Total sleep time obtained from actigraphy versus sleep logs in an academic sleep center and impact on further sleep testing

R. Robert Auger1,2
Ranji Varghese1
Michael H Silber1,3
Nancy L Slocumb1
1Center for Sleep Medicine,
2Department of Psychiatry
and Psychology, 3Department
of Neurology, Mayo Clinic College
of Medicine, Rochester, MN, USA

Background: While actigraphy has been deemed ideal for the longitudinal assessment of total sleep time (TST) by select groups, endorsement has not been universal and reimbursement is lacking, preventing its widespread use in clinical practice. This study compares longitudinal TST data obtained by actigraphy and logs preceding a clinical evaluation, and secondarily ascertains whether longitudinal TST impacts clinicians’ decisions to proceed with further sleep testing.

Methods: This was a retrospective, consecutive chart review spanning about 4 months in an academic sleep center. Eighty-four patients wore actigraphs in anticipation of clinical evaluations. Concomitant completion of sleep logs is routinely requested in this setting. Longitudinal TST data available in complete form was reviewed in a blinded fashion among a subset of these patients. A review of text from clinical notes of an expanded cohort with complete actigraphy data (regardless of the degree of completion of logs) enabled determination of the frequency and rationale for cancellation of prescheduled sleep testing.

Results: Of 84 actigraphy recordings, 90% produced complete data, and 30% produced fully completed logs. Among the subset with both available in complete form, significant mean TST differences were observed on weekends (7.06 ± 2.18 hours versus 8.30 ± 1.93 hours, \( P = 0.009 \)), but not on weekdays (7.38 ± 1.97 hours versus 7.72 ± 1.62 hours, \( P = 0.450 \)) for actigraphy and logs, respectively. Further analyses revealed poor agreement between the two measures, with predominantly increased TST estimation with logs. Among those with complete actigraphy data (±logs), testing was cancelled in 11 (15%), eight of whom (73%) presented with hypersomnia and three of whom (27%) presented with insomnia. Determination of insufficient sleep time was cited as the primary reason for cancellation (64%).

Conclusion: Actigraphy and sleep logs provided discrepant mean TST data on weekends only, and the latter predominantly estimated increased TST. Actigraphy was completed more reliably than logs. Longitudinal TST information influenced clinicians’ decisions to proceed with further testing, particularly among patients presenting with hypersomnia.

Keywords: sleep diaries, polysomnography, multiple sleep latency testing

Introduction

Since the publication of the American Sleep Disorders Association practice parameters pertaining to the use of actigraphy in 1995,1 the device’s prominence in the field has increased considerably. Actigraphically-derived data are included among the International Classification of Sleep Disorders, Second Edition (ICSD-2) diagnostic criteria for paradoxical insomnia, idiopathic hypersomnia, behaviorally-induced insufficient sleep syndrome and, with the exception of jet lag, all of the circadian rhythm sleep disorders.2 Reflecting its particular utility as a longitudinal tool, the ICSD-2 specifically cites actigraphy as a means of documenting stability of the sleep-wake schedule.
in virtually all of the circadian rhythm sleep disorders, and as a means of assurance of adequate total sleep time (TST) preceding interpretation of the multiple sleep latency test (MSLT), during the course of evaluations for narcolepsy and idiopathic hypersomnia. Evidence appears to support these ICSD recommendations, as the most recent American Academy of Sleep Medicine (AASM) practice parameters endorsed actigraphy as a means of characterizing sleep disturbances among individuals with insomnia, and as a method of ensuring “stable sleep patterns and adequate sleep duration prior to polysomnography (PSG) and MSLT”.3

Despite this and other AASM statements expressing superiority of actigraphy versus log-derived data, this endorsement is not ubiquitous within sleep medicine. Indeed, the ICSD-2 lists sleep logs as an equivalent alternative to actigraphy, and logs alone are recommended within the MSLT practice parameters as a means of assessing the sleep-wake schedule during the time period preceding the evaluation.5 Third-party payers have reacted accordingly. Although a Current Procedural Terminology Category I code was established for actigraphy in 2009, reimbursement for its use remains variable or nonexistent. As such, many sleep centers do not include it among their tools of assessment.

An historical review of the adult literature provides a useful backdrop. Numerous studies have validated actigraphically-derived sleep parameters with those of PSG (the reference standard).3,4,6,7 In their 2007 paper, the AASM Standards of Practice Committee concluded that actigraphy was highly correlated with PSG in those reports rated at higher evidence levels. Subsequent to analyzing Pearson r values for TST within various studies (predominantly among patient populations), the group calculated an average value of 0.71.3 There are comparatively few investigations that enable concomitant comparisons of solely log-derived data with PSG, particularly as these subjective reports often assist in actigraphic determinations of the sleep period (eg, references).8–13 Four such studies were identified, however, with varying use of statistical methods; all involved subjects with insomnia, recruited from either the community or clinical settings.14–17

In an early validation study by Hauri and Wisbey, 36 medication-free subjects (23 females, mean age 45 years) with various insomnia diagnoses were recruited from the community. Over a period of three laboratory nights, mean sleep duration obtained from PSG, logs, and actigraphy (type unknown, automatically scored with actigraphic scoring analysis)18 were compared.14 Results differed depending upon the diagnostic subgroup. Among those with insomnia comorbid with a mental disorder (n = 13) and psychophysiologic insomnia (n = 10), actigraphy reported longer mean TST than the subjective reports, and PSG held a middle ground, not statistically different from either method. For those with sleep state misperception (n = 8), no statistically significant differences in mean TST were observed between actigraphy-derived and log-derived data, but both were significantly lower than the PSG-determined mean TST. For all three of these subgroups, the actigraph best approximated PSG data in terms of absolute errors, such that inaccuracies reported by the device were only about half that provided by logs. The numbers were too small to evaluate differences between the remaining diagnostic subgroups (n = 5).

Chambers’ subsequent allegations of methodologic errors within the report by Hauri and Wisbey led to publication of his reanalysis of these data, which created controversy with respect to actigraphy’s purported superiority over logs in research and clinical settings.19 Chambers focused on the correlation of TST derived from each measure with PSG (not reported by Hauri and Wisbey), and found higher values associated with logs, such that they predicted nearly three times as much variance in PSG-TST than that predicted by actigraphy (r² = 0.36 and 0.13, respectively), despite logs’ consistent TST underestimation. Chambers nevertheless also reported a superior night-to-night (or within-subject) correlation of 0.81 between actigraph-derived and PSG-derived TST, suggesting a consistent device estimation error from night to night.

Vallieres and Morin subsequently performed a validation study that took place over a similar number of PSG laboratory nights, specifically investigating the use of actigraphy (Individual Monitoring Systems Inc, Arnold, MD, USA, automatically scored with software version 3.15a) for the assessment of primary insomnia among a recruited medication-free population (n = 17, 10 females, 41.6 ± 5.7 years [mean ± standard deviation]). Mean TST was not reported. Spearman analyses revealed a correlation of 0.71 between PSG-derived and actigraphy-derived TST, while the correlation between logs and PSG was insignificant.15

The two most recently published studies are the largest to date. One compared overnight sleep assessments accrued from actigraphy (AW64 Actiwatch™ [Mini Mitter, Bend, OR, USA], automatically scored with Actiware software version 5.0, at 30-second epochs, medium sensitivity settings, worn on nondominant wrist) and logs to PSG among depressed insomniacs recruited for a clinical trial investigating the effects of eszopiclone (n = 54, 36 females,
There were no significant differences with respect to mean TST recorded by PSG and actigraphy, and a moderate correlation was observed ($r = 0.54$). In contrast, significant differences were found between logs and PSG for all measured sleep variables, with underestimation of TST by a mean of $54.5 \pm 14.1$ minutes. Finally, Lichstein et al studied medicated and nonmedicated subjects with both primary and comorbid insomnia ($n = 57$, 31 females, aged 21–87 years). When comparing PSG-derived TST during one laboratory night to that estimated by logs and actigraphy (AW64 Actiwatch, automatically scored with Actiware Sleep version 3.3 software, at 30-second epochs, high sensitivity settings, worn on the dominant wrist), no significant mean differences were observed, a finding that persisted once results were adjusted for age and gender. Pearson correlations between actigraphy-PSG and log-PSG were both significant at 0.70 and 0.59, respectively.

In summary, validation studies for logs and actigraphy (the two most recent of which were published subsequent to publication of the current AASM guidelines) have shown either no mean TST differences or heightened accuracy of actigraphy. Correlation analyses reveal superiority of actigraphy in all but one study. However, all of these studies were performed with insomnia patients in the laboratory, precluding generalization of the results to longitudinal assessments in field settings, where subjects complete logs less reliably.

Only one retrospective study (also published subsequent to the 2007 AASM practice parameters) was identified that compared log-derived and actigraphy-derived sleep parameters among patients presenting with hypersomnia ($n = 54$, seven females, aged 30.7 ± 10.4 years). Bradshaw et al compared objective (Precision Control Design Octagonal Sleep Scoring Watch, automatically scored with ActionW version 2.4.20 software, at 60-second epochs, worn on the nondominant wrist) and subjective TST measured during a 2-week period preceding PSG/MSLT. Log estimates exceeded those predicted by actigraphy by a mean of 1.43 ± 1.31 hours nightly. Significant correlations were observed between actigraphy-derived mean TST and MSLT mean sleep latency ($r = -0.0515$), but no such correlations were observed with log-derived TST. Significant TST correlations were also described between overnight PSG and actigraphy in the laboratory ($r = 0.5469$), but no such log-PSG analyses were performed, precluding comparisons.

AASM practice guidelines (published prior to several of the studies cited above) presently endorse the use of actigraphy for longitudinal sleep assessments at the lowest evidence level, and its use among hypersomnolent patients (particularly in preparation for MSLT) is recommended as a specific area for future research. In recognition of this deficit, we compared actigraphy-derived and log-derived TST data among patients awaiting an initial sleep evaluation (including the extent to which information was available in complete form). In addition, we ascertained whether longitudinal TST information in general impacted clinicians’ decisions to proceed with further testing.

Materials and methods
The Mayo Center for Sleep Medicine (CSM) receives a large number of referrals from outside facilities. Depending upon the nature of the referring provider’s question (and the impression of the board-certified sleep physician who reviews the accompanying medical records), an actigraph may be mailed to the patient in anticipation of the initial appointment, typically for a duration of at least 7 days. Written instructions request that it be worn on the nondominant wrist “like a wristwatch,” and removed only when water exposure is anticipated. The serial completion of sleep logs is requested concomitantly, with specific instructions to record requested information each morning. Logs are designed to obtain longitudinal information regarding the subjective timing of sleep onset and offset, wakefulness after sleep onset, and TST, among other variables. The sole completion of logs is not routinely requested (ie, logs without accompanying actigraphy).

After approval by the Mayo Foundation Institutional Review Board, consecutive actigraphy data collected during an approximately 4-month span (determined by reviewing the billing record) were retrospectively reviewed. Inclusion criteria stipulated that the device was worn during a contiguous time period preceding the initial sleep consultation (minimum of 7 days). In cases where both log and actigraphy data were available in complete form, TST comparisons were performed by the same individual (RV).

During the time period of review, the CSM used AW64 Actiwatches, automatically scored with Actiware Sleep version 3.4 software, at 60-second epochs, and medium sensitivity settings. When logs are completed, they are used by CSM technologists both to assist with artifact rejection and to corroborate automated determinations of sleep onset and offset, which may result in manual adjustments to the software-determined values (methods reviewed).
Once the primary goal of our study was to compare TST derived solely from actigraphy and logs, the actigraph tracing and subjective data were reviewed separately.

Actigraphy-derived TST during a 24-hour period was determined by utilizing the visual scoring method described by Kripke et al. Their paper describes the marked visual contrast between motion artifact when a subject is awake (with almost continual movement of the wrists) versus the brief bursts of movement observed when a subject is asleep. Removal of the device is also readily discerned with this method. In a study that compared visual actigraphy scoring with PSG sleep/wake scoring, agreements of 91.6% and 96.3% were found for patients (predominantly with comorbid insomnia) and controls, respectively. Also using PSG as a reference standard, Hauri and Wisbey described superiority of manual scoring to four separate automated methods, among a population of insomniacs.

To standardize the calculation of actigraphically-determined sleep utilizing the visual method, a dedicated ruler was used. Cumulative sleep “length” was measured during a 24-hour period for each patient, and measurements were then converted to time units (the 24-hour time period was displayed as one continuous line on the actigraph tracings). Log-derived subjective TST was determined by subjects’ daily response to the question “How many hours did you sleep last night?” Analyses comprised both a weekday (Wednesday) and a weekend day (Saturday) to determine whether discrepancies existed. If more than one of these specified days was contained within a recording, the second day was chosen for analysis.

Data were analyzed and compared using JMP (SAS Institute, Cary, NC, USA) and MedCalc (Broekstraat, Belgium) software. Means were compared by matched pair analysis (confidence interval 95%), with \( P < 0.05 \) considered to be statistically significant. Values are summarized as the mean ± standard deviation. A subsequent analysis was performed (Lin’s concordance correlation coefficient) to ascertain agreement. Although there are other valid methods of agreement analysis, we felt that a Lin’s concordance correlation coefficient plot would provide the best visual impression.

Finally, a chart review was conducted to determine the frequency of cancellation of prescheduled PSG/MSLT as a result of acquired longitudinal TST data. Clinical notes were reviewed in detail, and an explicit statement by the provider was required to establish a correlation. Since we did not seek a comparison of the two assessment tools for this portion of the study, and since complete actigraphy data were available more commonly than complete log data, a larger number of subjects were included for this analysis. As such, we did not assess TST information obtained from sleep logs alone, as they are not routinely ordered without accompanying actigraphy within the CSM practice.

**Results**

Of the 84 patients initially identified, actigraphy data were noninterpretable (eg, due to device malfunction) among eight (10%), and 23 (27%) did not complete sleep logs to any degree. Complete sleep log and actigraphy data for both weekend and weekdays were available for 25 subjects (60% female, mean age 32.5 ± 17.3 years), with mean durations of observation of 14.4 ± 4.8 and 14.1 ± 4.7 days (actigraphy and logs, respectively). This subsample comprised the comparison cohort. Significant mean differences in TST were observed on weekends (actigraphy 7.06 ± 2.18 hours versus logs 8.30 ± 1.93 hours, \( P = 0.009 \)) but not on weekdays (actigraphy 7.38 ± 1.97 hours versus logs 7.72 ± 1.62 hours, \( P = 0.450 \)). For the agreement analyses, Lin’s concordance correlation coefficient was 0.257 (95% confidence interval, −0.1296 to 0.5761) for weekdays and 0.364 (95% confidence interval 0.03737–0.6203) for weekends. In the majority of instances, logs overestimated TST (60% and 68% for weekdays and weekends, respectively; Figures 1 and 2).

The analysis to determine the frequency of cancellation of anticipated testing included all subjects with complete

**Figure 1** Plot depicting agreement between TST in hours assessed by sleep logs versus ACTI on weekdays.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concordance correlation coefficient</td>
<td>0.2572</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>−0.1296 to 0.5761</td>
</tr>
<tr>
<td>Pearson ρ (precision)</td>
<td>0.2670</td>
</tr>
<tr>
<td>Bias correction factor ( C_0 ) (accuracy)</td>
<td>0.9635</td>
</tr>
</tbody>
</table>

**Abbreviations:** TST, total sleep time; ACTI, actigraphy.
Discussion

Our primary goal was to compare longitudinal TST assessments obtained by actigraphy and logs preceding a clinical evaluation. While the former has been deemed ideal (and perhaps superior) for this purpose by select groups within the sleep community,\(^1,4\) this endorsement has not been universal,\(^2,5,19\) and reimbursement is lacking, preventing the widespread use of actigraphy in clinical practice. While two prospective validation studies (utilizing PSG as a reference standard) demonstrated no significant mean TST differences derived from the two measures,\(^14,17\) a separate study described greater accuracy of actigraphy.\(^16\) In contrast with these short-duration investigations (which solely involved subjects with insomnia),\(^14,16,17\) our study was comprised primarily of patients complaining of hypersomnia undergoing longitudinal assessments (\(\geq 7\) days), was retrospective in nature, and did not use a PSG reference standard.

Our investigation is therefore most comparable with the aforementioned study by Bradshaw et al,\(^21\) although their cohort was larger, male-predominant, military-based, devoid of agreement analyses between actigraphy and logs, and included only patients with complaints of excessive daytime sleepiness. In addition, their group compared cumulative mean TST during the entire period preceding the evaluation, while we calculated means on the basis of a preselected weekday and weekend day. While daily measurements may have provided more accurate information, this was deemed to be overly labor-intensive, as our goal was to compare sole use of each measure in a longitudinal setting, necessitating painstaking visual scoring (and ruler measurements) of actigraphy tracings. Nevertheless, our reported weekend discrepancies in the two measures were quite similar to the discrepancies reported by Bradshaw et al,\(^21\) with increased log estimation of mean TST by approximately 75 minutes. Finally, while Bradshaw et al did not differentiate between assessments on weekdays and weekends, we identified significant differences only for the latter. The reason for this discrepancy is not clear, but strict routines enforced by work or academic schedules during weekdays could enhance the accuracy of subjective assessments. Regardless of the ultimate explanation, this finding should alert clinicians to the increased likelihood of discordance between subjective and objective TST reports during periods of relatively unrestricted sleep scheduling.

Perhaps a more important finding from our study was one of poor agreement between actigraphy and logs (a more accurate measure of convergence) regardless of the day of the week, with a predominance of increased estimation of subjective TST compared with that estimated by actigraphy. This proclivity for increased subjective longitudinal TST in field settings is a matter of concern, particularly in the setting of patient evaluations for hypersomnia. Bradshaw et al correlated actigraphy-derived longitudinal TST (but not log data) with decreased MSLT mean sleep latency, raising concerns that sole use of subjective reports could result in the misdiagnosis of a primary disorder of hypersomnia among those with chronic partial sleep deprivation.\(^21\) Previous laboratory-based studies describe similar effects with polysomnographically-measured acute and chronic sleep restriction,\(^26,27\) also reviewed by Arand et al.\(^28\) Indeed, we observed that prescheduled testing (PSG and/or MSLT) was most commonly cancelled among patients complaining of hypersomnia, due to clinicians’ impressions of insufficient

---

**Figure 2** Plot depicting agreement between TST in hours assessed by sleep logs versus ACTI on weekends.

**Abbreviations:** TST, total sleep time; ACTI, actigraphy.
sleep, as determined by actigraphy. However, since we were not able to assess whether decision-making was similarly affected when preceding TST was assessed solely by logs (the CSM practice does not operate as such), we cannot definitively claim superiority of actigraphically-obtained data for this purpose. Moreover, our methodology did not allow us to differentiate the degree to which the subsample of those who possessed both complete log and actigraphy data influenced the reported frequency of testing cancellation, as one might infer that these individuals were most conscien
tious, and perhaps more likely to obtain sufficient sleep. Future studies that compare dual versus partial “completers” of logs and actigraphy may reveal important demographic or other differences that could aid the discussion regarding the relative strengths of the assessment tools for the acquisition of longitudinal TST in the clinical arena.

Additional unique data from our study included the comparison of interpretable actigraphy and log information. Among those patients initially reviewed for study inclusion, 10% were deemed ineligible due to actigraphy malfunction, and 27% did not complete logs to any degree. We found only one other study that reported the frequency of device malfunction. In an investigation of predominantly male shift workers, 27% of actigraphy information (AMA model 32) was unavailable due to a variety of “difficulties,” including short battery life. Insufficient literature is available to determine the degree to which device failure depends on the location and duration of use, the type of device, the population of assessment, or associated clinical practices. Similarly, we are not aware of other studies that specifically report failure of log-derived TST. Sadeh noted that logs become less complete with increased duration of assessment. Other investigators report anecdotal observations of patients completing “days or weeks” of log entries while sitting in the waiting room. Since we compared actigraphy-derived and log-derived TST only among patients who provided both in complete form, we do not know whether partially completed logs exhibit a heightened degree of divergence, a scenario likely more frequently encountered by providers. Finally, while scheduled reminders or prompts might improve the quality of sleep log data, this may be impractical in a setting of longitudinal assessment, particularly in the pre-evaluation scenario described herein.

In addition to the limitations already described, other potential shortcomings warrant mention. As noted above, our study was retrospective and did not include a gold standard PSG reference. Even if future longitudinal studies were prospective, however, the use of continual PSG would presumably be cost-prohibitive and, unless ambulatory recordings were implemented, contrary to the field setting of interest. Moreover, subjects studied in the laboratory may complete logs more reliably than those studied in field settings for an extended period of time, potentially resulting in a mischaracterization of log accuracy. Our sole inclusion of prescheduled patients represents a selection bias, and we therefore are uncertain as to whether similar results would have been found if actigraphy and logs were ordered subsequent to clinicians meeting with patients. Relatedly, more rigorous instructions for completing logs in the setting of real time or pre-scheduled evaluations could result in a greater degree of agreement with actigraphy. Therefore, although the method of ascertainment described is consistent with a large portion of our CSM practice, it cannot be generalized to other sleep medicine practices.

Notwithstanding these limitations, our study fills a significant research gap with respect to the comparative use of actigraphy and logs for longitudinal TST assessments in clinical populations/field settings, particularly with respect to clinicians’ decisions to proceed with PSG/MSLT among those complaining of hypersomnia. Within the current CSM practice, it would appear that log estimations of TST are higher than those reported by actigraphy, particularly in the setting of an unrestricted sleep/wake schedule. In addition, actigraphy data are produced more reliably than logs.

In conclusion, while an AASM review suggested that logs and actigraphy be used in a complementary fashion, this is a luxury many cannot afford, due to current reimbursement practices. Although it might be expected that the objective and unbiased nature of data produced by the actigraph would necessarily be more accurate than those yielded by subjective assessment techniques during longitudinal assessments, this requires further rigorous confirmation. Proponents of actigraphy bear the burden of proof that it is superior to its less expensive alternative. Future studies should prospectively compare longitudinal actigraphy and log data among diverse patient populations, with a particular focus on evaluation of patients complaining of hypersomnia. Methodologic descriptions should include technical details for the reporting of actigraphy data (suggested guidelines are described in detail), and investigations should include measurements of impacts on objective outcomes (PSG/MSLT) and clinical decision-making. The results of our study, the study by Bradshaw et al, and future studies could impact reimbursement practices, and may result in revised MSLT practice parameters, with more detailed requirements.
for sleep logs and/or actigraphy during specified time periods prior to the formal sleep evaluation.

Acknowledgment
The authors thank Lori L Solmonson and Daniel L Herold for technical assistance.

Disclosure
The authors report no conflicts of interest in this work. They report no grant support for this manuscript.

References