A Comparison of Actigraphy and Polysomnography in Older Adults Treated for Chronic Primary Insomnia

Børge Sivertsen, PsyD; Siri Omvik, PsyD; Odd E. Havik, PhD; Ståle Pallesen, PhD; Bjørn Bjorvatn, MD, PhD; Geir Høstmark Nielsen, PsyD; Silvert Straume, BA; Inger Hilde Nordhus, PhD

1Department of Clinical Psychology and 2Department of Psychosocial Science, University of Bergen, Bergen, Norway; 3Norwegian Competence Center for Sleep Disorders, Bergen, Norway; 4Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway

Study Objectives: The present study explores the accuracy and clinical utility of actigraphy compared with polysomnography in older adults treated for chronic primary insomnia.

Design: Polysomnographic and actigraphic data were collected before and after treatment.

Setting: A university-based outpatient clinic for adults and elderly.

Participants: Thirty-four participants with chronic primary insomnia. Mean age was 60.5 years.

Interventions: Participants received either a manualized treatment package based on cognitive-behavior therapy and sleep management or hypnotic-drug treatment (7.5 mg zopiclone) for 6 weeks (these findings are reported elsewhere).

Measurements and results: Although the sensitivity of actigraphy to detect sleep was very high (95.2%), actigraphy performed poorly in detecting wakefulness (specificity: 36.3%), yielding on an overall level of accuracy of 83.1%. However, the level of actigraphy accuracy was dependent upon polysomnography-registered sleep efficiency. Actigraphy underestimated total sleep time and sleep-onset latency and consequently overestimated total sleep time and sleep efficiency. Compared with polysomnography, actigraphy captured only part of the treatment effects on total wake time and sleep-onset latency and failed to detect significant changes in sleep efficiency.

Conclusions: The present findings suggest that the clinical utility of actigraphy is still suboptimal in older adults treated for chronic primary insomnia and should, hence, be used in this clinical setting with the concurrent use of supplementary assessment methods.

Keywords: Insomnia, actigraphy, polysomnography, accuracy, sleep-wake perception

Citation: Sivertsen B; Omvik S; Havik OE et al. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. SLEEP 2006;29(10):1353-1358.

INTRODUCTION

SINCE THE INTRODUCTION OF POLYSOMNOGRAPHY (PSG) IN THE 1950S, THIS METHOD HAS BEEN REGARDED AS THE GOLD STANDARD FOR OBJECTIVE assessment of sleep, providing measures of both sleep and wake time, in addition to classification of sleep stages. Even though PSG offers extensive information on sleep behavior and sleep physiology, it is very expensive and can sometimes be too invasive to be used in clinical studies in which the primary focus is quantification of sleep time, wake time, or both. Actigraphy has been suggested as an alternative assessment method to PSG. Actigraphy consists of an accelerometer and memory storage, both fitted into a watch-like device to be worn around the wrist. Based on differences in movements associated with wakefulness and sleep, actigraphy provides an estimate of sleep-wake schedules. Being independent of patients’ ratings and personal judgments of their sleep, actigraphy may offer a less time-consuming and less expensive alternative to PSG. In addition, actigraphy is regarded as a less invasive measure.

Studies have shown actigraphy to correspond well with PSG in adults with impaired sleep (concordances ranging from 81% to 87%). However, the level of agreement is often reported as crude percentage or correlation coefficients. This may overestimate the level of agreement because an estimate of agreement based on crude percentage fails to take into account the chance agreement due to the high base rate of sleep during the night. Furthermore, agreement has been reported as an average across participants, and this may mask individual differences in correspondence between actigraphy and PSG. This possibility of interindividual differences should be explored because it may have implications for the use of actigraphy as a method in detecting individual changes in sleep behavior.

A more informative approach would be Tryon’s method of separately calculating and reporting sensitivity, specificity, and overall accuracy. In this view, sensitivity for sleep is the proportion of PSG-registered sleep epochs also identified as sleep by actigraphy, whereas specificity for sleep is the proportion of wake epochs correctly identified by actigraphy. By using this scheme, actigraphy has been shown to be more successful in detecting sleep (sensitivity), compared with detecting wakefulness (specificity). Only a few studies have examined to what extent actigraphy is able to detect treatment changes, as compared with PSG. Verbeek et al compared actigraphy, PSG, and sleep diaries in a sample of 20 middle-aged insomniacs, concluding that actigraphy overestimated the amount of sleep compared to PSG. Also, Vallières and Morin demonstrated that actigraphy was sensitive enough to detect treatment changed following cognitive-behavior therapy in a similar-aged cohort. Although not all sleep changes...
Participants and Procedure

Participants were recruited through newspaper advertisements to a randomized controlled study of the treatment of insomnia. Inclusion criteria were (1) age 55 years or older; (2) fulfillment of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) criteria for insomni, including difficulties initiating sleep, maintaining sleep, and/or early morning awakenings without returning to sleep; (3) duration of insomnia of at least 6 months; and (4) complaints of impaired daytime functioning. The following exclusion criteria were used: (1) use of hypnotic medication (including all types of benzodiazepines) the last 4 weeks before project start, (2) use of antidepressive or antipsychotic medications, (3) signs of cognitive impairment defined by a score less than 23 on the Mini-Mental State Examination, (4) presence of a major depressive disorder or other severe mental disorder as identified by clinical assessment based on the Structured Clinical Interview for DSM-IV (SCID-I), (5) presence of sleep apnea (apnea-hypopnea index > 15) or periodic limb movements in sleep (periodic limb movement index with arousal > 15), (6) working nightshifts and unable or unwilling to discontinue this work pattern, (7) unwillingness or inability to stop taking sleep medication before the start of the study, or (8) the presence of a serious somatic conditions preventing further participation.

Participants who responded to the advertisement (n = 92) underwent several screening processes before they were included. A short telephone interview (15 minutes) ensured that participants fulfilled the basic criteria for inclusion. Accepted participants (n = 75) met at the Department of Clinical Psychology, University of Bergen, for a diagnostic interview (SCID-I), screening for severe psychopathology and cognitive impairment. In all, 29 participants were excluded for the following reasons: sleep apnea or periodic limb movements in sleep (n = 16), spontaneous improvement before treatment start (n = 4), psychiatric condition (n = 3), not able to stop taking sleep medication (n = 1), and voluntary withdrawal (n = 5). After study inclusion, we excluded participants who did not complete the treatment protocol (n = 4) or forgot to wear the actigraph during the night of PSG registration (n = 3). We also had to exclude 5 participants due to technical problems with the actigraphs.

In all, 34 participants were included in the present study, as compared with 46 participants originally enrolled in the treatment program. The average age of included participants was 60.5 (SD = 4.5). Twenty-seven of the participants (79.4%) were under 65 years of age, yielding a skewness of 1.58 towards the lower end of the age distribution. The sample consisted of 17 men and 17 women, and years of education beyond compulsory primary and secondary schools were, on average, 5.3 years (SD = 2.5). Insomnia duration was on average 15.6 years (SD = 12.8). The sample mainly consisted of participants with maintenance insomnia; 27 of the 34 participants had a sleep-onset latency of less than 30 minutes.

Participants were enrolled in a 6-week treatment program and were randomly assigned to cognitive-behavior therapy, 7.5 mg of zopiclone, or placebo treatment. Participants in the placebo condition were randomly assigned to 1 of the 2 active treatment conditions after 6 weeks. Detailed information about the treatment protocols is reported elsewhere. Participants underwent 2 consecutive nights of PSG before treatment to allow for patient adaptation to the PSG montage. At treatment completion, 1 night of PSG was used because a recent study on healthy participants suggested that the so-called “first-night effect” is primarily present in the first night in the first assessment period. Because no treatment effect was found in the placebo condition, only data from the 2 active-treatment conditions are presented here.

The study was approved by the National Data Inspectorate, the Norwegian Medicines Agency, and The Regional Committee for Medical Research Ethics in Western Norway. Written informed consent was obtained from all participants included in this study.

MEASURES

Polysomnography

Sleep variables were assessed by ambulant clinical PSG before and after treatment. The PSG montage included electroencephalographic, electromyographic, and electrooculographic monitoring. Sleep stages, respiratory disturbances, and limb movements were scored according to standard criteria by 2 experienced technicians who were blinded to the participants’ condition. Respiration (air flow, tidal volume, and oxygen saturation) and anterior tibialis electromyographic readings were recorded to detect sleep apnea or periodic limb movements. All electrophysiologic signals were acquired using Embla A10 (Flaga-Medcare Somnologica® 3.2 software package, Reykjavik, Iceland).

Actigraphy

Actigraphic data were collected using the Actiwatches manufactured by Cambridge Neurotechnology and scored by the Actiwatches software (Actiwatches Sleep Analysis 2001 software, Version 1.19; Cambridge Neurotechnology Ltd., Cambridgeshire, UK). The sensitivity of the actigraph was set to the medium level, and 30-second epochs were used to allow epoch-by-epoch comparison with PSG. This actigraph had a button that enabled the participants to signal when they tried to fall asleep and when they got out of bed in the morning.

Data Analysis Plan

Agreement or disagreement between actigraphic data and PSG

SLEEP, Vol. 29, No. 10, 2006
There were no significant correlations between age and actigraphy (r = .03, df = 32, p = .80), these were pooled for the present analyses. There were no statistical significant differences between groups (pretreatment and posttreatment) for any of the clinical variables. Because there were no differences in actigraphy accuracy before or after treatment, we used the mean of all epoch classifications by PSG and actigraphy as our dependent variable. Paired-sample t-tests were used to examine the differences between pretreatment and posttreatment assessments. Differences between participants were not examined. To examine the relationship between the participants’ PSG-registered sleep efficiency and the accuracy level of their actigraphy recordings, Wilcoxon signed rank test was used. To examine to what extent actigraphy overestimated or underestimated the sleep measures, as compared with PSG, we used Wilcoxon signed rank tests to examine the differences between the pretreatment and posttreatment measures for the accuracy average across the 34 participants. Interindividual differences in level of accuracy were dependent on the quality of sleep, as defined by PSG-registered sleep efficiency (Figure 1). These 2 measures were highly correlated both before (r = 0.54, p = .001) and after treatment (r = 0.66, p < .001), indicating that the accuracy of the actigraph was significantly higher in participants with high sleep efficiency. However, PSG-registered sleep efficiency was not significantly correlated with either the sensitivity (r = 0.04, p = .81 and r = -0.17, p = .34) or specificity measures (r = 0.19, p = .28 and r = 0.27, p = .12) before or after treatment. Actigraphy accuracy was significantly higher in women (84.7%) than in men (81.0%) (t = 2.27, df = 32, p = .02), indicating that the accuracy of the actigraph was significantly higher in women (84.7%) than in men (81.0%) (t = 2.27, df = 32, p = .02). When collapsing the pretreatment and posttreatment measures, the sensitivity of actigraphy compared with PSG was good (sensitivity = 95.2%). However, the specificity of actigraphy was low (36.3%), yielding an overall accuracy of 83.1%. The overall accuracy level was slightly lower (81.9%) when based on all the 61225 epochs, as compared with taking the accuracy average across the 34 participants.

| Table 1—Number of epochs classified as sleep and wake using actigraphy and polysomnography
<table>
<thead>
<tr>
<th>Polysomnography</th>
<th>Sleep</th>
<th>Wake</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actigraphic sleep</td>
<td>45642</td>
<td>8690</td>
<td>54322</td>
</tr>
<tr>
<td>Actigraphic wake</td>
<td>2373</td>
<td>4530</td>
<td>6903</td>
</tr>
<tr>
<td>Total</td>
<td>48015</td>
<td>13220</td>
<td>61235</td>
</tr>
</tbody>
</table>

Table entries are frequencies of 30-second epochs.

Actigraphy and PSG in Older Insomniacs —Sivertsen et al

RESULTS

Preliminary analyses

There were no statistical significant differences between included or excluded participants on baseline demographic and clinical variables. Because there were no differences in actigraphy accuracy across the 2 treatment conditions at pretreatment (t = 0.68, df = 32, p = .50), these were pooled for the present analyses. There were no significant correlations between age and actigraphy sensitivity, specificity, or accuracy. Also, no differences in actigraphy accuracy was found between participants with sleep-onset insomnia (PSG-registered sleep-onset latency ≥30 minutes) and other types insomnia (t = 1.27, df = 32, p = .22).

Concordance Between Actigraphy and PSG

A comparison of all epoch classifications by PSG and actigraphy is presented in Table 1. When epochs scored by actigraphy deviated from the PSG classification of epochs, it was 3.6 times more likely that actigraphy misclassified the epoch as sleep than wake (8690 misclassified sleep epochs versus 2373 misclassified wake epochs). The sensitivity, specificity, and accuracy of the epoch-by-epoch comparisons of actigraphy and PSG are shown in Table 2. There were no statistically significant differences between the pretreatment and posttreatment measures on sensitivity and specificity. However, the accuracy after treatment (85.3%) was significantly higher than before treatment (80.8%), t = 2.84, df = 33, p < .01. When collapsing the pretreatment and posttreatment measures, the sensitivity of actigraphy compared with PSG was good (sensitivity = 95.2%). However, the specificity of actigraphy was low (36.3%), yielding an overall accuracy of 83.1%. The overall accuracy level was slightly lower (81.9%) when based on all the 61225 epochs, as compared with taking the accuracy average across the 34 participants.

Interindividual differences in level of accuracy were dependent on the quality of sleep, as defined by PSG-registered sleep efficiency (Figure 1). These 2 measures were highly correlated both before (r = 0.54, p = .001) and after treatment (r = 0.66, p < .001), indicating that the accuracy of the actigraph was significantly higher in participants with high sleep efficiency. However, PSG-registered sleep efficiency was not significantly correlated with either the sensitivity (r = -0.04, p = .81 and r = -0.17, p = .34) or specificity measures (r = 0.19, p = .28 and r = 0.27, p = .12) before or after treatment. Actigraphy accuracy was significantly higher in women (84.7%) than in men (81.0%) (t = 2.27, df = 32, p = .03) when taking the average of assessments before

Table 2—Statistical Measures of Epoch-by-Epoch Comparisons of Actigraphy and Polysomnography Before and After Treatment

<table>
<thead>
<tr>
<th></th>
<th>Before Treatment</th>
<th>After Treatment</th>
<th>Totala</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>94.9</td>
<td>5.3</td>
<td>77.6-99.9</td>
</tr>
<tr>
<td>Specificity</td>
<td>34.0</td>
<td>14.5</td>
<td>5.6-70.7</td>
</tr>
<tr>
<td>Accuracy</td>
<td>80.8</td>
<td>7.2</td>
<td>66.2-94.1</td>
</tr>
<tr>
<td>Epochs, no.</td>
<td>937.8</td>
<td>184.3</td>
<td>351-1230</td>
</tr>
</tbody>
</table>

Total number of epochs analyzed: a = 33163 epochs, b = 28072 epochs; c = 61235 epochs.
Estimation of Treatment Effects

Inspection of the mean differences indicated that actigraphy, as compared with PSG, captured only parts of the treatment effects on 3 of the 4 outcome variables. For example, compared with PSG, actigraphy detected only 60% of the changes in sleep efficiency (7.0% versus 4.2% improvement, respectively) from baseline to after treatment, and these improvements were not significantly intercorrelated (r = .07, p = .68). The same pattern was found for total wake time and sleep-onset latency, with actigraphy detecting 77% and 21%, respectively, of the changes from before to after treatment, and these improvements were not significantly different (7.0% versus 4.2% improvement, respectively) from baseline to after treatment. In contrast, actigraphy detected a 20% larger changes, although the magnitude of change was larger when measured with PSG (34%) than with actigraphy (7%).

DISCUSSION

The objective of the present study was to examine actigraphy sensitivity, specificity, and accuracy for sleep-wakefulness categorization, compared with PSG, in a sample consisting of older adults treated for chronic primary insomnia. We also wanted to explore to what extent actigraphy was able to capture treatment changes, as estimated by PSG. In summary, we found actigraphy to have a high level of sensitivity level (95.2%). However, the specificity of actigraphy, ie, the ability to detect wakefulness, was much lower (36.3%), yielding an overall accuracy level of 83.1%. Compared with PSG, actigraphy detected treatment changes on 3 of 4 outcome measures.

Although previous studies have shown a good concordance
between PSG and actigraphy (see for example, Tryon).\textsuperscript{21} many of these studies have merely reported bivariate correlation coefficients or percentage agreement, thus failing to correct for the expected chance agreement due to the high proportion of sleep epochs compared to wake epochs. Although the present findings suggest that the accuracy of the actigraphy in this study is similar or slightly better compared with the results from studies on the use of actigraphy in other sleep-related disorders,\textsuperscript{4,6} the low specificity remains problematic because it indicates that a high base rate of sleep may still produce high accuracy despite modest performance in detecting wakefulness. Furthermore, there were large individual differences on the specificity measure of the actigraph, which underscores its problems on capturing wake time in some persons. Another important finding is that the accuracy level (although not the sensitivity or specificity level) was dependent on the quality of sleep. This indicates that the level of accuracy in persons with severely impaired sleep (a larger proportion of wake epochs) will be considerably lower, thus reducing the applicability of actigraphy in these persons. However, the accuracy of the actigraph was also low in some participants with relatively good sleep (PSG-registered sleep efficiency > 80\%). To further improve the accuracy of actigraphy, the computer algorithms on which the actigraphs are based should be derived from a population that resembles the nature of the population being studied. An approach that weights nighttime movements differently has previously been attempted in other sleep disorders\textsuperscript{22} and may also enhance the ability of the actigraph to adequately capture sleep-wake patterns in elderly insomniacs.

Still, as noted by Tryon,\textsuperscript{23} one cannot expect the correspondence of actigraphy with PSG to exceed the reliability of PSG scorings. In an early study by Spiegel, the reliability of scoring stage-1 sleep on PSG was estimated to be only around 60%,\textsuperscript{24} suggesting that a substantial proportion of the difference between PSG and actigraphy (sleep-wake scoring) may be ascribed to individual differences in scoring of stage-1 sleep on PSG.\textsuperscript{21} However, in a more recent study, Whitney et al\textsuperscript{25} showed that the percentage agreement on wakefulness scoring across scorers was between 87\% to 90\%, suggesting a much higher reliability.

Despite systematic differences and errors between actigraphy and PSG, if actigraphy can still detect clinical improvements over time following treatments, the clinical utility of actigraphy can be supported. A few studies have demonstrated that actigraphy is able to detect treatment effects on several sleep parameters in middle-aged adults.\textsuperscript{10,11} Likewise, 2 small-scale studies of older adults have indicated a similar ability of actigraphy to detect changes following treatment. Our results are somewhat more ambiguous, with actigraphy failing to detect significant changes in sleep efficiency. This may partly be explained by the fact that actigraphy significantly overestimated both total sleep time and sleep efficiency before treatment and, conversely, underestimated the amount of wakefulness, leading to a ceiling effect. However, actigraphy was still able to identify significant changes in both total sleep time and total wake time, underscoring the importance of using multiple sleep measures in treatment studies. The present study has some methodologic limitations. First, due to technical problems with some of the actigraphs, the number of participants is lower than the number originally enrolled in the treatment study.\textsuperscript{16} The sample size was additionally reduced because some participants forgot to wear their actigraph on the night of PSG registration. We also excluded participants who did not complete the treatment protocol. Thus, we cannot rule out the influence of selection bias. Still, we consider this less likely because the participants lost in the study did not differ from those included. Second, the average sleep quality at pretreatment assessment was relatively adequate (PSG sleep efficiency of 81.8\%), which may limit the generalizability to persons whose sleep is more severely affected. Moreover, the results may not generalize to persons whose sleep problems include insomnia secondary to psychiatric or medical conditions. It should also be noted that both the hardware and software used for the registration and scoring in the present study may differ from those used in other studies, which may complicate direct comparisons of results across studies. Furthermore, we did not assess the stability of actigraphy over several nights. In a study by Hilton et al\textsuperscript{26} on 99 older healthy participants, the authors found significant internight and intrasubject variability, although they found no evidence of a “first-night effect.” Future research should also seek to explore the stability of actigraphy in samples of subjects with sleep problems. Finally, it should be noted that actigraphy might have other potential utilities, such as weekly progress reinforcement or compliance verification during treatment.

In summary, although recent reviews have suggested that actigraphy can be used as a reliable instrument for evaluating sleep patterns in adults,\textsuperscript{23,27,28} our findings indicate that this suggestion is not fully supported for older insomniacs. Future research should explore the circumstances under which and the individuals for whom actigraphy may capture sufficient sleep information to be used in a clinical setting without the concurrent use of PSG.

REFERENCES

13. Morin CM, Gramling SE. Sleep patterns and aging: Comparison of