Actigraphy Validation with Insomnia

Kenneth L. Lichstein, PhD1; Kristen C. Stone, MS2; James Donaldson, RPSGT, CRT3; Sidney D. Nau, PhD3; James P. Soeffing3; David Murray, PhD3; Kristin W. Lester, BA3; R. Neal Aguillard, MD4

1The University of Alabama, Tuscaloosa, AL; 2University of Memphis, Memphis, TN; 3Methodist University Hospital, Memphis, TN

Study Objective: Actigraphy, a method of inferring sleep from the presence or absence of wrist movement, has been well validated against polysomnography in trials with people without insomnia. However, the small amount of literature on validation with insomniacs has revealed an actigraphy bias toward overscoring sleep. The current validation trial with insomniacs used the largest number of subjects to date in such research and attracted participants with diverse demographic characteristics.

Design: People with insomnia slept 1 night in the laboratory while simultaneously being monitored by polysomnography, actigraphy (high-sensitivity algorithm of the Mini Mitter AW64 Actiwatch™), and morning sleep diary.

Setting: Sleep disorders center.

Participants: Participants were 57 volunteers from the community, 26 men and 31 women, ranging in age from 21 to 87 years. All participants satisfied conservative criteria for insomnia. The sample included subjects with primary insomnia, subjects with comorbid insomnia, and hypnotic users with current insomnia complaints.

Introduction

Actigraphy is a method of assessment that infers wakefulness and sleep from the presence or absence of limb movement. Actigraphy yields typical sleep-pattern measures, such as latency to sleep (SOL), wake time after sleep onset (WASO), and total sleep time (TST), but does not assess sleep architecture. The advantages of the actigraph—its objectivity, portability, and convenience—account for its widespread use in measuring sleep. The actigraph combines a movement detector and memory storage on a watch-like device worn on the wrist or ankle. The device is electrode-free and can be worn continuously day and night for periods longer than 1 week.

More sensitive than sleep diaries for documenting sleep fragmentation, the actigraph can also be used with people who cannot fill out sleep logs, such as infants and adults who cannot read or write. The actigraph is less expensive, noninvasive, and more conducive to repeated measures in comparison with polysomnography (PSG).

Conclusions: Actigraphy proved to be a satisfactory objective measure of sleep on 4 of 5 sleep parameters, but these results are specific to this particular instrument using this particular algorithm and should not be construed as a blanket endorsement of actigraphy for measuring insomnia.

Keywords: Actigraphy, insomnia, validate

Citation: Lichstein KL; Stone KC; Donaldson J et al. Actigraphy validation with insomnia. SLEEP 2006;29(2): 232-239.

In the last 2 decades, actigraphic and polysomnographic measures of sleep have been strongly correlated in normal sleepers and in sleepers with apnea, with correlation coefficients ranging from .89 to .98. Actigraphy is capable of identifying sleep patterns characteristic of sleep apnea and periodic leg movements and has been used extensively in intervention studies tracking patients’ sleep progress over time.

Insomnia, however, has been largely neglected in validation studies of actigraphy. Individual differences of movement patterns, especially among people with insomnia, hinder the potential of actigraphy. The main challenge for actigraphy with insomnia occurs when patients lie awake in bed motionless, which the actigraph scores as sleep. Only 4 studies have tested the ability of actigraphy to accurately score the sleep of people with insomnia. The first 2 actigraphy validation studies compared actigraphy and PSG on TST. Hauri and Bissey found that actigraphy overscored TST by an average 49 minutes compared with PSG in a study of 20 individuals. Jean-Louis et al reanalyzed these same data using the Actigraph Data Analysis Software—as opposed to the Actigraph Scoring Analysis used in the original study—reducing the average discrepancy to 25 minutes. The difference in the software of the 2 studies consisted of a change in the computer algorithms used to score sleep and wake. Jean-Louis et al developed an Actigraphic Sleep Threshold of 10 activity counts and a wake interval after arousal of 3 minutes (activity counts above the Actigraphic Sleep Threshold that lasted 3 minutes or less were scored as sleep because they were thought to be nonwaking arousals) for the Actigraph Data Analysis Software. Clearly, they found their algorithms advantageous over the algorithms developed by Sadeh et al for the Actigraph Scoring Analysis. Using a completely different algorithm created by Cole et al, Verbeek, Arends, Declerck, and Beecher found that actigraphs overscored TST by an average of 48 minutes compared with PSG in a study of 20 people.
middle-aged and older individuals complaining of insomnia.

More recently, 2 actigraphic validation studies were conducted that also revealed significant discrepancies between actigraphy and PSG. Kushida et al\(^1\) compared PSG and actigraphy on TST, sleep efficiency percentage, and number of awakenings in 100 middle-aged and older patients, each with at least 1 of the following sleep disorders: obstructive sleep apnea syndrome, upper airway resistance syndrome, periodic limb movement disorder, insomnia (n less than 15), narcolepsy, idiopathic hypersomnia, or restless legs syndrome. The validated algorithm used with the analysis software in this study modified the activity counts of each epoch according to the level of activity in the surrounding 2-minute time period. Actigraphic measures of number of awakenings were not significantly different from PSG, though TST was overscored 1.0 to 1.8 hours and sleep efficiency was subsequently overscored 12.1% to 29.1%. Kushida et al\(^2\) found that the actigraph overscored TST to a larger degree than subjective measures but found actigraphy to be more sensitive than sleep logs in detecting sleep fragmentation. These results reflect the entire sample and, therefore, do not speak directly to the accuracy of actigraphy with people with insomnia.

In a study of 17 middle-aged individuals diagnosed with chronic primary insomnia, Vallières and Morin\(^3\) found that actigraphy underestimated SOL compared with PSG. They did not report the scoring algorithms used. Their study revealed actigraphy to be more accurate than sleep diaries and recommended its use as a complement to sleep diaries.

The existing studies suggest actigraphy has a bias in overscoring sleep in people with insomnia. Given that actigraphy measures movement, it is noteworthy that across age groups, men have different patterns of nocturnal activity than do women,\(^4\) suggesting that sex could affect actigraphic accuracy. However, none of the existing studies have evaluated sex or as possible influences. Furthermore, all 4 validation studies of actigraphy with insomnia had small sample sizes of insomnia patients, only 1 of them evaluated actigraphic measures of WASO, and only 1 evaluated actigraphic measures of SOL. Accurate estimations of SOL and WASO are necessary for an adequate portrayal of insomnia.

The current study attempted to validate actigraphy with people with insomnia and incorporated methodologic controls not present in the existing literature. The current study included SOL and WASO as well as a larger sample size of people with insomnia, comparisons of actigraphy with subjective measures and PSG, and an analysis of the effects of age and sex to more clearly illuminate the limitations and potential of actigraphy and its role in assessing insomnia.

METHODS

Participants

Twenty-six men and 31 women (9 African Americans and 48 Caucasians), 21 to 87 years in age (eight 21- to 39-year-olds, twenty-two 40- to 59-year-olds, and twenty-seven 60- to 87-year-olds), participated in the study. This study was approved by the institutional review boards at both investigational sites (Methodist University Hospital and the University of Memphis), and participants signed appropriate consent forms.

Participants were recruited from media advertisements, from undergraduate psychology courses, and from a pool of volunteers from a previous insomnia study who had consented to being contacted for future studies. Telephone interviews were conducted with the volunteers to collect demographic information, to determine if they were interested in participating, and to determine if they met criteria, which consisted of the adult age restriction and a complaint of insomnia of at least 6 months’ duration.

Interviews also included questions designed to rule out subjects with narcolepsy, sleep apnea, restless legs, and periodic limb movement. Questions designed to rule out those with narcolepsy included, “Do you have sleep attacks during the day or paralysis at sleep onset?” To assess the presence of restless legs and periodic limb movement, participants were asked if their legs jerked during the night and if their legs felt restless before sleep onset. In addition to information about sex, age, weight, height, and smoking habits, the following questions are examples of those used to address sleep apnea: Do you fall asleep while watching TV, reading, or driving a car?; Has anyone ever told you that you seem to have difficulty breathing when you are sleeping or that you sometimes seem to stop breathing in your sleep?; Are you aware of a choking sensation when you sleep?; How much of a problem do you have with sleepiness during the day?; Has anyone ever told you that you snore? (How often during the night do you snore in any way?, How often during the night do you snore loudly and disruptively?, Does your snoring ever wake you?); How many naps do you take during a usual weekday? Depending on the answers to initial screening questions for sleep disorders, more specific questions were asked, including questions of available bed partners if permission was granted by the participants.

Diagnostic criteria for chronic insomnia\(^5\) include a current complaint of insomnia lasting at least 6 months and a report of impaired daytime functioning. Our participants had to meet these criteria as well as the self-reported quantitative criteria derived from 2 weeks of sleep diaries of a SOL or WASO that was at least 31 minutes long, occurring at least 3 nights a week.\(^6\)

Participants qualifying based on their responses to the telephone interview and who agreed to participate were given 2 weeks of sleep diaries to verify the quantitative criteria. To establish the complaint of impaired daytime functioning, 5 questionnaires were provided to volunteers along with the sleep diaries. The questionnaires and their accompanying cutoff scores are as follows: Epworth Sleepiness Scale,\(^7\) score of 7.4 or greater; Insomnia Impact Scale (shown elsewhere\(^8\)), score of 125 or greater; Fatigue Severity Scale,\(^9\) score of 5.5 or greater; Beck Depression Inventory,\(^10\) score of 10 or greater; and State-Trait Anxiety Inventory, Trait Scale,\(^11\) score of 37 or greater. Participants satisfying at least 1 of these criteria were judged to have a complaint of impaired daytime functioning. The justification for selecting these cutoff scores has been discussed elsewhere.\(^12\) After the telephone interview, an information meeting was held at which time participants signed informed consents.

Considering the majority of the insomnia population has a medical or psychiatric disorder in addition to insomnia,\(^13\) are taking a sleep active medication, or both, there were no screening criteria concerning comorbidity or substance use to test the full range of people with insomnia. Thus, our sample included individuals with insomnia independent of medical conditions, mental disorders, and substances, as well as individuals with insomnia dependent on these factors or coexisting with them. However, the presence of another sleep disorder may preclude the diagnosis of insomnia and, therefore, merit disqualification. Of the 68
participants originally recruited, 11 were disqualified: 7 due to discovery of sleep apnea during PSG, 3 due to technical difficulties with the actigraph, and 1 due to nocturnal seizures.

Based on the participants’ self-report of mental disorders, general medical conditions, and hypnotic use, the final sample consisted of 23 individuals with primary insomnia, 11 individuals (nonhypnotic users) with comorbid insomnia, 11 hypnotic users with current insomnia, and 12 hypnotic users with comorbid insomnia. Fifteen of the hypnotic users were taking 1 medication: benzodiazepines (6 subjects), nonbenzodiazepine agonists (5), sedating antidepressants (3), and 1 participant couldn’t remember. Eight participants were taking multiple hypnotics, combinations of the above.

When PSG revealed other sleep disorders or medical disorders, this information was shared with the participants. Participants were also compensated with 4 free treatment sessions for insomnia after completion of the study. The treatment consisted of well-established psychological interventions for insomnia: stimulus control, relaxation, and sleep hygiene. The techniques were practiced by the participant and monitored by the therapist with support sessions each lasting about 45 minutes once a week for 4 weeks. Treatments were conducted by graduate students in psychology.

**Instrumentation**

Mini-Mitter Company, Inc. provided 3 Mini Mitter actigraphy monitors (AW64 Actiwatch®) and associated software and hardware for this study. The Actiwatch looks like a wristwatch without a time face and is secured by a strap. It utilizes an accelerometer with a sensitivity of less than 0.01 g-force to measure motion by producing an electrical current corresponding to the degree of activity. The AW64 stores this electrical current as an activity count. The AW64 output is scored using proprietary software, Actiware Sleep v. 3.3 analysis software. The sampling frequency of the AW64 is 32 Hz. Actiware epoch length is the sampling interval in minutes. For example, the standard epoch length for sleep laboratory research is 30 seconds, which is denoted by Actiware as 0.5. To match PSG, this was the Actiwatch epoch used in this study. An event-marker button on the device can be depressed to mark such occurrences as bedtime and out-of-bed time. The Actiwatch will yield measures of SOL, number of awakenings, WASO, TST, and sleep efficiency (the ratio of TST to total time in bed × 100).

Pilot studies of Actiwatch sensitivity thresholds have revealed that, when actigraphy is compared with PSG, high actigraphic sensitivity (when an activity value of 20 or greater is considered wake) is more accurate than low sensitivity (when an activity value of 80 or greater is considered wake) in use with people with insomnia. Specifically, paired t-tests revealed discrepancies between the sensitivities. At high actigraphic sensitivity, the instruments did not produce significantly different mean values for any sleep measure. However, at low actigraphic sensitivity, there were significant differences between actigraphy and PSG on 3 mean sleep values: WASO, t (56) = -6.23, p < .001, TST, t (56) = 6.25, p < .001, and SE, t (56) = 5.72, p < .001. Therefore, the high sensitivity threshold was used to calculate the actigraphic measures of SOL, number of awakenings, WASO, TST, and sleep efficiency.

The scoring algorithm of Actiware Sleep v. 3.3 analysis software, similar to that used in the study conducted by Kushida et al., calculates the activity counts for a given epoch based on the activity level of the surrounding epochs, which are weighted based on their proximity to the given epoch. For instance, for a 30-second epoch, the activity counts of the surrounding 2 epochs are multiplied by one-fifth and the activity counts of the 2 epochs beyond those are multiplied by one-twenty-fifth. These values are then added to the activity counts of the given epoch for a total activity-count value. If this value is less than or equal to the designated threshold value, the epoch is scored as sleep.

Concomitantly, a Respiromics’ Alice 3 Infant and Adult computerized PSG system monitored participants’ sleep. Monitoring consisted of 2 electroencephalography channels, 2 electrooculography, and chin electromyography according to standard placements to score sleep stages. Supplementary channels included oxygen saturation level, bilateral anterior tibialis electromyogram, heart rate, thoracic strain gauge, and a nasal-oral thermistor.

To ensure that the epochs of PSG and actigraphy were synchronized, which is a critical feature of actigraphy validation studies, the same computer that housed the PSG system was used to program the Actiwatch data-collection start time for each participant. Furthermore, each PSG system and corresponding Actiwatch were programmed to start at the same time, approximately 9:00 PM for each participant. Using the same computer for the 2 instruments ensured that the instruments were operating on the same “clock” and, thus, keeping identical time.

PSG records were manually scored in 30-second epochs by a registered PSG technician according to the criteria of Rechtschaffen and Kales. A second technician randomly selected 25% of the records and scored them independently. Meetings between the technicians and the primary investigator compared the double-scored records, resolved discrepancies, and reviewed scoring standards to promote accuracy and consistency. PSG produced the same sleep-pattern values as actigraphy: SOL, number of awakenings, WASO, TST, and sleep efficiency.

Each participant completed a sleep diary (shown elsewhere) after the morning awakening. The diary provided a record of the time they entered bed the night before, how long it took them to fall asleep, number and duration of awakenings during the night, what time they awoke, and what time they actually got out of bed. From this information, we obtained the same 5 sleep-pattern measures calculated from actigraphy and PSG.

**Procedure**

During the information meeting, sleep diaries, daytime-functioning questionnaires, informed consent, and details of the study were explained to the qualifying participants. The participants had opportunities to ask questions and sign the consent. Sleep diaries and daytime-functioning questionnaires were screened upon completion (approximately 2 weeks later), and participants meeting the insomnia criteria spent 1 night at Methodist University Hospital, an accredited sleep disorders center, at which time their sleep was simultaneously measured using actigraphy, sleep diary, and PSG. Participants arrived at the sleep center around 8:30 PM, at which time the AW64 and the electrodes were placed on the participants, giving them time to become acclimated to the instrumentation and sleep center environment at a relaxed pace.

The actigraph was secured on the participant’s dominant wrist. Although more studies use nondominant rather than dominant wrist placement, it is, of yet, unclear which wrist yields the most...
Actigraphy Validation

The following section describes the results of analytic procedures testing for differences between instruments and evaluating the roles of age, sex, and hypnotics on those differences. Because each comparison was multiplied by the 5 sleep values, a Bonferroni adjusted \( \alpha = .05/5 \) or .01 was employed for every test.

### Differences Between Instruments

To address the question of whether or not the instruments produced significantly different mean sleep values, we performed five 1-way repeated-measures analyses of variance (ANOVA) comparing the 3 measurement instruments, actigraphy, sleep diary, and PSG. The ANOVA was performed for each sleep variable, SOL, number of awakenings, WASO, TST, and sleep efficiency. The mean sleep values and standard deviations for each instrument are shown in Table 1. The sleep variables producing significant differences among the means were SOL, F(2, 54) = 6.76, \( p < .002 \) and number of awakenings, F(2, 52) = 143.09, \( p < .001 \). Posthoc, Bonferroni adjusted t-tests on the repeated measures ANOVA for SOL revealed significant differences between sleep diaries and actigraphy, t(55) = -3.12, \( p = .003 \) and sleep diaries and PSG, t(55) = -3.63, \( p = .001 \). Posthoc, Bonferroni-adjusted t-tests on the repeated measures ANOVA for number of awakenings also revealed significant differences between sleep diaries and actigraphy, t(53) = 14.5, \( p < .001 \), and sleep diaries and PSG, t(53) = 14.9, \( p < .001 \). There were no significant differences between actigraphy and PSG on any sleep variable.

Pearson \( r \) values for each pair of instruments on each sleep variable are shown in Table 2 and were performed to determine if changes in the sleep values covary between instruments. Given the large number of \( r \) values calculated, we used \( p < .01 \) as our significance criterion. The \( r \) values for actigraphy and PSG revealed significant correlations on all measured sleep values except SOL. The \( r \) values for sleep diaries and actigraphy revealed significant correlations for TST and sleep efficiency only. Finally, the \( r \) values for sleep diaries and PSG revealed significant correlations for WASO, TST, and sleep efficiency.

Scatterplots of 2 instruments estimating the same variable may be instructive of the pattern of discrepancies across the range of measurement. We constructed such plots for the 5 sleep measures comparing the instruments of primary interest, actigraphy and PSG (see Figures 1-5). In these figures, the X and Y axes were constrained to be the same length, so that instances of perfect agreement between the 2 instruments would produce points on the diagonal and the distance of points from the diagonal would reveal the magnitude and direction of disagreement.

Figure 1 clearly explains the weak correlation between actigraphy and PSG for SOL. Most of the scores congregated within 0 to 30 minutes. This restricted range produced a circular scatterplot pattern and small \( r \) values, as would be expected. Within this range, note that most of the points occur to the left of the diagonal, indicating a bias for actigraphy to underestimate (register better sleep) compared with PSG SOL. Four outliers occur at actigraphy values of about 50, 80, 110, and 160 minutes. The last 3 reverse the bias pattern found with most of the subjects. With high actigraphy estimates of SOL, there is severe bias to overestimate SOL (register poorer sleep).

As predicted by the \( r \) values, Figures 2 to 5 show closer correspondence between actigraphy and PSG than does Figure 1, but informative patterns occur here as well. For number of awakenings (Figure 2), these 2 instruments are sensitive to awakenings and are capable of yielding high counts. A megaphone pattern

---

**Table 1—Sleep Values**

<table>
<thead>
<tr>
<th>Sleep Value</th>
<th>Actigraphy</th>
<th>Polysomnography</th>
<th>Sleep Diaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep-onset latency, min*</td>
<td>17.3(26.8)</td>
<td>17.5(12.6)</td>
<td>32.5(28.6)</td>
</tr>
<tr>
<td>Awakenings, no.*</td>
<td>33.2(15.2)</td>
<td>28.1(12.3)</td>
<td>3.3(1.9)</td>
</tr>
<tr>
<td>Wake after sleep onset, min.</td>
<td>67.2(45.3)</td>
<td>79.9(61.8)</td>
<td>57.4(72.0)</td>
</tr>
<tr>
<td>Total sleep time, min.</td>
<td>380.7(74.7)</td>
<td>366.1(87.1)</td>
<td>378.0(117.0)</td>
</tr>
<tr>
<td>Sleep efficiency, %</td>
<td>79.8(11.3)</td>
<td>77.0(14.1)</td>
<td>78.2(17.3)</td>
</tr>
</tbody>
</table>

Data are presented as mean ± SD.

* \( p < .01 \). For the measures that were significantly different across instruments, common superscripts denote no significant difference between means and disagreement between superscripts denote significant differences between means.

---

**Figure 1—A plot of actigraphy and polysomnography measures of minutes of sleep-onset latency (SOL). Each point represents the actigraphy and polysomnography estimate for 1 subject.**
can be discerned in the 20- to 60-count range, indicating that the magnitude of disagreement is correlated with the number of awakenings. Also, points are more densely distributed to the right of the diagonal, indicative of an actigraphy bias to overestimate number of awakenings compared with PSG. Increasingly larger disagreement between actigraphy and PSG can also be found with WASO starting at about the 50-minute mark, see Figure 3. At this juncture, actigraphy has a bias to underestimate WASO (see the majority of points congregating to the left of the diagonal with actigraphy WASO > 50). TST and sleep efficiency, Figures 4 and 5, show the most consistency across the range of measurement. Both these measures exhibit a mild to moderate bias for actigraphy to overestimate compared with PSG, as shown by a greater number of points to the right of the diagonal.

The Roles of Age, Sex, and Presence or Absence of Hypnotics

There were insufficient numbers of subjects to consider interactions between age, sex, and hypnotic use. Analyses were limited to main-effects tests. For each of the 5 sleep values, we computed a discrepancy score for each participant between each pair of instruments: actigraphy and PSG, actigraphy and sleep diary, and sleep diary and PSG, yielding 15 discrepancy scores.

Age

We correlated age with each of the 5 discrepancy scores within each of the 3 groups to determine if sleep-instrument agreement is related to age. Bonferroni adjustments were appropriately employed. Age was not significantly correlated with any discrepancy score.

Sex

We computed a Bonferroni-adjusted t-test on sex for each discrepancy score to determine if sleep-instrument agreement was related to sex. Although women had lower discrepancies on 14 of the 15 measures than men (the discrepancy between actigraphy and sleep diaries on WASO being the exception), there were no significant differences between men and women for any of the discrepancy scores.

Table 2—Pearson Correlations

<table>
<thead>
<tr>
<th></th>
<th>SOL A</th>
<th>SOL P</th>
<th>NWAK A</th>
<th>NWAK P</th>
<th>WASO A</th>
<th>WASO P</th>
<th>TST A</th>
<th>TST P</th>
<th>SE A</th>
<th>SE P</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOL P</td>
<td></td>
<td>.30*</td>
<td></td>
<td>.13</td>
<td></td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOL D</td>
<td></td>
<td></td>
<td></td>
<td>.49**</td>
<td></td>
<td>.07</td>
<td>.01</td>
<td>.48**</td>
<td>.29*</td>
<td>.46**</td>
</tr>
<tr>
<td>NWAK P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.70**</td>
<td>.59**</td>
<td>.43**</td>
<td>.36**</td>
<td>.48**</td>
<td></td>
</tr>
<tr>
<td>NWAK D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASO P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASO D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TST P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TST D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SE D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.
A, actigraphy; P, PSG; D, sleep diaries. SOL, sleep onset latency; NWAK, number of awakenings; WASO, wake after sleep onset; TST, total sleep time; SE, sleep efficiency.
Presence or Absence of Hypnotics

Of the 23 hypnotic users in the sample, only 11 used their hypnotic the night of testing, and these analyses considered this smaller group only. Given the small number of subjects whose data were available for these analyses, we focused on the comparison of greatest interest, discrepancy scores between actigraphy and PSG. Nine of 11 subjects took benzodiazepines or nonbenzodiazepine agonists, 1 took an antidepressant, and 1 took an over-the-counter medication. There were no significant differences between the discrepancy between actigraphy and PSG on any sleep value for the hypnotic-using group (those who used hypnotic medication on the night of the sleep study) than from the non–hypnotic-using group. Because the number of subjects whose data were available for these tests was so small, we computed effect sizes to explore the possibility of nonsignificant results due to low power. Using Cohen’s standards for $d$, the effect size for number of awakenings was trivial, $d = -0.09$. However, SOL, $d = -0.46$, approached a medium effect size, and WASO, $d = -0.74$, TST, $d = 0.80$, and sleep efficiency, $d = 0.78$, approached or met the criterion for a large effect size, a $d$ of 0.8, and, in every case, the actigraphy-PSG discrepancy for hypnotic users reflected greater overscoring of sleep by actigraphy compared with the scoring of the sleep of the subjects who did not use hypnotics.

DISCUSSION

Actigraphy measures number of awakenings, WASO, TST, and sleep efficiency with a degree of accuracy acceptable for insomnia clinical evaluation. However, for SOL, actigraphy does not mirror PSG consistently and is, therefore, not a valid measure of this sleep value. Strong correlations ($p < .01$) between actigraphy and PSG on all sleep measures, except SOL, show considerable consistency in the devices across participants. A lack of significant differences on any mean measure of sleep between the 2 instruments further points to the accuracy of actigraphy as a sleep measure for people with insomnia. Scatterplots revealed an actigraphy bias to underestimate or overestimate particular sleep variables, compared with PSG, and this pattern was sometimes unevenly distributed across the range of values.

Neither age nor sex affected the discrepancies between actigraphy and PSG. Though we did not obtain significant differences in actigraphy-PSG discrepancies in hypnotic users compared with those subjects who did not use hypnotics, only 11 participants were available for these underpowered analyses. Consideration of effect sizes suggests that hypnotics cause actigraphy to overscore sleep. These findings—in combination with the knowledge that hypnotic use decreases nocturnal motility $^{17}$—supports the idea that hypnotic use influences the ability of actigraphy to differentiate sleep and wake.

Actigraphy performed well on 4 of 5 sleep measures, but we did not achieve the robustness of findings found in validation studies with subjects without insomnia. For example, as stated in the introduction, these studies routinely observe correlations in the $0.8$ and $0.9$ range, and, though correlations with PSG in the current study were generally significant, the magnitude of effect was more moderate. There is room for improvement for actigraphy tracking of PSG in people with insomnia.

This study was powered a priori to detect large correlations ($r > .8$) or large differences ($d = 0.8$) among the 3 instruments for measuring the 5 sleep variables. As a result, the observed nonsignificant differences comparing means could be the result of low power rather than truly small differences. To check this possibility, we performed a series of posthoc power analyses for the pairwise differences among the 3 instruments for the 5 sleep variables. Those power analyses confirmed that the study was large enough to detect even medium-sized differences ($d = 0.5$) for 4 of the 5 sleep variables, though not for SOL. Those analyses also confirmed that the study was not large enough to detect smaller differences ($d < 0.4$) for any of the 5 sleep variables. We conclude from these posthoc power analyses that observed nonsignificant differences for WASO, TST, and sleep efficiency were not likely
due to inadequate power.

The nonsignificant correlation between actigraphy and PSG on SOL indicates random rather than systematic differences between the 2 instruments on this sleep value, pointing to the need for a more accurate actigraphic algorithm for measuring the SOL of people with insomnia. Because of the importance of SOL in insomnia research, this advance is critical to the aptitude of using actigraphy as an independent measure of sleep in people with insomnia.

The findings further indicate superior approximation to PSG of actigraphy over sleep diaries. Sleep diaries and PSG do not vary consistently across participants on SOL or number of awakenings. Furthermore, posthoc paired t-tests of the repeated-measures ANOVA indicated that, on SOL and number of awakenings, sleep diaries and PSG are discrepant, whereas actigraphy and PSG are not. Given that sleep diaries are subjective measures, the significant difference on number of awakenings is not surprising; recalling nocturnal awakenings is different from evaluating 30-second epochs. It may be argued that the great number of nighttime awakenings recorded by PSG and actigraphy do not register experimentally, are often very brief, and are of limited clinical significance. The common finding that people with insomnia self-report more WASO and poorer TST and sleep efficiency than are recorded by PSG was not replicated here, see Table 1. We cannot explain this anomaly with certainty. One possible explanation is that PSG data on insomniacs are often in the nonpathologic range. Our sample slept poorly in the lab, and this may have attenuated the discrepancy between these 2 measures.

One common component of all validation studies of actigraphy on people with insomnia is a comparison of TST. In our study, on average, actigraphy overestimated TST by 14 minutes. This discrepancy is less than the discrepancies in the 4 existing studies, which range from 48 minutes to over an hour. Understanding which variables contributed to the high degree of agreement between actigraphy and PSG of TST in our study is difficult due to the number of differences between all 5 of the validation studies.

The analytic differences between this study and earlier validation studies of people with insomnia represent, in part, the advantages of this study. For example, we used a varied group of people with insomnia on a number of characteristics. We did not exclude comorbid insomnia or people with insomnia who were taking hypnotic medication. All age ranges and both sexes were well represented. (Though, it should be noted that there were not enough African Americans in our study to evaluate differences that may arise due to ethnicity.) The sample allows for results that can be generalized to the population of people with insomnia and also allows for evaluation of subgroups, which helps clarify the utility of actigraphy according to individual attributes such as whether or not a person with insomnia takes hypnotic medication. This is the largest sample of people with insomnia to date used to validate actigraphy. Most importantly, additional variation between our study and existing studies includes different instruments and different scoring algorithms.

For the purposes of standardization and future cross-study analyses, it is critically important to emphasize that, in our study, “actigraphy” is represented by Mini-Mitter AW64 Actiwatch, analyzed using Actiware Sleep v. 3.3 algorithms at high sensitivity. Actigraphic performance is a function of the instrument sensitivity and the algorithm used to translate the data. Validation findings are linked to each unique combination of these instrument features and cannot be expected to generalize to other instruments or the same instrument utilizing alternative algorithms. As would be expected, data already exist showing that different algorithms applied to the same data produce different findings.2,5,12

Because sleep diaries and PSG measure different phenomena of sleep, so do actigraphy and PSG. Therefore, it is expected that actigraphy would not perfectly mirror PSG. Actigraphy does not measure physiologic sleep, as does PSG; it measures movement. Because PSG is the gold-standard measure and because brainwaves are, thus, the optimal measurable components of sleep, actigraphy assumes the role of a proxy measure of sleep. Actigraphy, as a proxy measure and under the conditions of this study, performs reasonably well, especially if accommodations or alternative algorithms can be validated to measure SOL in people with insomnia. Insomnia research would benefit from such corrections to actigraphic measures of SOL, as well as from comparison studies of the different actigraphic variables, such as sensitivity thresholds, wrist placement, and definitions of activity counts. Such comparison studies would yield needed standardization procedures for the use of actigraphy.

ACKNOWLEDGEMENTS

This research was supported by grants from the Mini Mitter Company, Inc. and the Methodist Healthcare Foundation.

REFERENCES