

Actigraphically based automatic bedtime sleep-wake scoring: Validity and clinical applications

A. SADEH, J. ALSTER, D. URBACH and P. LAVIE†

Sleep Laboratory, Unit of Behavioral Biology, Faculty of Medicine, Technion City, Haifa 32000, Israel

The validity of actigraphic data for bedtime sleep-wake discrimination was evaluated against conventional polysomnographic recordings. The discriminant function was developed on the basis of actigraphic data from nine normal subjects (91.8% agreement), and was validated on four normals (86.2%), 25 patients with sleep apnea syndrome (SAS) (85.7%), 16 insomniacs (78.2%) and 13 children (89.9%). The ability of this scoring procedure to discriminate between distinct clinical groups (normals, insomniacs and SAS) was tested and then validated with correct assignments of 73.5% and 64.6% (for calibration and validation samples, respectively) both significantly higher than chance. In addition, actigraphic data alone could significantly estimate the number of apneas in the SAS patients. These findings demonstrate the clinical and experimental value of ambulatory monitoring of motility in sleep.

Introduction

Since Szymansky [1] first introduced motility measurements in human sleep several methods have been introduced to measure body movements during sleep. These include photographic monitoring [2, 3], electroencephalographic recordings [4] and body or bed transducers [5-8]. Recently the use of wrist-worn actigraphs has gained considerable popularity. The actigraph enables long continuous monitoring of wrist movements with minimum disturbance or distortion of the natural circumstances in sleep and wake states [10].

The growing use of actigraphic recording of sleep-wake cycles for clinical and research purposes requires special effort to examine its validity and applicability to the basic distinction between wake and sleep.

Mullaney *et al.* [7] have demonstrated that actigraphic data provide a reliable and valid differentiation between sleep and wake. They reported a 94.5% agreement between actigraphic data and polysomnographic scoring. This is only slightly lower than the inter-rater reliability of polysomnographic scoring.

Webster *et al.* [11] took the next step of developing an automatic scoring system. Their results supported the validity of this system in differentiating between minutes of sleep and wake. Webster *et al.* reached a 93.04% agreement with EEG scoring. It should be noted, however, that since they validated their scoring system over 24 hours of recording per individual, this could inflate the 'hit' rates because of the relatively easily scored wake minutes during the day.

The first purpose of the study reported here was to examine the validity of sleep-wake automatic scoring of actigraphic data against traditionally scored polysomnographic data [12]. The specific aim was to develop an automatic scoring

† To whom correspondence should be addressed.

system that would combine two basic features: (a) capability to handle large data banks (data scored in 1 minute epochs for more than 1 week); and (b) a scoring system that would ensure a high proportion of correct classifications and accurate estimation of sleep efficiency.

When the validity of such a procedure was reasonably established, the next step was to test whether this scoring system could be utilized for an inexpensive, preliminary clinical screening of sleep disorders and for assessment of their severity.

Method

Sixty-seven subjects (13 normal and 54 patients including a subgroup of 13 children) participated in the first stage of the study. Children's ages ranged between 3 and 13 years (mean 9.1) and adults' ages ranged between 20 and 76 years (mean 43). All subjects were recorded in the sleep laboratory, for one night, in the conventional way, in addition to a wrist actigraph (Ambulatory Monitoring, Ardsley, NY, USA) worn on the non-dominant hand. The activity data, consisting of the number of zero crossings were registered in each 1 minute epoch. We used this epoch because it is commonly used in our laboratory for research purposes. Sleep stages were scored according to Rechtschaffen and Kales [12]. 'Sleep' consisted of stages 1,2,3-4 and REM. 'Wake' consisted of stages 0 and movement time. Each minute was scored according to its dominant stage. The actigraphic data were then subjected to discriminant procedures to test their ability to differentiate between minutes of sleep and wake.

Activity data from the first nine normal subjects taken from 'lights off' to 'lights on' were employed to identify the optimal discriminant function. All statistics were computed using the SAS statistic programs [13]. A stepwise discriminant procedure [14] was used to identify the five most efficient variables for the discriminant function. The input parameters for this procedure included activity level, minimum value and standard deviation during all combinations of up to 10 minutes before and after the given minute. After the best five variables were identified, they were subjected to the discriminant procedure [15] which provided the optimum discriminant function. The discriminant function was validated with different normal and clinical samples.

In the second stage of the study, which aimed at testing the possibility of differentiating between clinical groups via actigraphic monitoring, the children's group was excluded, since it was not clinically homogeneous. An additional group of 28 adults (10 normals, seven insomniacs and 11 sleep apnea syndrome; mean age 39.8) was included only in this part of the study. Thus, a total of 82 adults, 54 from the first stage of the study and 28 new subjects, was included in this analysis. The 82 subjects were divided, based on diagnostic interviews and polysomnographic recordings, into normals, insomniacs and sleep apnea syndrome groups. Fourteen variables were submitted to the stepwise discriminant analysis. These variables, which were assumed to represent sleep quality (activity level, sleep fragmentation, etc.), were computed based on the scoring procedures described in stage 1. Discriminant analysis was performed on variables computed from the actigraphic data on a randomly selected group of 34 subjects, and the validity of this function was tested on a validation sample of 48 subjects. Finally, in the sleep apnea syndrome group, the number of apneic episodes was submitted to stepwise regression analysis to assess the ability of the same actigraphically derived parameters to predict the severity of the sleep apnea syndrome.

Results

Stage 1

The stepwise discriminant analysis yielded the order of the best predictors for sleep-wake differentiation and their unique predictive value. The best five predictors were selected. The contribution of the next predictors, which were not included in the discrimination function, was lower than 1% (partial R square lower than 0.001 and lower than 1% contribution to the explained variance).

The discriminant procedures yielded the following sleep-wake discriminant function:

$$PS = 4.532 - (0.06828 * XO - 0.0385 * S_5 - 0.038 * S9 + 0.0298 * M2 - 0.0299 * S_2$$

If $PS \geq 0$ then stage=sleep, else stage=wake.

Where XO =number of zero crossings (NZC) of the classified minute; S_5 =standard deviation of NZC of the prior 5 min; $S9$ =standard deviation of NZC of the following 9 min; $M2$ =minimum value of NZC of the following 2 min; and S_2 =standard deviation of NZC of the prior 2 min.

The validity of this function was tested with the data of the four other normal subjects and 54 patients (subdivided into three categories: children, sleep apnea syndrome and insomnia). Tables 1 and 2 summarize the results of this sleep-wake classification and sleep efficiency indices, respectively, computed from this classification.

Table 1. Sleep-wake correct classifications. Number of correct classifications and percentages from the actual classifications.

Group	N	Sleep	Wake	Total
Normals calibration	9	2692 95.49%	237 63.54%	2929 91.76%
Normals validation	4	969 88.25%	176 76.19%	1145 86.16%
Patients children	13	4430 92.91%	402 66.01%	4832 89.86%
Patients insomnia	16	3320 95.43%	975 48.48%	4295 78.23%
Patients apnea	25	7486 92.06%	995 56.47%	8481 85.72%

Table 2. Sleep efficiency predictions. Polysomnographic and actigraphic mean sleep efficiency, Pearson correlations and significance level.

Group	N	EEG mean	Actigraph mean	Corr	P<
Normals	13	87.07	85.65	0.905	0.0001
Children	13	88.85	86.48	0.813	0.001
Insomnia	16	63.30	78.56	0.785	0.0005
Apnea	25	82.15	83.52	0.630	0.001

The total correct classifications were within the range 78.2–91.8% for the various groups. The 'hit' rate did not drop significantly from the calibration sample to the validation samples. The largest drop in the insomniacs group resulted from the relatively many false classifications of wake minutes as sleep minutes.

The correlations between sleep efficiencies calculated on the basis of the automatic classification, and those based on polysomnography, were 0.905 for normals, 0.813 for children, 0.735 for insomniacs and 0.630 for SAS (all highly significant at $P < 0.001$).

A visual inspection of individual scoring of specific subjects (figures 1, 2 and 3) reveals the focal areas of mismatch between polysomnographic and actigraphic scoring. From these examples it can be clearly seen that for the normal subject the two scoring procedures match almost perfectly. For the sleep apnea syndrome patient, however, who had 400 apneas, particularly during the last third of his sleep, there was obvious disagreement between the scorings. This mainly resulted from the fact that the polysomnographic scoring was insensitive to the multiple micro-arousals which accompanied the periodic terminations of the apneas. In the insomniac patient there was a very good agreement between the two scoring procedures, demonstrating the validity of the actigraphic recording to identify prolonged wake periods within sleep.

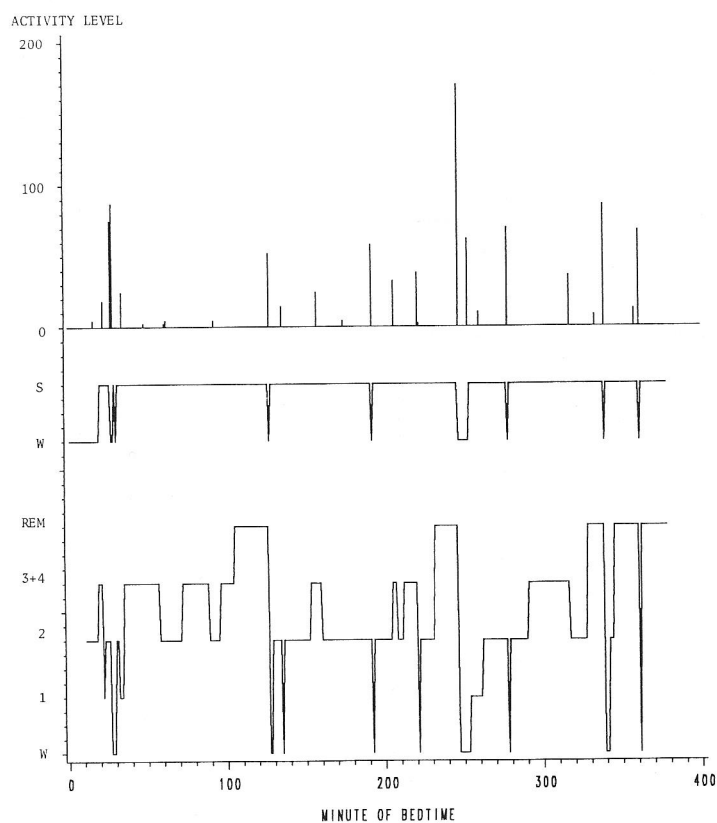


Figure 1. Actigraphic and polysomnographic scoring (normal).

Post-hoc analysis of the unique contribution of each parameter included in the discriminant function to the correct assignment probability revealed that the last variable entered (S_2) was relatively insignificant. The removal of any other parameter would lead to a significant drop of at least 1% of the correct assignment probability.

Stage 2

In the second stage of statistical analysis we tested the ability of our automatic scoring procedure to differentiate between the normal and two patient groups. Fourteen variables were submitted to the stepwise discriminant analysis. The selection of parameters included in the discriminant function followed the procedure in stage 1. The actigraph's parameters included in the analysis were total bed time (TBT), proportion of the number of transitions between sleep-wake minutes from TBT (TRS), number of successive minutes within each state (number of minutes that follow 10 min of the same state, STAB), standard deviations of activity levels in sleep state (SDS), number of transitions from one successive 10 min of sleep-wake state to another (STABR) and movement density (number of movements per minute, MD). Discriminant analysis performed on the data of the calibration sample of 34 subjects yielded the following discriminant functions:

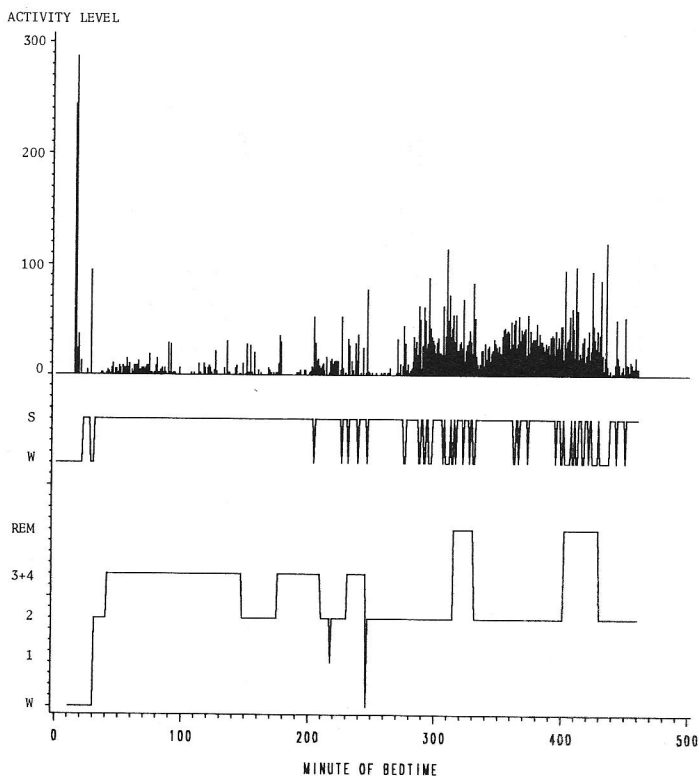


Figure 2. Actigraphic and polysomnographic scoring (sleep apnea).

$$PN = -87.90 - 0.310 * TBT + 6.39 * TRS + 25.05 * MD \\ + 1.53 * STABR + 0.808 * STAB + 0.507 * SDS$$

$$PI = -92.15 - 0.339 * TBT + 6.99 * TRS + 23.64 * MD \\ + 1.43 * STABR + 0.842 * STAB + 0.628 * SDS$$

$$PA = -98.52 - 0.273 * TBT + 6.29 * TRS + 29.19 * MD \\ + 1.34 * STABR + 0.799 * STAB + 0.557 * SDS$$

These are the three indices of the probability of belonging to each of the groups (normals, insomniacs and sleep apnea syndrome, respectively). The subject was assigned to the group of the largest index. The validity of this procedure was tested with the validation sample of 48 subjects (table 3). The 'hit' rate for the calibration sample was 73.53%, and for the validation sample, using the same discriminant functions, 64.6%. Both are significantly higher than random assignment probability (33.3%). When tested separately, the two validation groups (subjects who participated in the first stage and subjects included only in this stage) revealed a similar pattern of results. Furthermore, *post-hoc* analysis of the misclassified cases revealed that their diagnosis based on polysomnographic data itself was marginal.

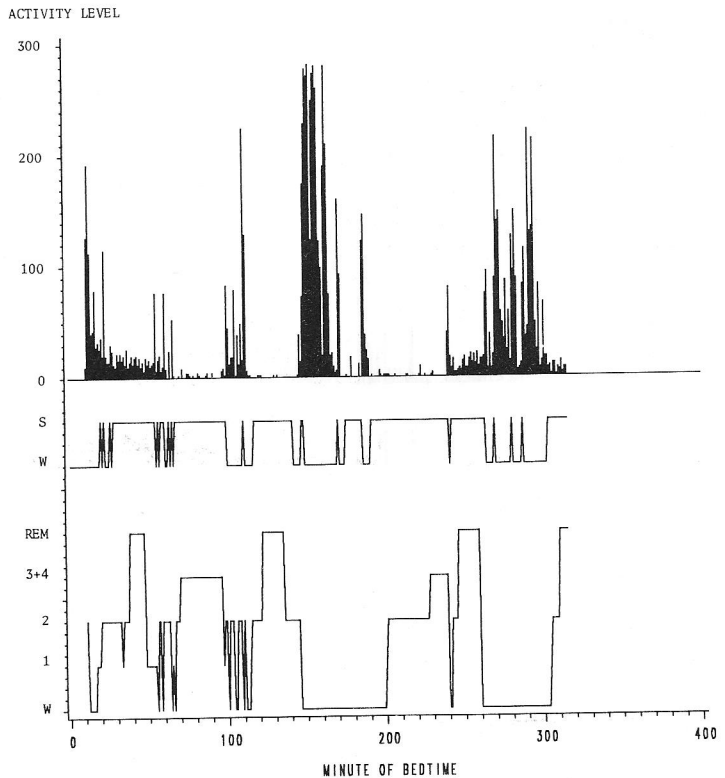


Figure 3. Actigraphic and polysomnographic scoring (insomnia).

Table 3. Clinical classification by actigraphic data (N=82).

From\ To Group\ Group	Calibration sample (N=34)			Validation sample (N=48)		
	Normals	Insomnia	Apnea	Normals	Insomnia	Apnea
Normals (N)%	6† 75%	1 12.5%	1 12.5%	9† 60%	3 20%	3 20%
Insomnia (N)%	2 20%	8† 80%	0 0%	2 15.4%	7† 53.8%	4 30.8%
Apnea (N)%	3 18.75%	2 12.5%	11† 68.75%	2 10%	3 15%	15† 75%

† Correct classifications.

Finally, we tested whether the actigraph could yield information on the severity of the sleep disorder in the sleep apnea subgroup. Stepwise regression analysis was employed to test which of the actigraphic parameters could be used to estimate the number of apneas detected via polysomnography. Two variables were found to be significant predictors: movement density and number of successive sleep minutes, which together explained 30.04% of the variability in the number of apneas ($p < 0.005$). The best single predictor was movement density which was significantly correlated with number of apneas, $r = 0.448$ ($p < 0.01$).

Discussion

Our results demonstrate an acceptable validity of the automatic scoring procedures for actigraphically obtained data in discriminating between sleep and wake states. A similar approach and results were demonstrated with relatively small samples by Cole and Kripke [16]. The accuracy of this procedure tended to fall within certain clinical groups (e.g., apnea) suggesting that prior identification of the clinical category membership, via actigraphic data or clinical interview, may contribute to developing special automatic scoring systems for specific clinical populations.

The clinical value of the actigraphic data and the automatic scoring system was further demonstrated in this pilot study by the significantly higher than chance probabilities of classifying subjects into three subgroups: normals, insomniacs and sleep apnea patients. The stability of the validation results of different sample groups lends additional support to the robustness of the discriminant function. Furthermore, the results show that apnea's severity could be significantly estimated by the actigraphic data alone.

Although the probabilities of misclassification were relatively high and further improvement is needed for meeting higher clinical standards, we believe that the advantages of actigraphic recordings for preliminary large scale screening studies and for clinical field studies outweigh this limitation. Such applications were successfully demonstrated in clinical settings with hypnotic drugs [5] and sleep apnea syndrome [17].

The present findings support the cost-effective value of the actigraph as a major tool in the clinical and experimental research of sleep-related issues. Sleep could be reliably differentiated from wake in subjects and patients lying in bed. Valuable clinical data could be extracted from activity data alone.

References

1. SZYMANSKY, J. S. (1922) Aktivitaet und ruhe bei denmenschen, *Zeitschrift für angewandte Psychologie*, **20**, 192-222.
2. HOBSON, J. A., SPAGNA, T. and MALENKA, R. (1978) Ethology of sleep studied with time-lapse photography: Postural immobility and sleep-cycle phase in humans. *Science*, **201**, 1251-1253.
3. LAVIE, P., ZOMER, J., PEISER, J., OVNAT, A. and CHARUZZI, I. (1982) The effects of obesity surgery on morbidly obese sleep apneic patients: A cinematographic and hypnopolygraphic investigation. In W. P. Koella (Ed) *Sleep 1982*, (Karger: Basle), pp. 343-345.
4. MUZET, A., BECHT, J., JACQUOT, P. and KOENIG, P. (1972) A technique for recording human body posture during sleep. *Psychophysiology*, **9**, 660-662.
5. BORBELEY, A. A., LOEPFE, M., MATTMAN, P. and TOBLER, I. (1983) Midazolam and triazolam: Hypnotic action and residual effects after a single bedtime dose. *Arzneimittel Forschung-Drug Research*, **33**, 1500-1502.
6. KAYED, K., HESLA, P. E. and ROSJO, O. (1979) The acti-culographic monitor of sleep. *Sleep*, **2**, 253-260.
7. MULLANEY, D. J., KRIPKE, D. F. and MESSIN, S. (1980) Wrist actigraphic estimation of sleep time. *Sleep*, **3**, 83-92.
8. HELFAND, R., LAVIE, P. and HOBSON, A. J. (1986) REM/NONREM discrimination via ocular and limb movement monitoring: Correlation with polygraphic data and development of a REM state algorithm. *Psychophysiology*, **23**, 334-339.
9. KRIPKE, D. F., MULLANEY, D. J., MESSIN, S. and WYBORNEY, V. G. (1978) Wrist actigraph measures of sleep and rhythms. *Electroencephalography and Clinical Neurophysiology*, **44**, 674-678.
10. REDMOND, D. P. and HEGGE, F. W. (1985) Observations on the design and specification of a wrist-worn human activity monitoring system. *Behaviour Research Methods, Instruments and Computers*, **17**, 659-669.
11. WEBSTER, J. B., KRIPKE, D. F., MESSIN, S., MULLANEY, D. J. and WYBORNEY, G. (1982) An activity-based sleep monitor system for ambulatory use. *Sleep*, **5**, 389-399.
12. RECHTSCHAFFEN, A. and KALES, A. (1968) In R. J. Berger et al. (Eds), *A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects* (Public Health Service, U.S. Government Printing Office, Washington, DC).
13. SAS Institute Inc. (1985) *SAS User Guide: Statistics, Version 5*, Ed. (SAS Institute Inc: Cary, NC).
14. KLECKA, W. R. (1980) Discriminant analysis. In *Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-019*, (Sage Publications, Beverley Hills).
15. RAO, C. R. (1973) *Linear Statistical Inference and Its Applications*, (New York, Wiley).
16. COLE, R. J. and KRIPKE, D. F. (1988) Progress in automatic sleep/wake scoring by wrist actigraph. *Sleep Research*, **17**, 331.
17. AUBERT-TULKENS, G., CULEE, K., HARMANT-VAN RIJCKEVORSEL, K. and RODENSTEIN, D. O. (1987) Ambulatory evaluation of sleep disturbances and therapeutic effects in sleep apnea syndrome by wrist activity monitoring. *American Review of Respiratory Diseases*, **136**, 851-856.