels is not shown.

sleep and wake bouts are consolidated into continuous blocks of circadian duration. Sleep efficiency is the percentage of bedrest time that subjects slept. Bedrest extended from the time subjects retired to bed with the lights off to the time they got up and lights were turned on.

It would be valuable to put confidence limits on the PV's obtained for each subject, but strong auto- and cross-correlation exhibited by the *act* and *psg* data invalidated the usual variance estimator for such data. The binary *psg* and *act* time series were therefore treated as a series of two-variable states defined by *psg* (sleep, S, or wakefulness, W) and *act* (sleep, s, or wakefulness, w). There were therefore four possible states: Ss, Ww, Sw, and Ws. A 4 x 4 matrix of transition probabilities among these states was calculated for each subject. In addition to being of interest in its own right, the matrices were used to test for the first-order Markov property, namely, that the probability of each transition (e.g. Ws $\rightarrow$ Sw) depends only the starting state (Ws) and not on any of the states that preceded it. Matrices were tested for the first-order Markov property by a chi-square test for order of dependence.[16] It was reasoned that a matrix with the Markov-1 property could then be used to generate a large number of simulated time series, from which PV's and their within-subjects vari-

ances could be calculated.

Finally, the orders of *psg* and *act* changes at times of falling asleep and waking up were examined. For each subject, a list was made of consecutive *psg act* states from which all repeats had been removed (i.e., state durations were ignored). Using a paired t-test, the number of Ww $\rightarrow$ Ws $\rightarrow$ Ss transitions as a proportion of the total number of two-step transitions was compared to the proportion of Ww $\rightarrow$ Sw $\rightarrow$ Ss transitions. Subjects were hypothesized to stop moving before they fell asleep more often than they stopped moving after falling asleep. A similar comparison was made between the frequencies of Ss $\rightarrow$ Sw $\rightarrow$ Ww and Ss $\rightarrow$ Ws $\rightarrow$ Ww to test the hypothesis that subjects started moving before rather than after waking up.

RESULTS

Fourteen subjects participated: five young men (21–24 years of age), five young women (22–35), two older men (both 70), and two older women (70–72). The data of an additional subject, a 70-year-old women, were excluded because she was sometimes found with wrist straps that were so loose that the recorders did not move with the wrist. None of the subjects showed any unusual features of sleep amount, sleep continuity or sleep-stage com-
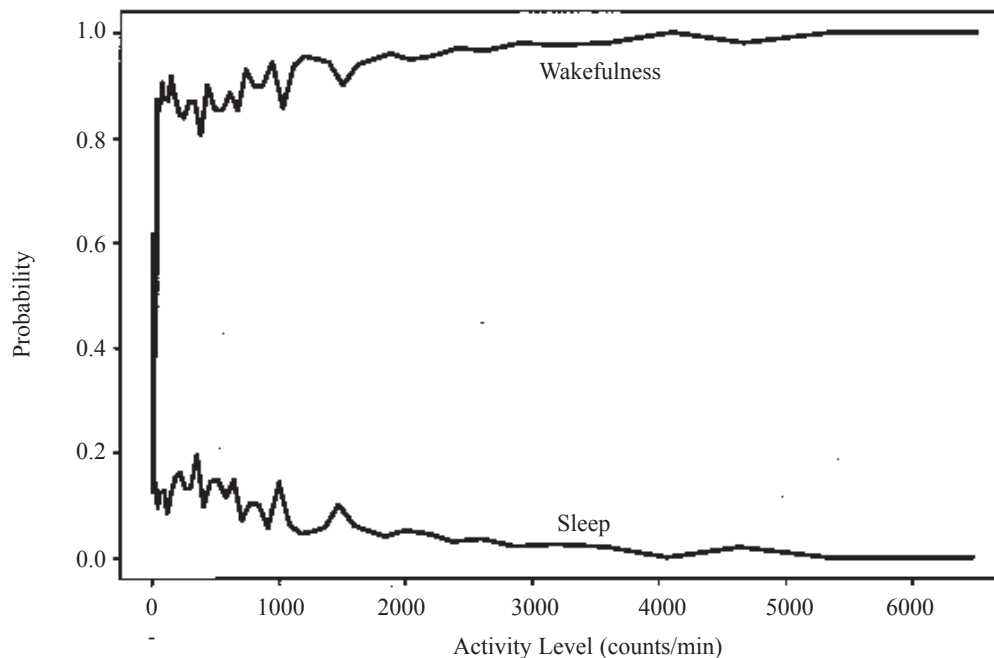
Empirical probabilities of sleep and wakefulness as functions of motor acitivty level

Figure 2—Empirical probabilities of *psg* sleep and wakefulness as functions of activity levels for subject "oy19." Abscissa shows *act* up to 99th percentile. The curve for wakefulness has been slightly displaced to the right for clarity. *Act* data were sorted into 100 ordered bins, and the probability of *psg* sleep was then calculated for each bin. The probability of wakefulness was 1 – probability of sleep. The highest probability of sleep (0.61 in this subject) corresponded to activity level = 0. The probability of sleep for all activity levels >0 was much smaller. Results were similar for all subjects.

position or any significant sleep disorders.

Circadian rest-activity cycles were clearly visible in the *act* data of all subjects with intervals of low activity coinciding with periods of *psg* sleep (Figure 1). The longest episodes of immobility (activity level=0) found in subjects during sleep ranged from 36 minutes to 88 minutes. Such episodes represented limb immobility rather than noncompliance, since recorders were not removed during periods of bedrest and sleep. Ninety minutes was therefore chosen as the cutoff between limb immobility and noncompliance. By this criterion, subjects had a mean of 0.4 episodes of noncompliance (range 0–3) over the seven-day experiment (8.6 minutes per solar day). Even if all of the artifactual zeroes were wrongly interpreted as sleep, the error in PVS for circadian samples was at most approximately 0.16%. The effect of noncompliance in this controlled laboratory study was therefore negligible.

The mean cumulative difference between the times told by the recorder and polygraph clocks over seven days was –4.58 to 1.25 minutes (95% confidence limits). PV's based on the last two days of data were not significantly lower than those of the first two days. Any loss of synchrony between the *act* and *psg* data over the course of the experiment was therefore slight.

The highest levels of motor activity were invariably associated with wakefulness. The next highest were usually associated with nonREM stage 1 (12 of 14 subjects), followed by REM sleep (11/14) and stage 2 (10/14). The lowest levels were associated with stages 3 or 4 (11 of 14 subjects). The measured *act* levels therefore corresponded to the depth of sleep.

The proportion of 0.5-minute *act* measurements equal to zero was large (group mean for circadian days = 52.0% for the CSA

recorder, 52.1% for the IM; for subjective nights, 84.5 for CSA, 89.7% for IM; for subjective days, 28.7% for CSA, 28.4% for IM). Most of the zeroes occurred in brief runs of 1 – 2 zeroes, especially during subjective days, when subjects were usually awake. Activity levels of zero were nevertheless the ones most likely to be associated with *psg* sleep (group mean=57.7%; for the data of one subject shown in Figure 2, 61%). Activity levels that were only slightly greater than zero were much less likely to be associated with sleep. A simple actigraphic prediction rule could therefore be adopted: predict sleep if motor activity = 0; predict wake otherwise. The simplest logistic regression model that performed reasonably well had a single term, the 20-minute moving average of *act* measurements.

PV's based on data from the two recorders were similar, and only results from the CSA recorder are presented. According to the MANOVA, PVS and PVW were significantly affected by day-night condition ($F_{(4,114)}$=91.0, p<.001), prediction method ($F_{(2,56)}$ = 110.3, p < .001) and interaction of condition with prediction method ($F_{(4,114)}$=17.6, p<.001). For the full sample (n=14), significant effects were again found for condition ($F_{(4,130)}$=80.7, p<.001), prediction method ($F_{(2,64)}$=130.6, p<.001) and their interaction ($F_{(4,130)}$=19.8, p<.001).

For both the simple-threshold and logistic-regression methods, mean PVS was highest in night samples (Table 2). Mean PVW was highest in day samples, approaching 100 %, and it was higher overall than PVS. PVS and PVW were each intermediate in circadian (day+night) samples. The samples, in turn, strongly differed in their base rates of sleep: Night samples consisted mainly of sleep (78.4% of observations), day samples included almost no sleep (1.0%), and circadian samples included less sleep
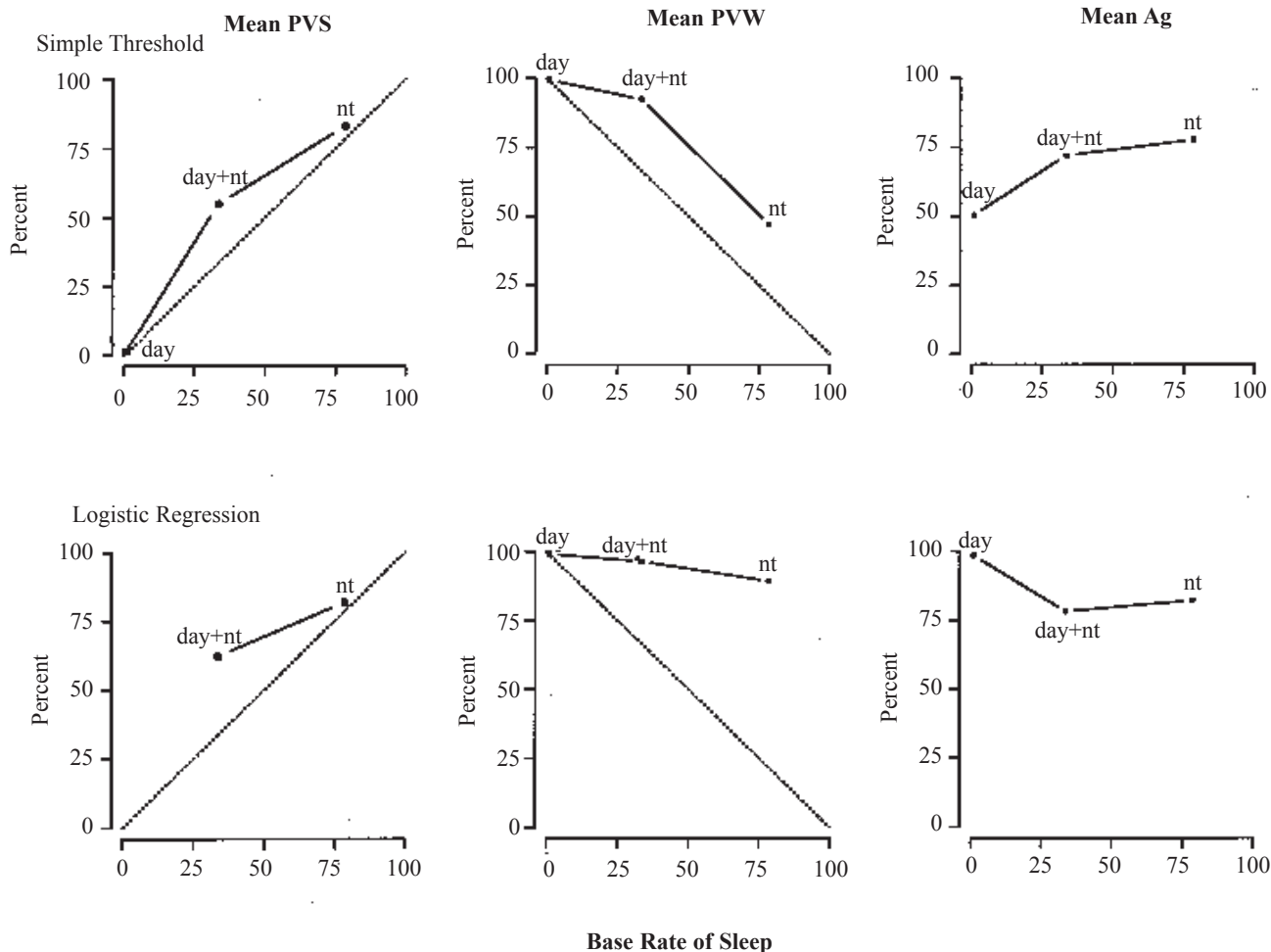
Figure 3—Dependence of mean predictive value for sleep (PVS), mean predictive value for wakefulness (PVW) and mean agreement rate (Ag) on percentage of observations for which *psg* indicated sleep (base rate of sleep). Predictive values were calculated by the simple threshold method and by logistic regression of *psg* state on a 20-minute moving average of *act*. A base rate of sleep was lowest for day samples, highest for night samples, and intermediate for circadian (day + night) samples. Dotted lines show PV expected by chance for various base rates.

than wakefulness (33.5%). The PV's therefore varied with the base rates of the states being predicted: PVS was relatively high in night samples consisting mainly of sleep, and PVW was high in day samples consisting mostly of wakefulness (Figure 3). Ag's were less strongly affected by base rates of sleep and wakefulness.

The simple-threshold method generally yielded the lowest PV's, but logistic regression raised PVS them to only 62.2% in circadian samples and lowered PVS to 81.6% in night samples. Its effect on PVS in day samples could not be judged, because logistic regression predicted no daytime sleep in most subjects, making PVS meaningless. Table 2 shows the logistic regression results for both the test sample (n=10) and the full sample (n=14), which included the four subjects used to develop the logistic regression model. The greatest specific effect of logistic regression was to raise PVW in night samples from 47.1% to 89.6%, thereby greatly reducing base-rate effects. Mean Ag reached only 76.9% for circadian samples and 82.0% for night samples.

Sleep varied among the subjects, who ranged widely in age. The lowest sleep efficiency (56.7%) occurred in a 70-year-old man, and the highest sleep efficiency (88.2%) occurred in a 23-year-old man. The age effects will be presented in detail separately.

Regardless of whether the simple-threshold or logistic-regression method was used, the activity recorders accurately predicted *psg*-based circadian period and consolidation but overestimated sleep rate and sleep efficiency (Table 3). The CSA recorder measured a mean sleep rate of 0.447 (equivalent to 10.7 hours of sleep per 24-hour day) vs. a *psg* mean of 0.341 (8.2 hours of sleep/day), and it overestimated sleep efficiency by the equivalent of 1.3 hours a night.

As already noted, the *act* and *psg* data were strongly auto- and cross-correlated. Transition probabilities among the four binary *act/psg* states showed that the concordant states (quiet sleep, Ss, and active wakefulness, Ww) were persistent, whereas the discordant states (quiet wakefulness, Ws and active sleep Sw), were

Table 2—Mean predictive values (%) and agreement rates (%) for three conditions and two prediction methods

	Simple threshold n=14†	---Logistic reg--- n=10*	n=14†
Days + nights			
PVS	55.0	62.2	62.2*
PVW	92.4	96.0	96.4**
Ag	72.1	76.9	78.2
Nights			
PVS	83.1	81.6	81.8
PVW	47.1	89.6***	89.3***
Ag	77.9	82.0	82.2***
Days			
PVS	1.4	-	-
PVW	99.5	99.5	99.2
Ag	50.4	98.6***	98.6***

†includes training sample
*p<.05 vs. predictive values (PV's) based on simple threshold method
**p<.01 vs. PV's by simple threshold
***p<.001 vs. PV's by simple threshold

transitional to and from other states (Table 4). Subjects more often stopped moving before falling asleep than after falling asleep (t=2.39, df=13, p=0.033), the mean interval being 6.2±3.9 (SD) minutes. Subjects also began to move before rather than after they woke up (t=3.19, df=13, p=.007), with a mean interval of 4.1±1.9 minutes. Direct transitions between Ss and Ww were rare. It could not be shown that the transition probabilities depended only on the state preceding the present state (the first-order Markov property).

DISCUSSION

Our data provide little support for actigraphy as a means of predicting the *psg*-determined sleep-wake state. The average probability that a sleep prediction in the course of a circadian day was correct was only 62.2%. Even predictions of sleep in night samples were incorrect nearly 20% of the time. This is much worse than the PVS of 90.2% calculated from data published by Cole et al.[8] PV's derived from the data of Sadeh et al.[7] were also higher than ours. Furthermore, in this first study of the predictive ability of act in daytime samples, neither of the prediction methods tested by us detected the rare episodes of daytime sleep. *Act* also overestimated the efficiency of sleep during self-selected bedrest periods and the amount of sleep during full, circadian cycles.

Several possible reasons for the differences between our findings and previous ones need to be considered. First, the *psg* data from successive nights or days were concatenated by us. Although careful records were kept and the procedure was automated, errors may have crept in. Second, differences between the clocks of the movement recorders and polygraph may have caused loss of synchrony as the experiment progressed. Both of these potential source of prediction error were made unlikely by showing that mean PV's based on the first two days were statis-

tically similar to those of the last two. A third possibility is that the *psg* recordings may have been misscored. To minimize this, technicians referenced their scoring decisions to a uniform, automated procedure (Sensormedics Inc.). Furthermore, their scoring was checked against the scoring of another technician. Graphs similar to Figure 1, as well as the inverse relationship of activity level to depth of sleep were additional evidence that the *psg* data had been scored correctly

The fourth potential source of prediction error was the insensitivity of the recorders. When movements are too sparse or slow to reach an activity recorder's detection threshold, they are represented as zero activity, and therefore, as sleep. In continuous use around the clock, both instruments chosen for this study recorded zeroes over 50% of the time. This may be compared with 35.5 % found by us[17] using the instrument that was used in previous *act-psg* studies (MiniMotionLogger, Ambulatory Monitoring, Inc.; 7-8). During subjective nights, when most mispredictions occurred, nearly 85% of *act* measurements were zero. By failing to identify small nocturnal movements, the recorders would have correctly predicted sleep whenever subjects moved in their sleep but incorrectly predicted sleep whenever they briefly woke. The proportion of correct sleep predictions (PVS) would therefore have been low. Although most current instruments appear to be capable of detecting such small movements, much of their storage (8 bits every 0.5 minute in the present case) is allotted to high *act* levels, which chiefly occur while subjects are up and about and are of secondary interest to sleep investigators and clinicians. Sleep-wake prediction would be improved by 12-bit A/D conversion in order to increase the number of activity levels, as well as by logarithmic or other nonlinear scaling of accelerometer signals prior to A/D conversion.

In addition to being insensitive, the recorders returned *act* levels that were unimodally distributed, whereas the recorder used in previous *act-psg* studies gave bimodal distributions.[17] If bimodal distributions reflect less overlap between the high *act* levels associated with wakefulness and the low levels associated with sleep, recorders that give bimodal distributions have greater ability to discriminate between the two states.[18]

The final explanation of the low PV's is that our subjects, who had no sleep complaints, were found to have insomnia in the laboratory. They had been instructed to use the bed only for sleep, but time in bed was not restricted. As a result, they often remained in bed longer than they were able to sleep, as shown by the low sleep efficiencies. Consolidation also was often low. We suggest that this insomnia explained the low PV's, because each bout of sleep was bounded by two sleep-wake transitions, each associated with *act* levels lower than those that consistently predicted wakefulness. Depending on their sensitivities, the recorders would have misrepresented some of these low *act* levels as zeroes and predicted sleep when subjects were either still awake (at times of sleep onset) or had just started to awaken (at times of sleep offset). This reasoning may also apply to the inaccuracy of *act* that has been observed in subjects with clinical insomnia or intentionally fragmented sleep.[8,19,20]

PV's are the measures that should be of greatest interest to investigators and clinicians who contemplate substituting *act* for *psg*, but they are subject to misinterpretation of two kinds. First, sleep scoring is not a fully reliable process. There are substantial variations between and within scorers, and actigraphy cannot be expected to be more accurate than the scored *psg* itself.[21] Second,

Table 3—Mean (SD) global sleep-wake measures

	PSG n=14	Act (simple threshold) n=14	Act (logistic reg) n=10	 n=14
Circadian period (min)	1507(55)	1513(66)	1524(64)	1520(57)
Sleep rate	0.341(0.05)	0.438(0.07)***	0.447(0.11)	0.472(0.10)**
Consolidation (min)	122.6(45.8)	96.5(31.2)	112.4(51.8)	127.5(58.2)
Sleep efficiency (%)	78.5(8.1)	86.2(5.7)**	90.6(4.0)***	89.6(4.8)***

**p<.01 by paired t-test vs. *psg* data; ***p<.001 by paired t-test vs *psg* data

Table 4—Transition probabilities among joint *psg-act* states for one subject

From	Ww	Ws	Sw	Ss
WW	0.821	0.170	0.003	0.006
Ws	0.348	0.558	0.012	0.083
Sw	0.038	0.098	0.281	0.584
Ss	0.007	0.044	0.087	0.861

W=*psg* wakefulness; S=*psg* sleep; w=*act* wakefulness; s=*act* sleep

PV's vary with the base rates at which "true sleep" and "true wake" observations occur in a sample. Our analysis of subsamples with strongly differing base rates highlighted this. Virtually perfect predictions of wakefulness were obtained in circadian and daytime samples, which consisted mainly or entirely of wakefulness, but the PV's of events with lower base rates, such as PVS in circadian samples and PVW in night samples, were much lower.

Base-rate effects are exemplified by the PVS of 81.6% obtained in night samples by the logistic-regression method. Being correct 81.6% of the time might look like moderate success but was only slightly better than the performance that could be expected from a malfunctioning or unworn instrument that recorded nothing but zeroes. Unless excluded as artifact, such zeroes would have been interpreted as sleep, and the PVs would have been 77.8%, because sleep accounted for 77.8% of the night samples. Although the difference between 81.6% (actual instrument) and 77.8% (hypothetically defective instrument) is small, it was statistically significant because the number of observations was large (n=8092, chi-square=3095, df=1, p<.001). Because base rates of sleep and wakefulness are not known a priori when *act* is performed independently of *psg*, *act* findings should always be reported with the day-night intervals over which they were made. A similar recommendation was made by an expert panel of The American Sleep Disorders Association.[9]

We also found that the accuracy of *act* predictions varied with the method used to convert *act* levels to sleep-wake predictions. The threshold method has been reported to work well,[22] but in our experience, mean PVS based on empirical thresholds was below 50% in night samples. It sharply improved when based on logistic regression.

To be useful as a means of emulating *psg*, *act* should also accurately predict global sleep parameters. These include the duration of sleep (more accurately, the sleep rate), the continuity of sleep and wakefulness (consolidation), and the ratio of time asleep to time spent trying to sleep (sleep efficiency). Our results were mixed: The two recorders estimated consolidation with fair accuracy, but they overestimated sleep rate and sleep efficiency. The CSA recorder overestimated sleep rate by the equivalent of 2.5 hours of sleep per 24-hour day, and it overestimated sleep efficiency by 1.3 hours of sleep per night. A tendency for *act* to overestimate sleep has been noted previously.[8] Overestimates of sleep imply that *act* predictions of sleep as a proportion of total observations are larger than the base rate of *psg* sleep. This is algebraically equivalent to saying that *act* incorrectly predicted sleep more often than it incorrectly predicted wakefulness. The incorrect predictions must have occurred mainly at night, because little or no sleep was predicted or found in the daytime. Therefore, *act* misinterpreted nocturnal wakefulness as sleep more often than it misinterpreted nocturnal sleep as wakefulness. The erroneous predictions of sleep were presumably made when *act* levels were lower than those that are predictive of wakefulness (Figure 2). We indeed observed subjects lying quietly in bed after waking for the "day," who later told us that they were deciding whether to get up. The sleep overestimates were therefore another consequence of the effective insensitivity of the recorders to low levels of activity.

The hypothesis considered in this paper is that *psg*, the standard method of discriminating sleep from wakefulness, can be emulated by measurement of motility, which is more easily obtained. This amounts to saying that *act* data are informationally equivalent to the multi-channel *psg*, at least for the purpose of sleep-wake discrimination. (*Act* cannot identify sleep stages, despite our finding that *act* levels roughly parallel the depth of sleep). We have pointed out[21] that sleep onset and offset each represents a spectrum of changes in electrophysiological variables (such as EEG, EOG, and EMG) and behavioral variables (such as body motility and arousal responses to auditory stimuli.[23] The hypothesis of information equivalence assumes that these variables are manifestations of a common sleep-wake state generator and are therefore correlated. We found that subjects stopped moving before they fell asleep and resumed moving

before they woke, but attempts to find rules relating successive, join *psg-act* states (the Markov-1 property) did not succeed. Our results with PV, sleep efficiency and sleep rate were also inconsistent with the information-equivalence hypothesis.

An alternative hypothesis is that the electrophysiological and behavioral variables have sources of variation other than a central sleep-wake generating process. That this is the case is apparent from the increases of motor activity that may occur during EEG sleep in sleepwalkers and patients with REM behavior disorder.[24] Even the constituent *psg* variables can have multiple sources of variation: alpha EEG activity may occasionally appear during slow-wave sleep, and spindles may occur during REM sleep. Changes in single variables, such as *act*, therefore cannot be used to uniquely predict central state or other centrally governed variables, such as the *psg* variables, EEG/EOG/EMG. From this point of view, *act-psg* discrepancies may be of biological origin and are not always technical problems.

Univariate markers of sleep-wake state other than *act* have been proposed,[23] but no observable variable can be said to have overriding theoretical validity. A multivariate definition of sleep and wakefulness therefore remains necessary. It was indeed by observing new correlations among EEG, EOG, and EMG that REM sleep was discovered,[25] and these variables have been used worldwide for decades to identify sleep and wakefulness.[1] These qualifications make the *psg* an apt standard by which to measure the predictive ability of *act*.

None of this detracts from the fact that activity levels are correlated with the states of sleep and wakefulness. Even the simple -threshold method for converting *act* data to sleep-wake predictions accurately defines the circadian sleep-wake cycle, as well as the rest/activity cycle, and *act* levels are highly correlated with sleep-wake state. It is rather the transitions between sleep and wakefulness that *act* often fails to detect. Thus, although *act* level correlates with subjective sleep disturbance,[11] and *act* predictions of wakefulness during the night were accurate nearly 90% of the time when based on logistic regression, no evidence has yet been presented that actigraphy can identify the times or durations of individual nocturnal arousals.

We conclude that the predictive power of actigraphy remains limited by the sensitivity of current recorders and, more significantly, by the likelihood that motility carries sleep-wake information that differs from that of the conventional *psg*. We have previously found substantial differences among activity recorders[17] and now between PV's based on several sample types and numerical prediction methods. Clearly, the early hope that actigraphy would provide a robust and convenient method of identifying sleep and wakefulness has not been realized. Our reservations about actigraphy as an accurate predictor of sleep and wakefulness do not preclude future advances that may make this possible. They certainly do not detract from its validity as a means of quantifying rest and activity levels and circadian rest-activity cycles. But it does not seem appropriate in 2001 to refer to inactivity defined by wrist actigraphy as "sleep" or to wrist activity as "wakefulness."

REFERENCES

1. Rechtschaffen A, Kales A. In: A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington, DC: United States Public Health Service, U.S. Government Printing Office, 1968.
2. Tryon WW. Activity measurement in psychology and medicine. New York: Plenum Press, 1991.
3. Aschoff J. Aktivitatsmuster der tagesperiodik. Naturwissenshaften 1957; 44:361-367.
4. DeCoursey PJ. Phase control in a rodent. Cold Spring Harbor Symp Quant Biol 1960;25:49-55.
5. Pittendrigh CS. Circadian rhythms and the circadian organization of living systems. Cold Spring Harbor Symp Quant Biol 1961;25:159-184.
6. Richter CP. A behavioristic study of the activity of the rat. In: Hunter WS, ed. Comparative psychology monographs. Baltimore, MD: Williams & Wilkins, 1922;1:1-55.
7. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep 1994; 17:201-207.
8. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992;15:461-469.
9. Sadeh A, Hauri PJ, Kripke DF, Lavie P. The role of actigraphy in the evaluation of sleep disorders. Sleep 1995;18:288-302.
10. Pollak CP, Stokes P. Circadian rest-activity rhythms in demented and nondemented elderly community residents and their caregivers. J Am Geriatrics Soc 1997;45:446-452.
11. Pollak CP, Stokes PE, Wagner DR. Nocturnal interactions between community elders and caregivers, as measured by cross-correlation of their motor activity. J Geriatic Psychiatry & Neurol 1997;10:168-173.
12. Tryon WW, Williams R. Fully proportional actigraphy: a new instrument. Behav Res Meth, Inst, and Comp 1996;28:392-403.
13. Chambers JM. Programming with data. A guide to the S language. New York: Springer-Verlag, 1998.
14. Winer BJ, Brown DR, Michels KM. Statistical principles in experimental design. 3rd ed., New York: McGraw-Hill, 1991.
15. Sokolove PG, Bushell WN. The chi square periodogram: its utility for analysis of circadian rhythms. J Theor Biol 1978;72:131-160.
16. Basawa IV, Rao BLSP. Statistical inference for stochastic processes. New York: Academic Press, 1980:63-64.
17. Pollak CP, Stokes PE, Wagner DR. Direct comparison of two widely used activity recorders. Sleep 1998;21:207-212.
18. Polmaecher T, Schulz H. The relation between wrist-actigraphic measures and sleep stages (abstract). Sleep Res 1987;16:55.
19. Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. Sleep 1992;15:293-301.
20. Stampi C, Broughton R. Ultrashort sleep-wake schedule: detection of sleep state through wrist-actigraph measures (abstract). Sleep Res 1988;17:100.
21. Tryon WW. Nocturnal activity and sleep assessment. Clin Psychol Rev 1996;16:197-213.
22. Monk TH, Buysse DJ, Rose LR. Wrist actigraphic measures of sleep in space. Sleep 1999;22:948-954.
23. Ogilvie RD, Wilkinson RT Behavioral versus EEG-based monitoring of all-night sleep/wake patterns. Sleep 1988;193-155.
24. Mahowald MW, Schenck CH. Dissociated states of wakefulness and sleep. Neurol 1992;42:44-51.
25. Aserinsky E, Kleitman N. Regularly occurring periods of eye motility and concomitant phenomena during sleep. Science 1953;118:273-274.