

## Further Validation of Actigraphy for Sleep Studies

Luciane de Souza MSc, Ana Amélia Benedito-Silva PhD, Maria Laura Nogueira Pires PhD, Dalva Poyares MD, PhD, Sergio Tufik MD, PhD, Helena Maria Calil MD, PhD

Department of Psychobiology, Federal University of São Paulo, São Paulo School of Medicine, São Paulo, Brazil

**Study Objectives:** Actigraphy is generally compared to polysomnography (PSG), which has been considered the gold standard for sleep studies. The objective of the present study was to evaluate the concordance between PSG and two previously proposed algorithms (Cole et al, 1992; Sadeh et al, 1994) to analyze actigraphic recordings. The minute-by-minute agreement rate was evaluated through calculation of sensitivity, specificity, and accuracy. Regarding the sleep parameters, the concordance was performed through the Bland and Altman technique.

**Design:** A night of adaptation to the sleep laboratory followed by simultaneous polysomnographic and actigraphic recordings throughout the night.

**Participants:** 21 healthy volunteers.

**Setting:** A sleep laboratory

**Interventions:** None

**Results:** Ninety-one percent of all PSG epochs were correctly identified by both algorithms, and this accuracy is reasonably satisfactory. The actigraphy was a sensitive method, with values of 99% and 97% for Cole's

and Sadeh's algorithms, respectively. However, actigraphy had a low specificity: 34% and 44% for Cole's and Sadeh's algorithms, respectively. The Bland and Altman technique showed that actigraphy systematically overestimated Sleep Latency, Total Sleep Time and Sleep Efficiency while it underestimated Intermittent Awakenings.

**Conclusions:** The results of this study show the utility of actigraphy as a useful method for assessment of sleep, despite its limitations regarding identification of waking epochs during sleep. The Bland and Altman concordance technique was revealed to be a powerful tool to evaluate how well actigraphy agreed with polysomnography. This technique, combined with calculations of sensitivity and specificity, appears to be the most adequate procedure for the assessment of concordance.

**Keywords:** actigraphy, polysomnography, sleep, validation

**Citation:** de Souza L, Benedito-Silva A, Pires M, et al. Further validation of actigraphy for sleep studies. *SLEEP* 2003;1:81-5.

### INTRODUCTION

SLEEP EVALUATION IN HUMANS HAS BEEN USUALLY PERFORMED WITH POLYSOMNOGRAPHY (PSG), A TECHNIQUE CONSIDERED THE GOLD STANDARD FOR SLEEP STUDIES. In fact, PSG has been used in clinical trials in spite of its limitations for longitudinal and more naturalistic studies. Actigraphs, instruments to measure wrist motor activity and also called activity monitors, provide an indirect assessment of sleep through the use of algorithms. These are automatic scoring methods specially developed to distinguish sleep from wakefulness. The use of actigraphy is considerably increasing in clinical studies due to its low cost, and possibilities opened for long-term studies without interfering with the volunteer's or patient's routine.

Studies employing a variety of algorithms have shown that wrist actigraphy correlates well with PSG recordings, and the agreement rates have been based on an epoch-by-epoch analysis. The algorithm developed by Cole et al<sup>1</sup> distinguishes sleep from wakefulness in samples of healthy volunteers and patients with sleep and psychiatric disorders, yielding an 88% agreement rate with PSG. The Sadeh et al<sup>2</sup> algorithm also gives high agreement rates with PSG ( $\pm 90\%$ ) in adults as well as in adolescents. Jean Louis et al<sup>3</sup> developed another algorithm, reporting a slightly higher agreement rate (97%) with PSG in healthy volunteers, associated with significant correlation coefficients for some sleep parameters, including a sample of patients with insomnia.<sup>4</sup> Despite the fact that agreement rates (epoch-by-epoch) and correlation coefficients

(sleep parameters) have been used to compare PSG and actigraphy in many studies,<sup>4-11</sup> the use of these techniques as a measure of concordance has not been considered fully appropriate.<sup>12,13</sup> First, it has not been taken into account that the testing periods, which are usually conducted during the night in healthy volunteers, are almost exclusively composed by sleep. Consequently, the probability of a good concordance is high. In fact, Sadeh et al<sup>6</sup> have shown that the concordance could be as high as 92% by just scoring all epochs as sleep.<sup>12</sup> Much more information would be achieved if the principles of sensitivity, specificity, and accuracy were applied.<sup>14</sup> Sensitivity would reflect the ability of the actigraphy to detect sleep when the PSG has also scored sleep; specificity, the ability of the actigraphy to detect wakefulness when the PSG did the same; and accuracy, the ability of the actigraphy to detect both sleep and wakefulness compared to PSG. In addition, the use of correlation coefficient has been criticized because it does not measure the concordance between variables but the strength of their relationship.<sup>13</sup> In other words, the correlation coefficient measures the degree of association between two quantities. Since it does not inform how closely they agree, a high correlation does not mean a good concordance.<sup>13</sup> More specifically, a perfect agreement occurs if the points lie along the line of equality in a plot of one variable against the other, but a perfect correlation can also be achieved if the points lie along any straight line, providing the slope is not zero or infinite.<sup>15</sup> In addition, correlation coefficient depends on the range of the true quantity in the sample: if this is wide, the correlation will be greater than if it is narrow.<sup>13</sup>

Bland and Altman have suggested a procedure to estimate concordance in a sample based on the plot of the difference against the mean of two measures from each subject.<sup>13,15</sup> Briefly, the mean difference between actigraphy and PSG measurements for all subjects would reflect the systematic difference between them, and the standard deviation of the mean difference indicates the random fluctuations around this mean.

The purpose of this study was to evaluate the concordance between PSG and two actigraphic algorithms (Cole et al, 1992; Sadeh et al, 1994). The actigraphic sensitivity, specificity, and accuracy were calculated from the algorithm, and the concordance among the sleep parameters was estimated with the Bland and Altman technique.

### Disclosure Statement

This study was supported by CAPES, CNPq and Fund of Incentive to Psychopharmacology Association (AFIP), Brazil.

Submitted for publication June 2001

Accepted for publication September 2002

Address correspondence to: Luciane de Souza, Departamento de Psicobiologia, UNIFESP, Rua Botucatu 862, 1º andar, 04023-062, São Paulo, SP, Brasil. Ph: 5511-55390155; FAX: 5511-55725092; e-mail: ludesou@psicobio.epm.br

## METHODS

### Subjects

Twenty-one healthy volunteers (14 women and 7 men) participated in this study, which began after the approval of the Institutional Ethics Committee and the volunteers' signature of a written consent form. Volunteers were submitted to a Structured Clinical Interview for diagnosis by the Diagnostic and Statistic Manual (DSM-IV)/Axis Disorder 1 - Non-Patient Edition<sup>16</sup> for exclusion of psychiatric pathologies. Absence of sleep disorders was verified with a diagnostic PSG, which also was performed to familiarize subjects with the laboratory setting, followed by the application of the Pittsburgh Scale for Evaluation of Sleep Quality, with its global score  $\leq 5$  being considered an inclusion criterion to the study.<sup>17</sup>

### Polysomnography

Recordings were carried out by a trained sleep technician using a sleep analyzer computer (SAC, version 9.2, Oxford Instruments, Inc., Oxon, UK). Electroencephalogram electrodes were placed according to the International 10/20 System. Thirty-second epochs were classified automatically, according to the criteria established by Rechtschaffen and Kales,<sup>18</sup> and visually inspected by a neurologist. The following parameters were analyzed: Sleep Latency (SL, in min), defined as the time from lights out until the onset of 3 consecutive minutes of Stage 1 or one epoch of any other stage; Total Sleep Time (TST, in min), defined as the actual time spent asleep; Intermittent Awakenings (IA, in min), defined as the total awakening time after sleep onset; Sleep Efficiency (SE, %), defined as the percentage of time between the sleep onset and final awakening, which was spent asleep.

### Actigraphy

The actigraph used was the Mini Motionlogger Actigraph - Basic 32 C (Ambulatory Monitoring, Inc., Ardsley, USA). Data were collected using the "zero crossing mode," in one-minute epochs. The following sleep parameters were analyzed: TST, IA, SE (same definitions as described for the PSG) and SL (in min), defined as the time elapsed from pushing the event button and the occurrence of five consecutive sleep epochs.

### Procedure

Volunteers came to the Sleep Laboratory on two consecutive nights, at least one hour before their usual bedtime. They were instructed to abstain from drinking tea or coffee for 48 h and alcohol for 72 h before their admission. The time alignment between PSG and actigraphy recordings on the second night consisted of pressing simultaneously the actigraphic event marker and the acquisition key of the sleep analyzer computer. The actigraph was worn in the nondominant arm.

### Data Analysis

Computerized analyses were performed according to the algorithms proposed by Cole et al<sup>1</sup> (Action 3 - Version 3.15, Ambulatory Monitoring, Inc., Ardsley, USA) and by Sadeh et al<sup>2</sup> (Action for Windows - Version 1.05, Ambulatory Monitoring, Inc., Ardsley, USA) for the estimation of actigraphic sleep parameters.

The algorithm of Cole et al<sup>1</sup> computes a weighted sum of the activity in the current minute, the preceding 4 minutes, and the following two minutes as follows:  $S = 0.0033(1.06a_n + 0.54a_{n-1} + 0.58a_{n-2} + 0.76a_{n-3} + 2.3a_{n-4} + 0.74a_0 + 0.67a_1)$ ; where  $a_n$ - $a_1$  are activity counts from the prior 4 minutes,  $a_0$  is the current minute, and  $a_1$  and  $a_2$  are the following two minutes. The current minute is scored as sleep when  $S < 1$ .

The algorithm of Sadeh et al<sup>2</sup> is computed as follows:  $PS = 7.601 - 0.065MW5 - 1.08NAT - 0.056SD6 - 0.073\ln(ACT)$ ; where PS is the

probability of sleep, MW5 is the average number of activity counts during the scored epoch and a window of five epochs preceding and following it; NAT is the number of epochs with activity level equal to or higher than 50 but lower than 100 activity counts in a window of 11 minutes, including the scored epoch and the five epochs preceding and following it; SD6 - is the standard deviation of the activity counts during the scored epoch and the five epochs preceding it;  $\ln(ACT)$  is the natural logarithm of the number of activity counts during the scored epoch + 1; If PS is zero or greater, the specific epoch is scored as sleep; otherwise, it is scored as wake.

The 30s epochs from the polysomnographic recordings were pooled in one-minute intervals and classified as wake or sleep, considering as wake two consecutive and discordant epochs.<sup>2</sup>

Sensitivity, specificity, and accuracy were calculated (epoch by epoch) for each individual recording, followed by obtaining their means. Sensitivity represents the proportion of all epochs classified as sleep in the PSG that were also identified by actigraphy. Specificity is the proportion of all epochs classified as wake by PSG that were also identified by actigraphy. Accuracy is the proportion of all the epochs correctly identified by the actigraphy. Sleep parameters obtained by actigraphy and PSG were analyzed by the Pearson's correlation coefficient and by the Bland and Altman technique.<sup>13,15</sup> Their technique considers A and B as measurements provided by actigraphy and PSG, respectively. For each subject, the average  $[(A + B)/2]$ , the difference  $(A - B)$ , the mean difference, and the standard deviation of the differences are calculated. The mean difference would be the estimated bias, that is, the systematic difference between methods: null mean differences reflect a perfect concordance, positive mean differences indicate an overestimation of the actigraphic measurement, and negative mean differences, an underestimation. The standard deviation measures random fluctuations around this mean, and the 95% limits of agreement (mean difference plus or minus 1.96 standard deviations) provide an estimate of how far apart measurements obtained by the two methods were likely to be for most individuals.

## RESULTS

Twenty-one healthy volunteers, 14 women and 7 men, aged between 18 and 33 years, completed the study. They all had a normal polysomnographic night sleep pattern. Table 1 shows the mean values and standard deviations obtained with polysomnographic and actigraphic recordings for SL, TST, IA and SE.

Actigraphy was a sensitive method, i.e., a high proportion of all epochs classified as sleep by PSG was also identified by actigraphy, yielding sensitivity values of 99% and 97% for Cole's and Sadeh's algorithms, respectively). However, actigraphy had a low specificity, i.e., a small proportion of all epochs classified as wake were so identified by actigraphy. Its specificity was 34% and 44% for Cole's and Sadeh's algorithms, respectively). Ninety-one percent of all PSG epochs were correctly identified by both algorithms, and this accuracy might be considered satisfactory.

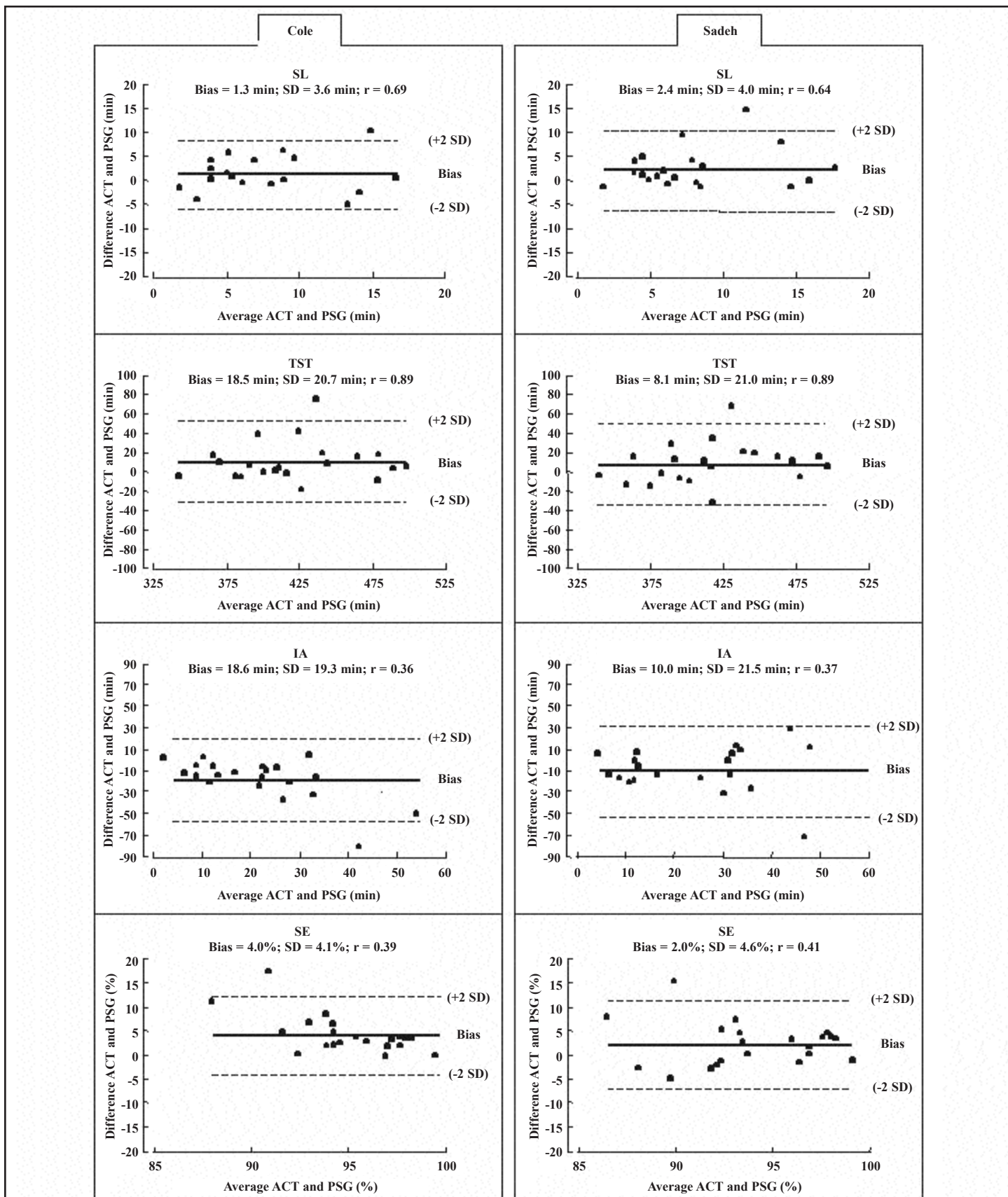
Figure 1 shows for each sleep parameter the plot of the difference between measurements by actigraphy and polysomnography against the average of the two methods from each subject. There was a lack of agreement between actigraphy and PSG for TST: the biases were 19 min and 8.0 min when estimated by Cole's and Sadeh's algorithms, respectively. With the Cole's algorithm, for 95% of the sample, the sleep duration was 23 minutes lower or 60 minutes higher than that measured by

**Table 1**—Sleep parameters (mean  $\pm$  sd) recorded by polysomnography and estimated by Cole's and Sadeh's algorithms applied to actigraphy recordings.

	Algorithms		
	PSG	Cole's	Sadeh's
Sleep latency (min)	6.9 $\pm$ 4.5	8.3 $\pm$ 4.7	9.4 $\pm$ 5.1
Total Sleep Time (min)	414.8 $\pm$ 43.2	433.3 $\pm$ 45.2	422.9 $\pm$ 47.9
Intermittent awakenings (min)	31.2 $\pm$ 20.6	12.6 $\pm$ 9.6	21.2 $\pm$ 17.7
Sleep Efficiency (%)	92.9 $\pm$ 4.4	97.1 $\pm$ 2.2	95.1 $\pm$ 4.1

PSG. With the Sadeh algorithm, these values were 34 and 50 minutes respectively. Despite these large differences, the correlation coefficients ( $r = 0.89$ ) were high for both algorithms.

The discrepancies between methods were not so high for the estimation of SL. The mean differences were 1.3 and 2.4 minutes when Cole's and Sadeh's algorithms were applied. The use of Cole's algorithm result-



**Figure 1**—Difference against average of PSG and ACT (Cole to the left, Sadeh to the right) sleep measurements, with 95% limits of agreement (broken lines). Differences are shown in the y-axis and averages in the x-axis. Mean bias, standard-deviation and correlation coefficients are shown in the right corner. SL - Sleep Latency (min); TST - Total Sleep Time (min); IA - Intermittent Awakening (min); SE - Sleep Efficiency (%). Abscissa data are expressed in minutes, except for SE, which is in percentage.

ed in an estimate ranging, for 95% of the sample, from 6 min lower to 8.5 min higher than the PSG measurement. With the Sadeh's algorithm, these values were 5.6 min and 10.4 min, respectively. Despite the good agreement observed with both algorithms, the correlation coefficients ( $r = 0.69$  and  $0.64$ ) for SL were only moderate. Probably, the narrow range of variation has accounted for it.

The duration of wakefulness during the night showed systematic biases of  $-18.6$  and  $-10.0$  min for the Cole's and Sadeh's algorithms. The actigraphic measurements of IA, for 95% of the sample, according to Cole's algorithm were 57 min lower and 20 min higher than the PSG estimate, whereas for the Sadeh's algorithm these values were 53 min and 33 min, respectively. This lack of agreement was partially detected by the low correlation coefficients ( $r = 0.36$  and  $0.37$ , respectively). However, it should be noted that this wide concordance interval could be explained by the results obtained from two volunteers. They remained awake for a long period of time, but this was only partially detected by actigraphy. In fact, when their results were excluded from the analysis, the mean bias decreased (40% on average) and the concordance intervals shortened (35% on average), and the correlation coefficients also increased to a higher value to approximately 0.5.

A good agreement was observed for SE. Both algorithms tended to estimate slightly higher values: the biases were 4.0% and 2.0% for Cole's and Sadeh's algorithms, respectively. The sleep efficiency with the Cole's algorithm for 95% of the sample was 4.1% lower and 12.3% higher than the PSG values; and with the Sadeh's algorithm, these values were 7.0% and 11.4%. Despite the good agreement between methods, the correlation coefficients ( $r = 0.39$  and  $0.41$ ) were low for both algorithms. Again, the narrow range of variation might have accounted for it.

## DISCUSSION

In the present study, applying automatic sleep scoring to motor activity resulted in a good accuracy (91%) with both the algorithms (Cole's and Sadeh's) in comparison to PSG. These findings are in accordance with those previously obtained in healthy volunteers and patients with sleep and/or psychiatric disorders.<sup>1,2,8,10,11</sup> However, as already mentioned, general concordance calculations are insufficient, inasmuch as they do not consider the fact that most of the testing period are occupied by sleep, consequently raising the chance of high concordance.<sup>12</sup> In this context, calculation of the sensitivity and specificity expanded the information about the concordance between methods. Both algorithms showed high sensitivity, i.e., they detected most of the epochs classified as sleep by PSG. However, they were not specific, since they detected only a small proportion of waking epochs. This limitation was even great for the Cole's algorithm, in agreement with a previous report.<sup>14</sup>

The Pearson's correlation coefficients however were high for sleep latency and sleep duration, but for sleep efficiency and intermittent awakenings, they were not satisfactory, in consonance with previous studies.<sup>1,10</sup> These discrepancies might result from the small variability of these parameters in healthy volunteers, a factor influencing the magnitude of Pearson's correlation coefficient, and thus limiting its use as a concordance measurement.<sup>13,15</sup>

The Bland and Altman technique showed that actigraphy systematically overestimated sleep latency, total sleep time, and sleep efficiency, while it underestimated the awakenings.

The overestimation of the total sleep time might be a direct consequence of the limited capacity of the actigraphy to identify waking epochs during sleep. The large concordance interval for the awakenings throughout the night supports this idea. Thus, the discordance between the methods will occur when the subjects wake up but remain motionless. Tryon<sup>12</sup> had already pointed out that the discordance between actigraphy and PSG takes place mostly at the transitions between wake and sleep. These occasions would characteristically correspond to the sleep onset and to the intermittent awakenings during the night. However, our findings show that the actigraphy well detected the time taken to fall

asleep but not the awakenings throughout the night, although it had a small influence on the concordance between PSG and actigraphy regarding sleep efficiency. In fact, discrepancies between PSG and both algorithms were higher for the values of sleep efficiency lower than 90% and intermittent awakenings higher than 40 minutes. Furthermore, differences between actigraphy and PSG did not appear to be dependent on the magnitude of sleep latency and sleep duration. These findings corroborate the observation that the accuracy of actigraphy diminishes as the sleep probability decreases.<sup>19,20</sup>

Overall, both algorithms gave similar estimates of sleep latency and sleep efficiency, but the systematic biases for awakenings during the night and for sleep duration were smaller for Sadeh's than for Cole's algorithm.

In general, the available data indicate that automatic sleep scoring of motor activity is a useful method for sleep assessment despite its limitations regarding estimation of some sleep parameters. These limitations rely on how to decide if someone is awake or asleep (automatic scoring algorithms) and not on the activity measurement itself. However, regarding the sleep-wake pattern during daily routine, the actigraphy seems to add more naturalistic information. Thus, it might be helpful for monitoring sleep-wake patterns of insomnia, rhythm disorders, and inadequate perception of sleep in longitudinal studies. Actigraphy might even replace the self-assessment of sleep in some circumstances as it has been reported that patients with insomnia not only underestimate the duration of sleep but also overestimate the sleep latency.<sup>22</sup> The American Academy of Sleep Medicine<sup>23</sup> considers that actigraphy might help the assessment of the sleep-wake pattern of insomnia patients for extend periods, thus providing data not usually measured by PSG.

Finally, we hope that the present study might provide to the investigator, or to the clinician, information about the advantages and limitations of actigraphy. Furthermore, it would be important to emphasize that is up to the investigator or to the clinician to decide the extent to which concordance between actigraphy and polysomnography might influence research or clinical practice.

## REFERENCES

1. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep* 1992;15:461-69.
2. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: A empirical test of methodological issues. *Sleep* 1994;17:201-07.
3. Jean-Louis G, Von Gizycki H, Zizi F, et al. Determination of sleep and wakefulness with the Actigraph Data Analysis Software (ADAS). *Sleep* 1996;19:739-43.
4. Jean-Louis G, Zizi F, Von Gizycki H, Hauri P. Actigraphic assessment of sleep in insomnia: Application of the Actigraph Data Analysis Software (ADAS). *Physiol. Behav* 1999; 65: 659-63.
5. Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. *Sleep* 1992;15:293-301.
6. Sadeh A, Alster J, Urbach D, Lavie P. Actigraphically based automatic bedtime sleep-wake scoring: Validity and clinical applications. *J. Ambul. Monitoring* 1989;2:209-216.
7. Kim FC, Matsumoto M, Lewy AJ, Sack RL. Validation of the ActiLume wrist actigraphy monitor using polysomnography. *Sleep Res* 1997;26:673.
8. Kripke DF, Mullaney DJ, Messin S, Wyborny VG. Wrist actigraphic measures of sleep and rhythms. *Electroencephalogr. Clin. Neurophysiol* 1978;44:674-76.
9. Mason WJ, Kripke DF. Comparison of the actiLume and EEG for identifying total sleep time and wake after sleep onset. *Sleep Res* 1995;24A:482.
10. Mullaney DJ, Kripke DF, Messin S. Wrist actigraphic estimation of sleep time. *Sleep* 1980;3:83-92.
11. Zisapel R, Oxenberg A, Tarrasch R, Zisapel N. Wrist activity-based monitoring of nocturnal sleep: Validation of a novel scoring algorithm. *Sleep Res* 1995;24:506.
12. Tryon WW. Activity and Sleep. In: *Activity measurement in psychology and medicine*. New York, Plenum Press, 1991:179-181.
13. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-310.
14. Blood ML, Sack RL, Percy DC, Pen JC. A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. *Sleep* 1997;20:388-95.
15. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. *Lancet* 1995;346:1085-87.
16. First MB, Spitzer RL, Gibbon M, Williams JBW. Structured Clinical Interview for DSM-IV Axis I Disorders - Non-patient Edition (SCID-I/NP, Version 2.0). Biometrics Research Department, New York, 1996.
17. Buysse D J, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatr. Res* 1989;28:193-213.
18. 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.

19. Hauri PJ. Wrist actigraphy in insomniacs – *Sleep Res* 1989;18:239.
20. Levine B, Moyles T, Roehrs T, Fortier J, Roth R. Actigraphic monitoring and polygraphic recording in determination of sleep and wake. *Sleep Res* 1986;15:247.
21. Kubicki St, Holler 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 Stagers. *Sleep* 1989;12:140-49.
22. Frankel BL, Coursey D, Buchbinder R, Snyder FS. Recorded and reported sleep in chronic primary insomnia. *Arch Gen Psychiat* 1976;33:615-23.
23. Sadeh A, Hauri PJ, Kripke DF, Lavie P. ASDA Report – Practice Parameters for the Use of Actigraphy in the Clinical Assessment of Sleep Disorders. *Sleep* 1995;18:285 – 87.