

Issues of Validity in Actigraphic Sleep Assessment

Warren W. Tryon, PhD

Fordham University

Objectives: The Standards of Practice Committee of the American Sleep Disorders Association has supported the use of actigraphy in the assessment of sleep disorders. Pollak et al disagree. The objective of this paper is to identify and critically evaluate several theoretic and methodologic issues that are central to these divergent views regarding the valid use of actigraphy for sleep assessment.

Design: Critical review, analysis, and comment.

Setting: N/A

Patients/Participants: N/A

Interventions: N/A

Measurements and Results: N/A.

Conclusions: (1) Coefficients of the validity of actigraphy exceed those associated with common medical tests and the best psychological tests. (2) Reasons why actigraphy should be held to a substantially higher empirical standard than common medical tests and the best psychological tests have yet to be advanced. (3) Differences between actigraphy and polysomnography are not random and can be reduced. (3a) Sleep onset is a gradual rather than a discrete process. Actigraphy keys on an earlier

phase of the sleep-onset process than does polysomnography, resulting in systematic rather than random differences. (3b) A sleep switch device can be used to substantially increase the accuracy of sleep-onset times. (3c) The residual unreliability of polysomnographic data explains a portion of the differences between actigraphy and polysomnography. Actigraphy cannot be expected to agree more completely with polysomnography than polysomnography does with itself. (4) Complete agreement between actigraphy and polysomnography has been presumed, but achieving such a limit is theoretically impossible. Some lower maximum agreement limit should be identified. (5) Conclusions that actigraphy is not an accurate sleep-wake indicator and that it is inappropriate to infer sleep from actigraphy data are not supported by the findings. Conclusions reached, including caveats, by the Standards of Practice Committee remain supported.

Key Words: actigraphy; polysomnography; sleep-wake validity; evaluation studies; sleep; monitoring, ambulatory; monitoring, physiologic

Citation: Tryon WW. Issues of validity in actigraphic sleep assessment. *SLEEP* 2004;27(1):158-65.

INTRODUCTION

ACTIGRAPHY HAS BEEN USED TO STUDY SLEEP FOR AT LEAST THE PAST 30 YEARS, SINCE KUPFER ET AL¹ REPORTED SIGNIFICANT AND SUBSTANTIAL CORRELATIONS BETWEEN WRIST ACTIVITY AND ELECTROENCEPHALOGRAM (EEG)-MEASURED MOVEMENT AND WAKEFULNESS IN 1972. Sufficient supportive data existed by 1995 to enable the Standards of Practice Committee of the American Sleep Disorders Association² to support the use of actigraphy in evaluating certain aspects of sleep disorders. Representing an opposing view, Pollak et al concluded that their data "... disqualify actigraphy as an accurate sleep-wake indicator"^{3p957} and that "... it does not seem appropriate in 2001 to refer to inactivity defined by wrist actigraphy as "sleep" or to wrist activity as "wakefulness."^{3p965} Were the conclusions of the Standards of Practice Committee of the American Sleep Disorders Association² premature, incorrect, or both premature and incorrect? Are the conclusions drawn by Pollak et al³ fully supported by their data or do they go beyond their data? This article identifies and critically evaluates several theoretic and methodologic issues that are central to the controversy surrounding the use of actigraphy to assess sleep versus wakefulness.

CONCEPT AND DEFINITION OF SLEEP

Operational definitions define concepts in terms of the measurement procedures used to quantify them. However, the domain of the operational definition may not be exactly congruent with the conceptual

domain of interest. For example, one can define intelligence as what an intelligence test measures, but no intelligence test adequately assesses all facets of intelligence. Several different intelligence tests exist, and they are highly intercorrelated, indicating the presence of a common factor, but this common factor is not coextensive with the concept of intelligence. The Wechsler scales for adults and children arguably constitute the industry standard for measuring intelligence, but they are not seen as complete or ideal measures of intelligence. Does this issue pertain to sleep? What is the conceptual domain of sleep? Is this conceptual domain completely measured by polysomnography (PSG)? Are we in complete agreement that sleep is, and only is, what PSG measures? Rechtschaffen and Kales⁴ may have standardized terminology and techniques and provided a scoring system regarding human sleep, but did they also resolve the conceptual question as to the nature of sleep? If so, then how closely must a proxy sleep measure agree with PSG to be considered useful? Does the degree of agreement differ depending upon measurement purposes, ie, screening versus diagnosis? These questions are addressed below.

EMPIRICAL CONTEXT

Testing

Perhaps the easiest question posed above concerns the degree of relationship that a test and its validating criterion must have in order to be considered useful in clinical and research contexts. Criterion validity is established when a test correlates significantly with an external criterion.⁵ Criterion studies of actigraphic sleep-scoring validity have almost entirely been conducted against PSG. Construct validity is established when a test correlates significantly with another measure of the same phenomenon,⁵ or when test scores bear a statistically significant relationship to other variables in a predicted way, or when test scores track therapeutic changes in the predicted way. The validity of psychological tests is no longer considered to be a fixed property of the test but a property of the application of a particular test to a particular objective in a particular context, including the specific participants tested.⁵ Meyer et al⁶ summarized the validity evidence from more than 125 meta-analyses

Disclosure Statement

No significant financial interest/other relationship to disclose.

Submitted for publication November 2002

Accepted for publication September 2003

Address correspondence to: Warren W. Tryon, PhD, Department of Psychology, Fordham University, Bronx, NY 10458-5198; Tel: 718-817-3787; Fax: 718-817-3785; E-mail: wtryon@fordham.edu

covering over 800 data sets regarding the validity of psychological and medical tests. Table 1 presents validity coefficients for selected medical tests reported by Meyer et al.⁶ representing both the least and most valid medical tests reported. The least valid medical procedure reported explains less than 2% of criterion variance, whereas the most valid medical procedure reported explains just over 70% of criterion variance. The best medical tests reported by Meyer et al.⁶ fail to predict at least 30% of criterion variance. That they did not include the most valid medical tests available does not negate the validity coefficients of the accepted medical tests they do report. Intelligence tests are among the most valid psychological tests. Correlations of the Wechsler Adult Intelligence Scale with other well validated intelligence tests such as the Stanford-Binet cluster around .80.^{5p221} Hence, about 64% of the variance in test scores can be explained, leaving approximately 36% of the variance unexplained. Correlations between Kaufman Brief Intelligence Test scores and Wechsler Adult Intelligence Scale full-scale IQ scores range from .61 to .88.⁷ These correlations account for from 44% to 77% of the variance, leaving approximately 56% to 23% of the variance unaccounted for. Hence, 23% to 56% error characterizes some of the most valid psychological tests.

The maximum validity coefficients associated with personality testing stem from instances where a shorter version of a test is correlated with a longer version of the same test. The dominant view of personality structure is that it is composed of the following 5 factors: Neuroticism,

Extraversion, Openness, Agreeableness, and Conscientiousness. The NEO-PI-R is the current gold standard for measuring these 5 factors.⁸ The NEO-FFI is a short form of the NEO-PI-R.⁸ It was developed by

Table 1—Validity coefficients for medical tests

Entry #	Medical Test	r	% variance explained
24	Identifying balance deficits due to vestibular impairment using platform posturography	.13	1.69
42	Prediction of death or myocardial infarction within 1 week of vascular surgery using 4 preoperative cardiac tests	.20	4.00
54	Predicting hip fracture from bone mineral density measurements	.25	6.26
65	Diagnosis of acute appendicitis using C-reactive protein tests	.28	7.84
74	Detecting acute infectious diarrhea using fecal leukocyte results	.30	9.00
80	Detecting breast cancer within 1 year from mammograms	.32	10.24
88	Identifying dental cavities from x-rays	.36	12.96
98	Identifying endometrial cancer in postmenopausal women using endovaginal ultrasound	.39	15.21
102	Detection of prostate cancer in men 60 to 72 years of age by prostate specific antigen	.40	16.00
109	Diagnosis of coronary artery disease by cardiac fluoroscopy	.43	18.49
121	Sensitivity of total serum cholesterol levels due to dietary changes	.50	25.00
133	Identification of deep venous thrombosis using ultrasound	.60	36.00
138	Detection of rheumatoid arthritis using Immunoglobulin-G using antiperinuclear factor scores	.68	46.24
141	Measuring cardiac stroke volume using thoracic impedance scores	.81	65.51
144	Measuring arterial oxygen saturation using finger or ear pulse oximetry	.84	70.56

Adapted from Table 2, Meyer GJ, Finn SE, Eyde LD, et al. Psychological testing and psychological assessment: A review of evidence and issues. *Am Psychologist* 2001;56:128-65. Copyright © 2001 by the American Psychological Association. Adapted with permission. The percentage of variance explained is determined by squaring the correlation coefficient and multiplying by 100. The percentage of variance unexplained is 100% minus the percentage of variance explained.

Table 2—Possible Agreements and Disagreements Regarding Classification when Screening or Assessing with Actigraphy and Polysomnography

Actigraph Determined	PSG Determined	
	Normal	Abnormal
Normal	A) Hit	B) Error
Abnormal	C) Error	D) Hit

PSG refers to polysomnography.

Table 3—Potential Contributions of Actigraphy by Diagnostic Category*

DYSSOMNIAS	Potential Role of Actigraphy
<i>Sleep Disorder Classification</i>	
<u>Intrinsic Sleep Disorders</u>	
Insomnia	Actigraphy needs to be supplemented with a sleep switch in order to identify people who lay quietly awake. ¹⁷
Sleep-state misperception	Patients with sleep-state misperception will claim that they do not sleep at all some nights, and both actigraphy and PSG will find that they do sleep much more than they report, thereby correctly identifying persons with this disorder.
Sleep apnea	May produce actigraphy records scored as wake. Abdominal actigraphy can be used to obtain independent evidence of epochs of violent breathing associated with the termination of apnea.
Periodic Limb Movement Disorder	May produce actigraphy records scored as wake. Ankle actigraphy can be used to obtain independent evidence of periodic limb movements during sleep.
<u>Extrinsic Sleep Disorders</u>	These sleep disorders depend upon contextual information that is not available from polysomnography or actigraphy. The 2 examples presented below illustrate the need to query potential patients about many issues and to obtain sleep-diary information whether one uses polysomnography or actigraphy.
Environmental Sleep Disorder	Can result from excessive noise, temperature or a bed partner with a sleep disorder, or attempting to sleep in a strange place such as a hospital.
Inadequate Sleep Hygiene	Entails irregular sleep habits, excessive napping, and sleep-incompatible behaviors.
<u>Circadian Rhythm Sleep Disorders</u>	Mistlberger and Rusak ⁵³ use activity-level figures to present basic scientific findings regarding circadian rhythms in mammals.
Time-zone change (jet lag)	Continuous actigraphy will track the activity-rest cycle and reveal changes associated with time-zone travel. This effect was nicely illustrated by Gruen's 28-day westerly trip around the world. ⁵⁵
Shift-work sleep disorder	Continuous actigraphy will track the activity-rest cycle and reveal changes associated with shift-work changes.
Irregular sleep-wake pattern	Continuous actigraphy will track the multiple cycles of sleep and wake associated with temporally disorganized and variable episodes of sleep and waking behavior. Ancoli-Israel et al ⁵⁴ used actigraphy to identify waking and sleeping 2 to 3 times per hour in nursing-home patients.
Delayed sleep phase syndrome	Continuous actigraphy will track the delay of the major sleep period in relation to the desired sleep time.
Advanced sleep phase syndrome	Continuous actigraphy will track the advance of the major sleep period in relation to the desired sleep time.
Non-24-hour sleep-wake disorder	Continuous actigraphy will track the consistent 1- to 2-hour daily delays in sleep onset and wake times.
<u>PARASOMNIAS</u>	
<u>Arousal Disorders</u>	
Sleepwalking	Sleepwalking is likely to be detected by concurrent waist actigraphy if not by wrist actigraphy.
Sleep terrors	Sleep terrors and nightmares will likely produce wakefulness sufficient to be detected by actigraphy.
<u>Sleep-wake Transition Disorders</u>	Actigraphy is not informative for these disorders.
<u>Parasomnias Usually Associated with REM Sleep</u>	Actigraphy cannot measure sleep stages.
<u>Nightmares</u>	Actigraphy can document resulting awakening to within 30 seconds and can assess how long it took for the people to recompose themselves for sleep.

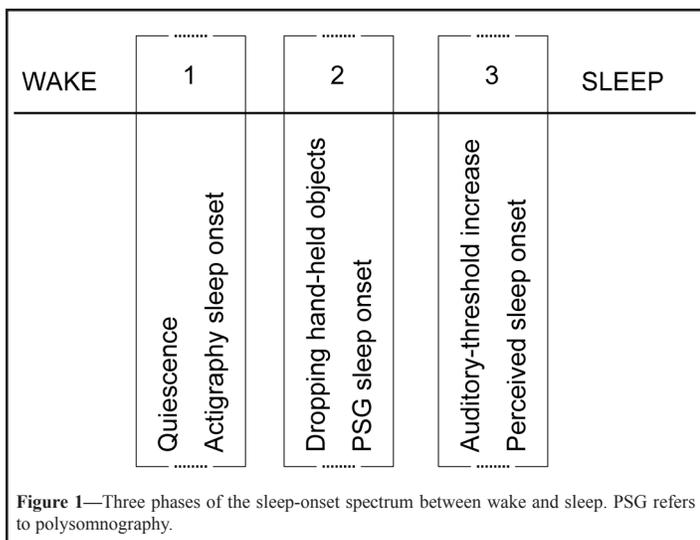
*Based on Thorpy MJ. Classification of sleep disorders. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 2nd ed. Philadelphia: WB Saunders; 1994:426-36.

selecting the subset of the 12 best items for each factor—those that correlated most positively or negatively with the same NEO-PI-R scales. Investigators⁸ reported correlations of .92 for Neuroticism, .91 for Openness, .90 for Extraversion, .87 for Conscientiousness, and .77 for Agreeableness. These correlations account for 85%, 83%, 81%, 76%, and 59% of variance, respectively. These very high correlation coefficients result because the short form of the test is a subset of the longer form of the test. This would be like using a subset of PSG channels to predict results from all PSG channels. Substantially lower correlations would likely result if different items were used to measure the same constructs. Hence, these personality correlations constitute an upper bound on the size of expected validity coefficients. Notice that even these unusually high validity coefficients leave 15%, 17%, 19%, 24%, and 41% of the variance unexplained. Hence, 15% to 41% error characterizes part-whole relationships among personality tests.

Tryon's⁹ Table 3 identifies 14 studies that have validated wrist actigraphy against PSG. His Table 4 presents correlations between actigraphy and PSG measured total sleep time, percent sleep, sleep efficiency and wake after sleep onset. These correlations ranged from .72 to .98 for Total Sleep Time, .82 to .96 for Percent Sleep, from .56 to .91 for Sleep Efficiency, and from .49 to .87 for Wake After Sleep Onset. Tryon's⁹ Table 5 presents percent agreement between actigraphy and PSG measured sleep and wake. They ranged from 78.8% to 99.7% for sleep and from 48.5% to 79.8% for wake. Pollak et al.³ reported results within this range. Their percent agreement results were 82.0% for nights, 98.6% for days, and 76.9% for 24-hour periods using logistic regression in a cross validation sample of 10 participants leaving 18.0%, 1.4%, and 23.1% agreement unaccounted for. This range of 2% to 23% error also falls within the acceptable errors associated with accepted medical, and the best intelligence, and personality tests.⁶ It therefore follows that actigraphy is as valid a sleep-wake indicator as common medical tests are valid indicators of pathology and the best psychological tests are indicators of intelligence and personality.

SLEEP LOGS

It has long been known that sleep logs can overestimate sleep latency and severely underestimate total sleep time.¹⁰⁻¹⁴ Some patients insist that they have not slept at all even after obtaining a full night's sleep as verified by PSG.¹⁵ Sleep-state misperception is diagnosed when a complaint of excessive sleepiness or insomnia is accompanied by a normal PSG.¹⁶ Discrepancies of this magnitude exceed the errors summarized above for psychological and medical tests and exceed the "errors" associated with actigraphy, yet sleep logs have not been "disqualified" as a sleep indicator, as actigraphy has been.³ Arguably, the errors associated with actigraphy are not sufficiently large so as to preclude actigraphy from detecting differences between objective and subjective sleep that characterize



the diagnosis of sleep-state misperception. The fact that sleep-state misperception can only be detected by comparing subjective and objective sleep indicators means that some objective method should always be used to supplement the subjective report, and PSG data are not always available.

THEORETIC CONTEXT

Actigraphy has been evaluated against a dichotomous sleep-versus-wake distinction that PSG presumably measures without error. Hence, all actigraphy minus PSG discrepancies have been attributed to random measurement error associated with actigraphy. This inference is flawed in several ways discussed below: sleep onset is a gradual rather than a discrete process that entails a series of changes. Actigraphy and PSG key on different phases of the sleep-onset process, resulting in systematic rather than random differences. A sleep switch device can be used to reduce these systematic errors. Even if the conceptual issue is resolved to mean that sleep is, and only is, what PSG measures, the empirical fact remains that PSG sleep scoring is not perfectly reliable. The unreliability of stage 1 sleep scoring alone accounts for an important fraction of the differences between actigraphy and PSG. Actigraphy is a single-channel measurement system, whereas PSG is a multichannel measurement system. It is unlikely, even in principle, for a univariate system to fully duplicate a multivariate system unless the multivariate system is completely redundant, which PSG and actigraphy are not. By professional agreement, PSG is the only accepted means by which to diagnose various sleep disorders and therefore has value other than as a sleep-wake indicator. Behavior therapy for sleep disorders and large-scale studies of sleep in various populations require an inexpensive sleep-wake estimator for which wrist actigraphy, especially if augmented with the sleep switch device,¹⁷ is a well-validated and reasonably economical method.

SLEEP-ONSET SPECTRUM

Previously reviewed empirical evidence⁹ warrants Figure 1, which presents a visual summary of 3 sleep-onset phases whose orderly sequence comprise the sleep-onset spectrum (SOS). The gaps between blocks show that not all sleep-onset events are represented. The ellipses (...) indicate that each phase is of variable width in time and duration for every person depending upon: (1) current life events—we all sleep better some nights than others; (2) developmental processes, frequently referred to as aging, that typically increase SOS duration, and (3) various diseases that can temporarily or chronically change, typically increase, SOS duration. On the contrary, PSG data are scored for wake, 4 stages of sleep, and rapid eye movements. Wake is contrasted with various degrees of sleep, implying that sleep-onset is a discrete process. This implication is supported by the fact that actigraphy is validated against PSG-based sleep-versus-wake distinctions without considering the SOS. The validity of such studies depends upon the validity of the discrete sleep-versus-wake dichotomy. The presumed upper limit of 100% agreement assumes that sleep is properly understood as distinct from wake and that both PSG and actigraphy can potentially identify both discrete states simultaneously. Otherwise, obtained results must be compared to a lower theoretic maximum agreement value. Ogilvie and Wilkinson¹⁸ and Harsh and Ogilvie¹⁹ describe sleep onset as a gradual process that entails a predictable progression of behavioral, physiologic, and psychological events. They use the term *sleep-onset period*. Tryon^{9,20} prefers the term SOS to emphasize the orderly sequence of events associated with sleep onset. Three major sleep-onset phases can be identified.

Phase 1: Quiescence and Immobility

Inactivity is necessarily the first sleep-onset phase. No organism goes to sleep before it becomes inactive. This SOS phase begins with inactivity. Actigraphy-identified sleep onset follows shortly, depending upon the sleep-scoring algorithm used. Good sleepers pass through this phase

rapidly, thereby causing actigraphy to identify sleep onset nearer to where PSG identifies sleep onset. Poor sleepers move through this SOS phase slowly, allowing for a large temporal gap to arise between actigraphy-identified sleep onset and the beginning of SOS Phase 2.

Phase 2: Decreased Muscle Tone

The leading edge of SOS phase 2 entails muscle-tone decreases to a point where hand-held objects are dropped.²¹⁻²⁵ Dropping hand-held objects was the gold standard against which electroencephalogram (EEG) sleep stage 1 criteria were validated.²⁶ This behavioral correlate of PSG sleep onset must continue to be acknowledged as valid if PSG sleep-onset measures are to continue to be accepted. One cannot invalidate an important validation criterion without also invalidating that which the criterion supports. Sleep onset has more recently been studied using this criterion. Ogilvie and colleagues²⁷ used a hand-held “dead-man” switch requiring 90 g of pressure to maintain closure to measure the drop point. Franklin²⁸ and Viens et al²⁹ described a similar apparatus suitable for home use. Hauri¹⁷ evaluated an inexpensive commercially available sleep switch device (RMP, Inc., Boynton Beach, Fla) that was found to correlate ($r_{23} = .98, P < .001$) in a sample of 19 insomnia patients and 6 normal sleepers with PSG-measured sleep onset. The arithmetic mean difference between sleep onset as defined by the sleep switch device and the first 10 minutes of solid sleep was -.8 minutes. The mean of the absolute differences, without regard to sign, was 5.1 minutes. The largest single deviation for any sleeper was 13.5 minutes. Use of the sleep switch device probably alters sleep-onset cognitions and therefore may change the sleep-onset process it intends to measure. However, the empirical results reported above, in conjunction with the fact that the sleep switch device simulates the dropping of hand-held objects—the gold standard against which PSG sleep-onset was validated—are reasons for continued clinical and research interest in it.

Phase 2: EEG Sleep Stage 1

The EEG changes that constitute the Rechtschaffen and Kales⁴ sleep-onset criteria occur soon after muscle tone has decreased and mark the second portion of this SOS phase, but some degree of consciousness (wake) remains. Ogilvie and Wilkinson,³⁰ while recording PSG data, instructed subjects to squeeze a hand-held microswitch when they heard a tone. Their results indicated that 99.3% of the tones were correctly responded to during EEG stage wake, 72.2% during stage 1 sleep, 24% during stage 2 sleep, and 5.3% during stage 3 sleep. That correct responding was not zero during stage 3 sleep questions the view that the measurement domain of PSG-defined sleep completely overlaps the conceptual domain of sleep.

Phase 3: Auditory Threshold Increase and Perceived Sleep Onset

Auditory threshold rises rapidly within 1 minute of the first EEG sleep spindle.³¹ Increases in the auditory threshold take place mainly during EEG stage 2 sleep.³² Subjects no longer respond to their name when it is spoken softly, to a light touch, or to normal external stimuli.^{30,33} Self-reported sleep onset, perceived sleep onset, appears to happen after increases in the auditory threshold occur.³²⁻³⁵ Espie and coworkers³⁶ found that self-reported sleep onset occurred concurrent with auditory-threshold increase. Lichstein et al³⁷ may be the only study indicating that self-reported sleep onset happens before auditory-threshold increases occur.

SLEEP-ONSET SPECTRUM PHASE PRIVILEGE

It is arguably inappropriate to define sleep onset as a discrete event that occurs at a single point in time given that the evidence indicates that sleep onset is a gradual and continuous multivariate process that takes place over a half hour or more in poor sleepers. Selecting a single point to dichotomize wake from sleep is therefore an arbitrary choice.

Rechtschaffen³⁸ cautioned investigators against privileging 1 measurement basis over all others. He stated, “... physiological measures derive their value as indicators of sleep from their correlations with the behavioral criteria, not from any intrinsic ontological or explanatory superiority.”^{38p5} He further concluded, “Any scientific definition of sleep that ignores the behaviors by which sleep is generally known unnecessarily violates common understanding and invites confusion.”^{38p4} Kleitman³⁹ cited a dozen studies showing discrepancies between behavioral and EEG sleep criteria and questioned using EEG as the sole basis for defining sleep. It remains an empirical question as to whether dichotomizing wake from sleep at the SOS point measured by actigraphy and at the point measured by PSG correlate comparably with clinical changes and outcomes and with other measures of interest.

ACTIGRAPHY VERSUS PSG DISCREPANCIES

Actigraphy is theoretically associated with SOS phase 1, whereas PSG is theoretically associated with SOS phase 2. Six studies confirm this phase difference by reporting that actigraph sleep-onset criteria are met *prior to* EEG stage 1 sleep criteria.⁴⁰⁻⁴⁵ The sleep-onset period is protracted in patients with insomnia. Hence, actigraphy sleep-onset times and PSG sleep-onset times should and do agree less well in patients with insomnia than in normal sleepers. For example, Hauri¹⁷ reported average actigraphy sleep-onset time of 5.5 minutes (SD = 6.2 min) versus mean PSG sleep-onset time of 34.2 minutes (SD = 33.3 min) to the first 10 minutes of solid sleep in patients with insomnia. That actigraphy sleep onset *systematically* precedes PSG sleep onset demonstrates that it validly keys on an earlier SOS phase, and, hence, actigraphy minus PSG differences are not random measurement error. Nor can these differences be entirely ascribed to actigraphy, as will be shown.

Hauri¹⁷ reported that the average sleep-onset time, as measured by sleep switch device, was 32.4 minutes (SD = 30.7 min). Hence, the sleep switch device can be used in conjunction with actigraphy to obtain sleep-onset times that more closely correspond with PSG-measured sleep-onset times. These differences of $34.2 - 5.5 = 28.7$ minutes can be reduced down to $34.2 - 32.4 = 1.8$ minutes. Hauri¹⁷ also reported that sleep-onset times from the sleep switch device correlated ($r_{23} = .98, P < .001$) with PSG-derived sleep-onset times in a sample of 19 insomnia patients and 6 normal sleepers.

Sleep-wake scoring-validity studies should include as much wake as sleep in order to avoid achieving close agreement between actigraphy and PSG on the basis of base rate alone. If only sleep is studied, then scoring all epochs sleep will artifactually produce high agreement. If only wake is studied, then scoring all epochs wake will also artifactually produce high agreement.

In principle, the pattern of transitions, both forward and backward through sleep-onset phases wake, 1, 2, 3, and sleep depicted in Figure 1 may account for most of the differences between actigraphy and PSG: (1) barring artifact such as, but not limited to, cosleeping with an active partner and/or sleeping in a waterbed where movements by the sleeper or cosleeper can reverberate, sleeping in a vibrating or rocking bed, and/or sleeping with one's wrist on his or her chest or abdomen, thereby picking up breathing artifact; (2) barring medication artifact such as, but not limited to, prescribed and/or illicit drugs or substances, alcohol, caffeine, and/or nicotine; and (3) barring artifact such as, but not limited to, diseases that produce involuntary movements including, but not limited to, restless legs syndrome and periodic limb movements disorder. For example, assuming that sleep is scored at fixed epochs, the following pattern, if frequently repeated, is likely to produce large differences in actigraphy and PSG: W...W, 1, 1, 1, ..., 1, 1, 1, 2, 3, S, ..., S, 3, 2, 1, 1, 1, ..., 1, 1, 1, 2, 3, S, ..., S, where the ellipsis represent continuations of the indicated state. It is unlikely that long continuous periods of high activity will be scored as sleep by either method. It is also unlikely that long continuous periods of deep sleep will be scored as wake by either method. Hence, actigraphy can consistently distinguish sleep from wake in the extreme, contrary to the conclusion by Pollak et al.³ Short periods

of active wake and deep sleep provide a greater challenge. Shorter periods of rest and light sleep constitute the greatest challenge because they may entail the SOS. Most discrepancies between actigraphy and PSG derive from traversing the SOS in both directions because they key on different SOS phases. The SOS seems to be traversed less rapidly from wake to sleep than from sleep to wake, thereby providing more opportunity for disagreement.

The SOS also informs our understanding of deviations between actigraphy and sleep logs and between PSG and sleep log in that perceived sleep onset may be caused by an increase in the auditory threshold. How else is a person to detect sleep onset with their eyes shut? That auditory-threshold increases occur after PSG sleep onset, which occurs after actigraphy sleep onset, requires discrepancies between actigraphy and the sleep log to be larger than between PSG and the sleep log. One factor that moderates this relationship is the extent to which memory forms for wake as the sleeper moves from SOS phase 3 toward wake and the extent to which memory forms regarding the frequency and duration of sleep-onset and sleep-offset events. Behavior observers typically record what they see soon after the event occurs while they are awake and attentive to minimize memory factors. Sleepers must necessarily recall events that took place hours ago when they were losing consciousness or may have been only partly conscious, depending upon how far back through the SOS they came. Another moderating factor is the extent to which the person is motivated to record all that can be recalled. This includes regularly completing the sleep log immediately upon awakening and spending sufficient time to obtain maximum recall.

RELIABILITY OF PSG

Actigraphy cannot be expected to agree more completely with PSG than PSG does with itself. Scoring of PSG sleep is conducted by human raters trained to implement Rechtschaffen and Kales's⁴ criteria. Automatic sleep stages are available but are not yet as good as hand scoring.⁴⁶ Interrater PSG sleep-scoring agreement values range from 80% to 98%.¹⁸ Whereas the reliability of scoring stage 2 sleep is approximately 90%,^{47p62} the reliability of scoring stage 1 sleep can be as low as 60%. Hence, up to 40% of EEG stage 1 scoring, and therefore sleep-wake scoring, and differences between PSG and actigraphy can be attributed to the unreliability of the scoring of stage 1 sleep. This 40% difference frequently equals or exceeds all of the sleep-onset differences between actigraphy and PSG.

REQUIRED AGREEMENT

The minimum degree of agreement between actigraphy or any other putative sleep measure, including sleep logs, and PSG in order to be considered valid and/or useful has not yet been determined. This minimum level may well vary depending upon the intended application. For example, a lower minimum level of agreement may be appropriate for screening purposes in a research context versus assessment in a clinical context. Should actigraphy be required to demonstrate agreement with criterion measures that exceed requirements for common medical tests and the best psychological tests? If so, then these reasons should be provided in future validation studies.

CHANGES BEFORE AND AFTER TREATMENT

It has been noted^{2,48} that even if systematic deviations between actigraphy and PSG are considered errors, actigraphy can possibly accurately track clinical change over time, such as response to sleep-hygiene instruction. A study by Coffield and Tryon⁴⁹ reported that wrist actigraphy was able to track sleep improvement in successfully treated, verified by psychometric testing, hospitalized patients and was also able to detect residual sleep problems at discharge.

The comparative ability of PSG and actigraphy to track clinical change has not been investigated but bears directly on the clinical and research validity and utility of actigraphy. Simultaneous PSG and acti-

graphy data have not been systematically compared in patients whose sleep is expected to change over time. It would be informative to evaluate the extent to which correlations between actigraphy and PSG in both clinical and control subjects are equal to PSG test-retest correlations in control subjects. Put otherwise, to what degree is the difference (*error*) between actigraphy and PSG equal to PSG test-retest difference (*reliability*) in control subjects?

LIMITATIONS

It is important to note that actigraphy is not a unitary methodology. Multiple vendors offer a range of actigraphs with different operating characteristics. The present variability in actigraph hardware probably exceeds the variability of contemporary polygraphs. Actigraph vendors provide a variety of sleep-scoring software; some score sleep in a single pass, whereas others use rescoring rules. It is important to note that current data constitute lower-bound estimates to the degree that these differences augment error variance. This variability may be reduced through standardization, thereby possibly reducing discrepancies between actigraphy and PSG below current levels. Some sleep-scoring software is better validated against PSG than are others. On the contrary, PSG sleep-scoring rules have been standardized.⁴ Polysomnography is frequently conducted with a technician or observer present who can verify the conditions under which data are obtained. Home PSG shares with actigraphy the need for participants to keep logs regarding the conditions under which data are recorded.

The limitations discussed above and throughout this article impact our ability to use actigraphy clinically. It is presently unknown to what extent "mistakes" are made using actigraphy when considering PSG as the errorless standard. One can construct a variety of scenarios depending upon how frequently and how completely one moves back and forth across the SOS for each night data are recorded, which can range from a few days to multiple weeks in the case of actigraphy. Given that all possibilities have some probability of occurrence, this issue reduces to a question of what these probabilities are and an informed answer requires further empirical research.

Visual inspection of actigraphy records over 7 to 14 nights may be sufficient to objectively confirm patient complaints of disrupted sleep if sleep is seriously disturbed. It remains an open empirical question as to how slight a sleep disturbance actigraphy can detect, and that partly depends upon how many nights of data one collects. A small disturbance that consistently repeats for 14 consecutive nights is more likely to be noticed than if such a sleep disturbance were limited to 1 or 2 nights. Early PSG researchers recommended averaging results over 3 to 5 nights, but current practice more generally involves 2 to 3 nights. Measurement reliability is known to be a function of test length, with longer tests being more reliable. The Standards of Practice Committee noted in conjunction with repeated measurements that "Even if absolute values of actigraphy such as sleep efficiency and total sleep time are imprecise, trends may be accurately reflected."^{2p286} The ability of actigraphy to conveniently extend over multiple nights is a methodologic asset that should be fully exploited whenever possible. Hence, the most comprehensive test of the clinical utility of actigraphy will compare extended actigraphy against standard sleep-laboratory evaluation. Polysomnography may also be extended through home assessment, but the primary comparison is with standard PSG sleep-laboratory evaluation, as professionals are currently comfortable with this method. These and related limitations of actigraphy are discussed elsewhere.⁵⁰⁻⁵²

Existing actigraph sleep-scoring algorithms use multivariate statistics based on the general linear model. It is possible that connectionist neural-network pattern-recognition scoring systems that implement nonlinear multivariate statistics might be more effective. They could be trained by examining activity recordings during windows centered on PSG-detected transitions from wake to sleep and sleep to wake in many good and poor sleepers examined over multiple nights.

Actigraphy is a single-channel measurement system, whereas PSG is a multichannel measurement system. A univariate system can only fully

duplicate a multivariate system when all multivariate channels are completely redundant. Polysomnography and actigraphy are not entirely redundant, and therefore it is unreasonable to expect actigraphy to completely duplicate PSG results. Studies published to date have compared agreement between actigraphy and PSG statistics to 100% and therefore have incorrectly presumed that complete agreement is at least theoretically possible. Some lesser upper theoretic limit is arguably more appropriate.

ASSESSMENT AND SCREENING

One of the reviewers posed the following question: "What do we do with measures that are more "accurate" for some types of sleep-disordered patients than others when the purpose is to assess (screen) whether they are patients or not and what kind of disorder they might have?" The same reviewer then made the following statement: "If the correspondence between actual sleep and actigraph-estimated sleep systematically varies as a function of group membership, we are in a deep hole. This is particularly apparent for insomniac patients but is also the case for any type of patient with disrupted sleep or movement disorders during sleep. This systematic variability (between actigraphy and PSG) limits the usefulness of actigraphy for research studies aimed at comparing measures between some groups or for clinical studies aimed at screening some patients." The main concern here is that the systematic differences between actigraphy and PSG are not constant but may vary as a function of sleep and movement disorders and that the degree of difference ranges from small in the case of normal sleepers to large in some clinical cases. This problem is particularly acute when comparing means of a group with an unspecified sleep disorder and normal controls asking the question: "What type of sleep disorder is present?" The fact that the difference between actigraphy and PSG could be large, medium, or small confounds interpretations based on observed differences between means. This certainly is a "deep hole" and should therefore be avoided by not attempting to make such inferences.

An alternative approach is to make clinical and/or research decisions on a case-by-case basis. Table 2 specifies the 4 comparative decisions that would arise if a screening or comparative assessment study were conducted using both actigraphy and PSG. Normal sleepers, Cell A, constitute the largest fraction of participants in screening studies, and their actigraphy data contain the least error. Existing data indicate that actigraphy and PSG agree quite well for normal sleepers, which means that consistent decisions will likely be made for persons without sleep disorders. By inference, all other participants are judged to have some form of sleep disorder, the specifics of which are to be decided on the basis of subsequent testing. This approach to screening capitalizes on the issue of variable actigraphy-PSG differences by identifying good sleepers where agreement between actigraphy and PSG is high.

The main concern expressed by the reviewer pertains to Cell D, where a question is raised as to how frequently actigraphy and PSG can agree regarding sleep disorders, given the systematic but variable actigraphy-PSG difference. Table 3 considers the relevance of actigraphy to the sleep disorders for which it is most appropriate. The entries in the right-hand column describe how actigraphy informs the diagnosis of each sleep disorder. Six observations about this table are noteworthy. First, 15 sleep disorders are involved, which documents a wide range of applicability. Second, all of the figures presented by Mistlberger and Rusak⁵³ in their chapter on circadian rhythms in mammals are of activity level, which means that actigraphy is especially suited for assessing circadian-rhythm disorders in people when actigraphs are worn 24 hours per day for 1, 2, or more weeks. Ancoli-Israel et al⁵⁴ reported severe fragmentation of circadian rhythms in nursing home residents. Gruen⁵⁵ used actigraphy to track circadian changes during a westerly trip around the world. Findings based on PSG, if any, are absent from the Mistlberger and Rusak⁵³ chapter, making actigraphy-PSG discrepancies moot. Third, Hauri's¹⁷ sleep switch can substantially reduce the discrepancies between actigraphy and PSG. Fourth, PSG utilizes multichannel assessment, and it may be necessary to use more than 1 channel of actigraphy

to reach correct decisions. Ankle actigraphy may be a useful supplement to wrist actigraphy to independently identify periodic leg movements during sleep. Abdominal actigraphy may be a useful adjunct to wrist actigraphy to independently identify violent incidents of respiration associated with the termination of sleep apnea. Fifth, extrinsic sleep disorders rely on information that neither PSG nor actigraphy can provide, thus rendering actigraphy-PSG differences moot for diagnosing these sleep disorders. This observation argues against the type of blind between-group comparisons found problematic above and raises the necessity of querying people regarding activities of daily living and obtaining sleep-diary information. Sixth, none of the test results indicated in Table 3 are necessarily negated by the fact that actigraphy-PSG differences vary across sleep disorders. In sum, Table 3 identifies contributions that actigraphy can make to 15 sleep disorders. This information appears relevant to screening and assessment purposes. These contributions are not compromised by systematic variation in actigraphy and PSG across sleep disorders.

HOW ACCURATE DO YOU NEED TO BE?

Highly accurate assessment procedures are frequently also the most expensive procedures. Less-accurate assessment procedures are more cost effective and may be preferred, depending upon the cost of making various screening errors. A cost-benefit analysis could result in choosing a somewhat less accurate but far less expensive method. The human cost of making decision errors must also be considered when determining what level of accuracy is required in a particular situation.

Systematic Versus Random Error

Systematic variation is more tractable than is random variation because the direction of bias is known. Variation in the degree of difference is less problematic when the sign of the difference remains constant than when it varies randomly. Knowing that actigraphy always identifies sleep onset before PSG means that actigraphy **underestimates** sleep-onset latency and **overestimates** total sleep time and, therefore, percentage of sleep and sleep efficiency. This means that actigraphic measures establish upper bounds to sleep-onset latency and lower bounds to total sleep time, percentage sleep, and sleep efficiency. It is possible to effectively reason within these limits.

Knowing that movement disorders can produce actigraphic data that will be scored wake while corresponding PSG data are scored sleep means that actigraphy **underestimates** total sleep time and, therefore, percentage of sleep and sleep efficiency. Hence, movement disorders partially reduce the overestimates of these sleep variables noted above. The possibility that movement disorders could potentially "correct" actigraphy records from insomniacs is problematic. One possible solution is to supplement wrist actigraphy with ankle actigraphy to independently determine leg movements as noted above.

Artifact

The relationship between inactivity and sleep and the ability to infer sleep from inactivity is threatened by at least the following measurement artifacts: (1) sleeping with an active bed partner, (2) sleeping in a waterbed or rocking or vibrating bed, (3) sleeping with his or her wrist on the chest or abdomen, (4) having a concurrent movement disorder that does not consistently wake the person, (5) ingesting medications that produce movement but do not wake the person, and (6) lying very still while awake for extended periods of time. Minimizing or eliminating these artifacts increases the probability of a correct inference. These artifacts impair inference more than the fact that systematic actigraphy-PSG differences vary across sleep disorders.

GENERALIZABILITY

Researchers and clinicians conduct tests in order to generalize to the patient's natural situation. If the results of PSG pertained only to sleep in

sleep laboratories and not to the home environment, then it would provide little useful information for either scientific or clinical purposes. External validity and generalizability are 2 terms psychologists use when discussing the applicability of psychological test data, and these concepts pertain to this discussion. External validity must be demonstrated rather than presumed. While PSG is sometimes conducted in the home, the extent to which this assessment procedure is reactive and alters the behavior being assessed is not known because there is no way to know how the person sleeps, in terms of PSG criteria, without using PSG methods. The necessary electrodes and associated equipment needed to conduct PSG may alter sleep in unknown ways. Part of the differences between actigraphy and PSG may be traceable to different methodologies. Actigraphy is less invasive and can be continued for a longer time, which means that people can get used to wearing actigraphs, especially if they are wearing them 24 hours per day.

While collecting data under laboratory conditions in the presence of a technician who can vouch for the conditions under which data were collected is important in order to validate data-collection procedures, the presence of a third party may modify the sleep environment in important ways that do not generalize to when the person sleeps without the technician and without electrodes and with their bed partner in their own bed. Unattended home PSG suffers from the same uncontrolled assessment problems, as does unattended home actigraphy.

Is Variability Diagnostic?

Polysomnography is frequently conducted over 3 nights with the first night's results discarded. Variability in sleep measures is therefore reduced to a single difference based on the 2 remaining nights. Actigraphy can be conducted for extended time periods, and therefore many more differences can be calculated. Three types of variability can be calculated. The first type of variability is the standard practice of summing the squared deviations of scores about the mean and dividing by $N-1$ observations to calculate variance and perhaps taking the square root to calculate the standard deviation. The second type of variability entails calculating and squaring differences across consecutive nights. One week of actigraphy, 7 nights, yields 6 consecutive differences. Two weeks of actigraphy, 14 nights, yields 13 consecutive differences. The third type of variability entails calculating and squaring all possible differences between nights resulting in $N(N-1)/2$ independent differences. One week of actigraphy provides 21 such differences. Two weeks of actigraphy provides 91 such differences. Perhaps some sleep disorders leave, and can be identified by, a variability signature. If so, then this approach may provide an end run around the problem that actigraphy-PSG differences vary across sleep disorders.

Empirical support for the utility of assessing sleep stability is provided by Gruber et al.⁵⁶ who obtained actigraphically measured sleep variables over 5 consecutive school nights in 38 boys with attention-deficit/hyperactivity disorder, aged 6 to 14 years, and 64 control boys, where morning rising time was determined by the time school started. Results of multivariate analysis of covariance and univariate analyses of means failed to find significant differences regarding sleep-onset time, sleep duration, true sleep (sleep time excluding periods of wake), sleep percentage, night wakings, and the longest sleep period. However, multivariate analysis of covariance on the standard deviations of sleep measures yielded significant group differences ($F = 4.87, P < .0001$). Subsequent univariate analyses revealed that variability in sleep duration ($F = 27.77, P < .0001$) and variability in true sleep ($F = 27.15, P < .0001$) significantly differentiated the 2 groups. Neither analysis of mean subjective sleep measures or their standard deviations yielded statistically significant results. The authors concluded that instability of the sleep-wake system might be characteristic of children with attention-deficit/hyperactivity disorder. Actigraphy enables investigators to further empirically examine this possibility, as self-report measures seem insensitive to such variability.

CONCLUSIONS

The following conclusions can be reached: (1) Validity coefficients published to date, including those by Pollak et al.,³ exceed the validity coefficients associated with common medical tests and the best psychological tests. (2) If there are reasons why actigraphy should be held to a substantially higher empirical standard than are common medical tests and the best psychological tests, these reasons have yet to be advanced. (3) Differences between actigraphy and PSG are not random. (3a) Sleep onset is a gradual rather than a discrete process. Available evidence indicates that actigraphy keys on an earlier phase of the sleep-onset process than PSG does. The resulting differences between actigraphy and PSG are therefore systematic rather than random. (3b) A sleep switch device¹⁷ can be used to supplement actigraphy to substantially reduce differences in the sleep onset measured by actigraphy and PSG. (3c) The reliability of human PSG-derived sleep scoring is not perfect, especially for stage 1 sleep. The residual unreliability of PSG data explains a portion of the differences between actigraphy and PSG. Future validation studies of actigraphy should report the reliability of PSG sleep-wake determinations by rescoring all PSG records a second time by a second technician who is blind to the results of the first technician. Actigraphy cannot be expected to agree more completely with PSG than PSG does with itself.

(4) Complete concordance between actigraphy and PSG has been presumed by investigators who compare actigraphy-PSG agreement statistics to 100%, but such a limit is theoretically possible only when all channels of the multichannel measurement system (PSG) are completely redundant with each other and completely redundant with the single-channel measurement system (actigraphy). Such a claim has not been made and should not be presumed. The theoretic upper limit of actigraphy-PSG agreement is therefore less than 100% (by an unknown amount) on this basis alone. Future research should estimate the maximum degree of agreement based upon the different numbers of measurement channels used. (5) Conclusions that actigraphy is not an accurate sleep-wake indicator and that it is inappropriate to infer sleep from actigraphy data conflict with empirical data to the contrary. Conclusions reached by the Standards of Practice Committee remain supported, including their caveats.

REFERENCES

1. Kupfer DJ, Dete TP, Foster FG, Tucker GJ, Delgado J. The application of Delgado's telemetric mobility recorder for human studies. *Behav Biol* 1972;7:585-90.
2. Standards of Practice Committee. Practice parameters for the use of actigraphy in the clinical assessment of sleep disorders. *Sleep* 1995;18:285-7.
3. Pollak CP, Tryon WW, Nagaraja H, Dzwonczyk R. How well does wrist actigraphy identify states of sleep and wakefulness? *Sleep Res* 2001;24:957.
4. Rechtschaffen A, Kales A, eds. *A Manual of Standardized Terminology, Techniques, and Scoring System for Sleep Stages of Human Subjects*. Los Angeles: Brain Information Service/ Brain Research Institute, UCLA; 1968.5.
5. Anastasi A, Urbina S. *Psychological Testing*. 7th ed. Upper Saddle River, NJ: Prentice Hall; 1997.
6. Meyer GJ, Finn SE, Eyde LD, et al. Psychological testing and psychological assessment: A review of evidence and issues. *Am Psychologist* 2001;56:128-65.
7. Spreen O, Strauss E. *A Compendium of Neuropsychological Tests*. 2nd ed. New York: Oxford University Press; 1998.
8. Costa Jr. P T, McCrae RR. *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc; 1992.
9. Tryon WW. Nocturnal activity and sleep assessment. *Clin Psychol Rev* 1996;16:197-213.
10. Baekeland F, Hoy P. Reported vs. recorded sleep characteristics. *Arch Gen Psychiatry* 1971;24:548-51.
11. Carskadon MA, Dement WC, Mitler MM, Guilleminault C, Zarcone VP, Spiegel R. Self-reports versus sleep laboratory findings in 122 drug-free subjects with complaints of chronic insomnia. *Am J Psychiatry* 1976;133:1382-8.
12. Frankel BL, Coursey RD, Buchbinder R, Snyder F. Recorded and reported sleep in chronic primary insomnia. *Arch Gen Psychiatry* 1976;33:615-23.
13. Freedman R, Papsdorf J. Biofeedback and progressive relaxation. Treatment of sleep-onset insomnia: A controlled, all-night investigation. *Biofeedback Self Regul* 1976;1:253-71.
14. Hauri P, Fisher J. Persistent psychophysiological (learned) insomnia. *Sleep* 1986;9:38-53.
15. McCall WV, Edinger JD. Subjective total insomnia: An example of sleep state misperception. *Sleep* 1992;15:71-3.
16. Hauri PJ. Primary insomnia. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 2nd ed. Philadelphia: WB Saunders; 1994:494-9.

17. Hauri P. Evaluation of a sleep switch device. *Sleep* 1999;22:1110-7.
18. Ogilvie RD, Wilkinson RT. Behavioral versus EEG-based monitoring of all-night sleep/wake patterns. *Sleep* 1988;11:139-55.
19. Harsh JR, Ogilvie RD. Introduction: A first sketch. In: Ogilvie RD, Harsh JR, eds. *Sleep-Onset: Normal and Abnormal Processes*. Washington, DC: American Psychological Association; 1994:xvii-xxviii.
20. Tryon WW. *Activity Measurement in Psychology and Medicine*. New York: Plenum; 1991.
21. Carskadon MA, Dement WC. Normal human sleep: An overview. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 2nd ed. Philadelphia: WB Saunders; 1994:16-25.
22. Chase MH, Morales FR. The control of motor neurons during sleep. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine* 2nd ed. Philadelphia: WB Saunders; 1994:163-175.
23. Loomis AL, Harvey EN, Hobart G A. Cerebral states during sleep as studied by human brain potentials. *J Exp Psychol* 1937;21:127-44.
24. Perry TJ, Goldwater BC. A passive behavioral measure of sleep-onset in high-alpha and low-alpha subjects. *Psychophysiology* 1987;24:657-65.
25. Snyder F, Scott J. The psychology of sleep. In: Greenfield NS, Sternback RA, eds. *Handbook of Psychophysiology*. Toronto: Holt, Rinehart & Winston; 1972:645-708.
26. Blake H, Gerard RW, Kleitman N. Factors influencing brain potentials during sleep. *J Neurophysiol* 1939;2:48-60.
27. Ogilvie RD, Wilkinson RT, Allison S. The detection of sleep-onset: behavioral, physiological, and subjective convergence. *Sleep* 1989;12:458-74.
28. Franklin J. The measurement of sleep-onset latency in insomnia. *Behav Res Ther* 1981;19:547-9.
29. Viens M, De Koninck J, Van den Bergen R, Audet R, Christ G. A refined switch-activated time monitor for the measurement of sleep-onset latency. *Behav Res Ther* 1988;26:271-3.
30. Ogilvie RD, Wilkinson RT. The detection of sleep-onset: behavioral and physiological convergence. *Psychophysiology* 1984;21:510-20.
31. Bonnet MH, Moore SE. The threshold of sleep: perception of sleep as a function of time asleep and auditory threshold. *Sleep* 1982;5:267-76.
32. Bonato RA, Ogilvie RD. A home evaluation of a behavioral response measure of sleep/wakefulness. *Percept Mot Skills* 1989;68:87-96.
33. Lindsley OR. Operant behavior during sleep: A measure of depth of sleep. *Science* 1957;126:1290-1.
34. Birrell PC. Behavioral, subjective, and electroencephalographic indices of sleep-onset latency and sleep duration. *J Behav Assess* 1983;5:179-90.
35. Lichstein KL, Hoelscher TJ, Eakin TL, Nickel R. Empirical sleep assessment in the home: A convenient, inexpensive approach. *J Behav Assess* 1983;5:111-8.
36. Espie CA, Lindsay WR, Espie LC. Use of the sleep assessment device (Kelley and Lichstein, 1980) to validate insomniacs' self-report of sleep pattern. *J Psychopathol Behav Assess* 1989;11:71-9.
37. Lichstein KL, Hoelscher TJ, Eakin TL, Nickel R. Empirical sleep assessment in the home: a convenient, inexpensive approach. *J Behav Assess* 1983;5:111-8.
38. Rechtschaffen A. Sleep-onset: conceptual issues. In: *Sleep-Onset: Normal and Abnormal Processes*. Washington, DC: American Psychological Association; 1994:3-17.
39. Kleitman N. *Sleep and Wakefulness*. Chicago: University of Chicago Press; 1963.
40. Cole RJ, Kripke DF. Progress in automatic sleep-wake scoring by wrist actigraph. *Sleep Res* 1989;18:331.
41. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. *Sleep* 1992;15:461-9.
42. Hauri PJ, Wisbey J. Wrist actigraphy in insomnia. *Sleep* 1992;15:293-301.
43. Mullaney DJ, Kripke DF, Messin S. Wrist-actigraphic estimation of sleep time. *Sleep* 1980;3:83-92.
44. Stampi C, Broughton R. Ultrashort sleep-wake schedule: Detection of sleep state through wrist actigraph measures. *Sleep Res* 1989;18:100.
45. Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborny G. An activity-based sleep monitor system for ambulatory use. *Sleep* 1982;5:389-99.
46. Carskadon MA, Rechtschaffen A. Monitoring and staging human sleep. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 2nd ed. Philadelphia: WB Saunders; 1994:943-60.
47. Spiegel R. *Sleep and Sleeplessness in Advanced Age*. New York: SP Medical & Scientific Books; 1981.
48. Sadeh A, Hauri PJ, Kripke DF, Lavie P. The role of actigraphy in the evaluation of sleep disorders. *Sleep* 1995;288-302.
49. Coffield TG, Tryon WW. Construct validation of actigraphic sleep measures in hospitalized depressed patients. *Behav Sleep Med* (in press).
50. Sadeh A, Sharkey KM, Carskadon MA. Activity-based sleep-wake identification: An empirical test of methodological issues. *Sleep* 17:201-7.
51. Chambers MJ. Actigraphy and insomnia: A closer look Part 1. *Sleep*. 1994:405-408.
52. Hauri PJ, Wisbey J. Actigraphy and insomnia: a closer look Part 2: a reply to Dr. Mark J. Chambers. *Sleep* 1994:408-10.
53. Mistlberger, RE, Rusak, B. Circadian rhythms in mammals: Formal properties and environmental influences. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 2nd ed. Philadelphia: WB Saunders; 1994:277-85.
54. Ancoli-Israel S, Parker L, Sinaee R, Fell RL, Kripke DF. Sleep fragmentation in patients from a nursing home. *J Gerontol: Med Sci* 1989;44:M18-21.
55. Gruen W. Wrist-actigraph measurements of sleep-wake patterns during a 28-day westerly trip around World. Poster 490 presented at the 5th International Congress of Sleep Research. 1987, 691.
56. Gruber R, Sadeh A, Raviv A. (). Instability of sleep patterns in children with Attention-Deficit/Hyperactivity Disorder. *J Am Acad Child Adolesc Psychiatry* 2000;39:495-501.
57. Thorpy MJ. Classification of sleep disorders. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine*. 2nd ed. Philadelphia: WB Saunders; 1994:426-36.