# Validation of the Fatigue Science Readiband™ Actigraph and Associated Sleep/Wake Classification Algorithms

Russell, C.A., PhD, Caldwell, J.A., PhD, Arand, D., PhD, Myers, L.J., PhD, Wubbels, P., BS, and Downs, H., PhD

Requests for reprints should be addressed to
the principal author at **chris@archinoetics.com**

**Archinoetics, LLC**
700 Bishop St, Suite 2000
Honolulu, HI 9681

## Background

Fatigue associated with 24/7 operations is increasingly recognized as a serious hazard in the modern industrialized world. The combination of shift work, prolonged duty cycles, and poor sleep continuously challenges the basic biological capacity of personnel, and failure to manage fatigue associated with these factors can place people and organizations at risk in terms of both safety and health. In the past, the effects of fatigue have been grossly underestimated (Rosekind, 2005); however, there is increasing awareness that fatigue is a substantial problem. For instance, whereas drowsiness on the job once was thought to account for only 3-7 percent of transportation mishaps, more recent detailed examinations indicate that as many as 10-20 percent of accidents are most likely fatigue related (Akerstedt and Haraldsson, 2001).

## The dangers of fatigue

The heightened recognition of fatigue as a safety issue has been facilitated by evolving scientific knowledge regarding the impact of prolonged sleep deprivation and chronic sleep restriction on alertness, health, and performance. It is now clear that the average adult needs 7-9 hours of sleep every day in order to function at the highest level of effectiveness, and that failing to fulfill this genetically predetermined sleep quota rapidly leads to substantial cognitive impairments from which full recovery is more slow and difficult than previously thought (i.e., two fairly-recent sleep-restriction studies showed that full recovery following 7-14 days of chronic sleep loss required longer than 3 days even after subjects were again able to obtain 8 hours of sleep per night. In addition, recent studies have shown that despite beliefs to the contrary, people are unable to “train themselves” to function at full capacity on less than their biologically-determined sleep need. Instead, chronically insufficient sleep creates a “sleep debt” that cumulatively degrades vigilance and attention, slows cognition, deteriorates short term memory, decreases frontal lobe functions, increases unpredictable and involuntary sleep onsets, and generally impairs waking cognition and performance (Banks and Dinges, 2007; Bonnet, 1994; Dinges, 1992; Horne, 1988; Horne, 1993; Koslowsky and Babkoff, 1992; Naitoh, 1975; Thomas, Sing, and Belenky, 1993).

Unfortunately, despite the mounting evidence about the dangers of inadequate sleep, sleep restriction remains a growing problem in industrialized societies. The National Sleep Foundation (2008) recently reported that US adults are sleeping only 6 hours and 40 minutes per night during the work week—an amount which falls significantly below the 8 hours recommended by experts. The reasons for such widespread sleep restriction have not been fully illuminated, but a rise in the number of hours typically worked by adults is at least partially to blame (Jacobs and Gerson, 2004) since this work-time commitment has not been offset by a decrease in other obligations associated with family, social, and other responsibilities. In fact, a definitive link has been established between adult work hours and decreased sleep (Caruso et al., 2006). Given the growing “do-more-with-less” mentality that is pervasive in today’s fast-paced world, there is every indication that the situation will likely get worse long before it gets better.

### Fatigue and on-the-job performance

From a performance and safety standpoint, chronic sleep loss is an important concern. For instance, a recent study of train operators indicated that fatigued drivers perform less consistently, utilize more fuel, and commit more speeding violations than well-rested drivers—negatively impacting both efficiency and safety (Dorrian, Hussey, and Dawson, 2007).

Studies of highway crashes typically indicate that drowsy drivers account for approximately 100,000 accidents, 1,357 fatalities, and 71,000 injury crashes in the U.S. each year (National Highway Traffic Safety Administration, 2009), but some investigators consider this an underestimate, suggesting instead that drowsy driving could account for over 1 million accidents per year (Rosekind, 2005). Aviation safety statistics suggest that approximately 4-to-7 percent of U.S. commercial aviation mishaps are fatigue-related (Kirsch, 1996). A study of interns revealed that those working five or more extended shifts (those greater than or equal to 24 hours) were three times more likely to make fatigue-related errors resulting in a patient fatality (Barger, et al., 2006), and a survey of health-care workers consisting primarily of nurses revealed that 19 percent believed fatigue contributed to a worsening of patients' conditions (Rosekind, 2005). An assessment of occupational injuries to a representative sample of U.S. adults indicated that the injury rate associated with overtime (and by inference, sleep loss) was 61 percent higher than the rate in jobs without overtime; working 12 hours a day increased the hazard rate 37 percent; and working 60 hours a week increased the work-related illnesses or injuries by 23 percent (Dembe et al., 2005).

### Fatigue and personal well-being

From a personal well-being standpoint, chronically insufficient sleep has been associated with health risks, social/familial problems, and workplace difficulties. For example, people who suffer sleep loss from insomnia have been found to utilize healthcare resources at a higher rate than their well-rested peers. And, they suffer from an overall reduction in quality of life, impaired day-to-day functioning, increased depression and/or anxiety, an elevated risk of alcohol or drug abuse, and a host of other difficulties (Thase, 2005). Stoller (1994) blames $80 billion in annual productivity- and accident-related losses on impaired daytime alertness associated with insomnia, and Walsh and Ustun (1999) attribute $12 billion in annual healthcare expenditures to sleep disruptions. The average 6-month health- and work/absenteeism-related cost differential between patients with insomnia versus those without insomnia is approximately $1,200.00 (Ozminkowski, Wang, and Walsh, 2007). Insomniacs are more likely than well-rested people to experience industrial or automotive accidents; have work-related performance problems; manifest high rates of absenteeism; suffer neurocognitive impairments; and fall victim to depression (Roth, 2007). Despite the probability that differences exist between people who lose sleep due to insomnia and those who lose sleep intentionally (to expand their waking hours), there is evidence that the fatigue suffered by both groups has similar effects.

## Growing fatigue awareness

In light of the above statistics, it is interesting to consider that a complete appreciation of the importance of sleep for health, safety, mood, productivity, and quality of life has become widely recognized only during the last 2 decades in spite of the fact that sleep deprivation has been studied for over 100 years (Balkin et al., 2008). The occurrence of recent and highly-publicized fatigue-related industrial and transportation mishaps has helped to bolster public and professional awareness of the problem. As a result, a new and widespread societal focus on the development and implementation of fatigue risk management systems (FRMS), especially for safety-sensitive occupations, has developed. Unfortunately, effective real-world counter-fatigue initiatives have been slow to progress.

## Barriers to fatigue management

The primary barrier to effective fatigue management in applied settings appears to be the result of inadequate fatigue measurement strategies. While there are subjective fatigue questionnaires such as the Epworth Sleepiness Scale (ESS), electrophysiological tests such as the Multiple Sleep Latency Test (MSLT), ocular assessments such as the Fitness Impairment Test (FIT), and reaction-time tests such as the Psychomotor Vigilance Task (PVT) that are useful for judging sleepiness (fatigue) levels (Mathis and Hess, 2009), it is unfortunately the case that these assessments are not adequate for making day-to-day fitness-for-duty determinations. In some cases, the tests are not feasible from a timing standpoint, and in other cases, a lack of employee compliance has frustrated efforts to implement these types of fatigue measurements.. And since there are no quick and reliable physiological markers for fatigue and no easy to use "fatigue Breathalyzers," the levels of worker impairment from fatigue cannot be readily evaluated in the same manner as impairment from drug or alcohol intoxication (for which employees often are already routinely tested). There is no specific test that can quickly evaluate on a day-to-day basis whether individual employees reporting for work are "fit for duty" from a fatigue standpoint, and thus it is difficult to characterize the extent of overall organizational fatigue problems, and to establish the degree to which any implemented fatigue-mitigation efforts have been or are being successful.

## A method for accurately measuring and monitoring fatigue

There is a combination of tools that can unobtrusively assess performance capabilities and fatigue risk by tracking the work/rest schedules, sleep quantity, and sleep quality of people and then processing these data through a validated performance-readiness/fatigue-risk prediction model. The utility of this strategy lies in the fact that 1) fatigue is known to stem primarily from insufficient and/or disrupted sleep in combination with circadian factors (Caldwell, Caldwell, and Schmidt, 2008), 2) the fatigue impact of different sleep/wake schedules and different levels of sleep quality/quantity have been well-established by over 100 years of scientific research (Balkin et al., 2008; Roth, 2004), 3) wrist actigraphy is able to accurately and unobtrusively track sleep information and aspects of circadian rhythms useful for predicting fatigue states (Sadeh and Acebo, 2002), and 4) sleep and circadian information can be processed through computerized mathematical models which translate this information into accurate predictions of alertness, performance, and fatigue risk (Van Dongen, 2004; Hursh et al., 2004).

### The connection between sleep, circadian rhythms, and fatigue

As noted earlier, according to sleep experts, most adults need 8 hours of sleep per night in order to function optimally, and losing even small amounts of sleep each night will seriously degrade alertness, performance, and vigilance (Balkin et al., 2000; Carskadon, and Roth, 1991). The body's internal clock often compounds the impact of systematic sleep loss because the physiological tendency to sleep at night and to be awake during the day is powerful, and difficulties occur when schedule changes cause personnel to work against this tendency. Night workers typically experience greater performance and alertness problems than their day-working counterparts simply because they are striving to maintain wakefulness at a time during which the body is programmed for sleep. Furthermore, simply altering the normal sleep/wake cycle, either through night work or time zone changes, impairs the ability to obtain adequate recovery sleep, and this typically leads to a further exacerbation of any existing cumulative sleep debt.

**Problems associated with disrupted sleep**

According to Roth (2007), numerous studies have proven that reductions in total sleep time and/or degraded sleep continuity are associated with impaired alertness, and a wide array of performance and memory deficits. Sleep loss likewise has been shown to increase the risk of becoming involved in an accident, developing psychiatric problems, missing work, and over-utilizing health care resources. Especially significant is the fact that chronic sleep restriction has even been associated with increased morbidity.

**The use of actigraphy to monitor sleep and fatigue.**

Given the importance of sleep for general wellbeing and the importance of sleep and circadian factors for operational performance effectiveness, it is necessary to be able to monitor at least some aspects of the sleep and the body clocks of people in order to make predictions about their ability to function effectively in the workplace. The “gold-standard” for measuring sleep is polysomnography—a procedure which involves monitoring the electroencephalographic, electromyographic, and electrooculographic activity of humans. Sleep also can be assessed via the more operationally-feasible technique of wrist actigraphy. Actigraphy is a relatively non-invasive method of monitoring rest/activity cycles and sleep via a small wrist-worn device which contains accelerometers that track the frequency of wrist movements, and processing this information through various algorithms to establish sleep/wake and sleep quality measures.

Although a comprehensive discussion of actigraphy, the different types of available actigraphs, and the various methods for processing actigraphy data is beyond the scope of this paper, a review of the published literature demonstrates that wrist-worn actigraphy has long been used to monitor sleep duration and sleep quality in circumstances where polysomnography is impractical. The Standards of Practice Committee of the American Academy of Sleep Medicine (2007) concluded that actigraphy provides an accurate estimate of sleep patterns in healthy people. It can also assist in the assessment of people suspected of experiencing certain sleep problems to include shift work disorder, jet lag disorder, obstructive sleep apnea, insomnia, hypersomnia, and various other circadian rhythm and sleep disorders. Actigraphy also is viewed as a useful method for establishing how well patients with circadian- rhythm disorders and insomnia respond to therapeutic interventions, and some investigators consider actigraphy a “better fit” than polysomnography for conducting long-term personal sleep-habit assessments (Merilahti et al., 2007).

A fairly recent comprehensive review established that actigraphic measures of sleep and polysomnographically-derived measures of sleep correlate 80-85% in both normal and patient populations (American Sleep Disorders Association, 1995). While not as accurate as polysomnography, there is general agreement that actigraphy, with its ability to record continuously for long time periods, is more reliable than patients’ subjective recall of the frequency or nighttime awakenings or the duration of nightly sleep. Also, although actigraphy by itself cannot diagnose specific sleep disorders, it is clearly useful for identifying poor sleep quality and for predicting associated deteriorations in daytime alertness (i.e., increased fatigue) resulting from nighttime sleep patterns. It is this latter fact that makes actigraphy so useful for conducting fatigue-risk assessments in applied settings.

### Mathematical modeling to predict effectiveness and risk

Since it is possible to track sleep and circadian information via wrist actigraphy, and since the manner in which sleep and circadian processes affect fatigue and performance is well-understood, it is possible to process sleep/wake data through mathematical models to: 1) predict the times at which performance is most likely to be safe and acceptable versus the times at which performance will be compromised; and 2) to establish the cumulative effects of different work/rest schedules on overall performance capability and overall accident risk (Mallis et al., 2004).

One such model is the Sleep Activity and Task Effectiveness (SAFTE) model (Hursh et al., 2004) which can accurately estimate the impact of scheduling factors and sleep history on both safety and productivity. The SAFTE model is currently the model of choice for the US Army and the US Air Force, and in comparison to other widely available models, the SAFTE model has been independently validated as the most accurate predictor of sleep-restriction on subjective fatigue ratings and objectively measured performance (Van Dongen, 2004). In addition, the SAFTE model has been shown to accurately predict the impact of sleep and scheduling factors on human-factors accident risk (Hursh et al., 2006) and the impact of the these factors on human factors accident severity (Hursh, personal communication, 2008).

The SAFTE model mathematically simulates the primary physiological processes that determine the level of fatigue (i.e., performance effectiveness) at any given point in time. It contains a circadian process that represents the manner in which the body clock influences both performance and sleep regulation; a sleep-reservoir process that represents the way in which recovery sleep is affected by hours of sleep, hours of wakefulness, current sleep debt, the circadian timing of sleep and any type of sleep fragmentation (awakenings during a period of sleep); and a sleep-inertia process that simulates the brief period of grogginess that occurs upon awakening from sleep. The model calculates performance effectiveness based on the current balance of these three processes in a manner generally similar to the homeostatic model of Folkard and Akerstedt (1991).

As noted earlier, performance effectiveness follows the circadian rhythm such that it peaks in the early evening (around 2000 hours), and troughs in the early morning (about 0400 hours). Performance likewise is affected by the level of the sleep reservoir such that a low reservoir is typically accompanied by low performance. Sleep propensity is negatively correlated with alertness. These rhythmic patterns are temporarily upset when people move to another time zone or alter their work schedules, and the SAFTE model also mimics this process via representation of a temporary performance degradation that accompanies readjustment while simultaneously gradually realigning the circadian rhythm with the new activity pattern. In order for the model to make the best predictions, accurate sleep/wake information is required, and in applied settings this information can best be provided by correctly-processed actigraphy data.

## Validation of the Fatigue Science Actigraph-Based Sleep/Wake Scoring Method

Given that 1) the connection between sleep/circadian factors and fatigue has been well-established, that 2) actigraphy in general has been validated as an acceptable method for measuring sleep/wake patterns and sleep quality in normal healthy individuals and in people with certain types of sleep problems, and that 3) the SAFTE model has been validated as an accurate predictor of performance effectiveness and fatigue risk provided that accurate historical sleep/wake information is available, the only remaining issue is the degree to which a specific actigraph and its associated data-processing algorithms are capable of accurately reflecting the actual sleep/wake patterns of typical individuals.

The present report will focus upon the validation of the Fatigue Science actigraph (the ReadiBand™) coupled with the patented Fatigue Science actigraph data-processing strategy. The process is directly compared to other sleep/wake algorithms including those approved by the Standards of Practice Committee of the American Academy of Sleep Medicine (Littner et al, 2003) for use in the judging of sleep quality. This is a critical component of operational fatigue-risk management since accurate sleep/wake information is essential for accurate fatigue-prediction modeling.

**The basics of actigraphy**

Actigraphy involves the utilization of a small portable device (in this case, a wrist-worn accelerometer) to sense arm movements a number of times per second and to store this information for subsequent analysis. This is accomplished via a small accelerometer capable of sensing movement along any one of three axes. The accelerometer is sampled 16 times per second, each second the actigraph is worn, to ensure high resolution. Each time a limb movement occurs, the accelerometer registers this information and stores it in an adjacent memory chip. The Fatigue Science actigraph can continuously record actigraphic data for a period of up to 30 days, but shorter recording episodes are more typically used. Once the desired data collection period has elapsed (i.e., 1 to 30 days), the data are downloaded onto a personal computer, and the number of zero crossings (number of times during which any type of limb movement occurred) per 1-second period is calculated for each second the actigraph was worn. This information is subsequently passed to classification algorithms which determine whether the level of movement is indicative of a typical awake state during which the actigraph wearer is going about normal daily activities, or a typical sleep state during which the wearer is lying down asleep. Actigraphy testing has been used in a variety of research studies to evaluate sleep/wake cycles in patients and normals, to determine circadian rhythm information, and to determine the effects of interventions designed to enhance sleep. The basic assumption underlying actigraphy testing is that people will physically move about more frequently when they are up and awake versus when they are in bed asleep. Furthermore, it is assumed that people who are in bed will move less when they are sleeping soundly than when they are experiencing sleep disturbances associated with environmental or other disruptive factors.

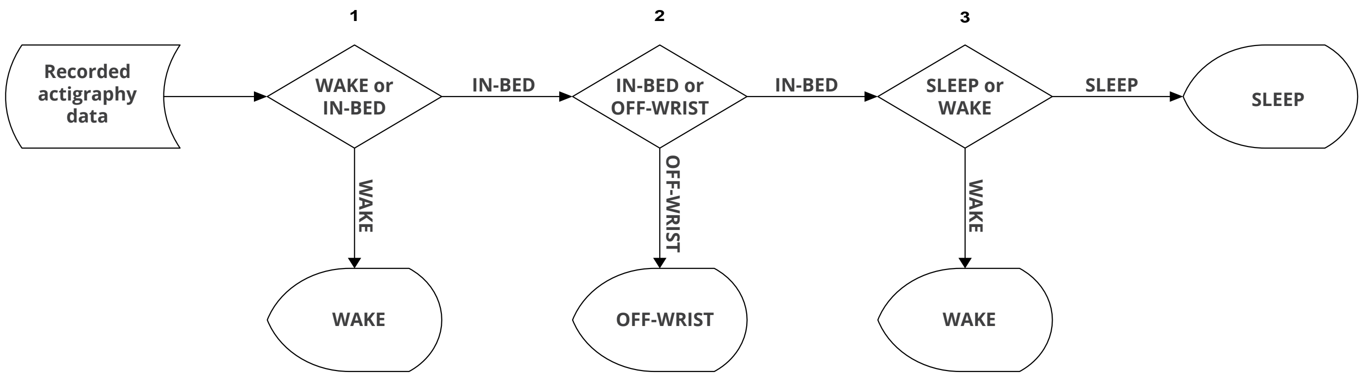
Actigraphy cannot classify sleep/wake activity as accurately as polysomnography because sleep is in fact an empirically determinable physiological state, and actigraphy is only capable of inferring the presence or absence of this state through a surrogate measurement (i.e., physical movement) rather than through direct examination. Polysomnography directly provides physiological sleep/wake information based on electroencephalographic, electromyographic, and electrooculographic events. Thus, it is always possible that a patient could "fool the actigraph" by quietly lying in bed (appearing to be asleep) while his/her brain was actually actively awake, his/her eyes were open and moving, and his/her muscles were taut but still. And this in fact is the challenge of accurately scoring sleep/wake states via actigraphy, and is in fact that reason that some researchers claim the specificity of actigraphy is too low to accurately detect wakefulness during sleep (Pollak, Tyron, Nagaraja, & Dzwonczyk, 2001; Paquet, Kawinska, & Carrier, 2007). In other words, their concern is that actigraphy-based sleep/wake detection algorithms tend to score many of a wearer's wake periods as sleep. Nonetheless, the Standards of Practice Committee of the American Academy of Sleep Medicine (Littner et al, 2003) has concluded that, generally speaking, actigraphy is reliable and valid for judging the quality, quantity, and timing of sleep in normal, healthy populations (a position that was reiterated in the Academy's more recent position paper (Morgenthaler et al, 2007).

**Validation studies for the Fatigue Science actigraph**

Although actigraphy is generally accepted as a valid way to record sleep/wake information in environments which preclude polysomnography, it should be noted that there are a number of specifically different actigraphs on the market today, and each actigraph relies on different internal components and different data processing strategies to make accurate sleep/wake determinations. In the present report, we will offer evidence that our approach provides improved sleep/wake determination in comparison to other currently-available actigraphic collection and processing systems, addressing the arguments of those opposed to actigraphy by specifically demonstrating the improved specificity of our algorithm. In summary, the purpose of the present report is to document the validation of the Fatigue Science actigraph—The ReadiBand™.

The validation effort consisted of two separate studies. The first study focused on ensuring that the automated data-processing algorithm accurately differentiated between times when the wearer of the actigraph was awake and moving about (characterized by relatively high activity scores) versus times when the wearer was either lying in bed or had entirely removed the actigraph from his/her wrist (both of which are characterized by relatively low activity scores). Essentially, the desired goal is to automatically label four states: 1) Wearer awake and out of bed, 2) ReadiBand off wrist, 3) Wearer awake and in bed, and 4) Wearer asleep and in bed. The most important states in terms of calculating fatigue risk are amount of wake and amount of sleep and the consistency and timing of the sleep/wake cycle. Algorithms were developed to make these determinations, and the accuracy or validity of these algorithms were determined via two studies. The first study ensured the accuracy of making the determination “Awake and out of bed” versus “In bed—Awake or asleep” versus “ReadiBand off wrist” (steps 1 and 2 as shown in figure 1). The second study ensured that the “In bed” data were correctly classified as “In bed—Awake” versus “In bed—Asleep” (step 3 in figure 1). Each of these studies will be discussed in turn.

**Figure 1.** High level overview of the three step state determination logic




**Modeling WAKE-AND-MOVING-ABOUT, OFF-WRIST, and IN-BED states (Study 1).**

Overview: The first portion of the sleep scoring algorithm developed for the Fatigue Science ReadiBand was designed to automatically extract one of the following states based solely on actigraph data: 1) Awake and moving about, 2) In bed—Awake or Asleep, or 3) Actigraph off wrist. The first validation study ensured agreement between the results of human-scoring of the data versus the results of algorithmic scoring of the data. High classification accuracy for these states improves the subsequent sleep/wake detection algorithm. Many validation efforts for actigraphy compare only the actigraphy data that is collected during the periods in which polysomnography is collected, i.e., the typical empirically-determined sleep periods. However, this approach is flawed due to the fact that actigraphy is generally collected 24 hours a day (when the wearer is asleep and awake, in bed and out of bed, etc.) for periods exceeding 2-weeks to assess routine sleep/wake cycles, sleep quantity and quality, and fatigue risk. Algorithms for scoring sleep and associated fatigue risk must account for the entire period of evaluation and not merely the sleep periods which represent only a fraction of the available data. Napping in the midst of what might be thought of as typical awake periods is considered a recommended fatigue countermeasure and the sleep that occurs during these naps should be considered in any fatigue-risk or generalized daily-sleep assessment.

Subjects: A mixed collection of de-identified actigraph data previously acquired by Fatigue Science for software/hardware functionality tests from 180 participants wearing the sleep band for four or more consecutive days was used.

Apparatus: Actigraphy data were collected from the Fatigue Science ReadiBand actigraph. The wrist-worn ReadiBand, which contains a 3D accelerometer sampled at 16 Hz, a storage chip, and a 1.5 V battery, was worn on the non-dominant wrist continuously for 24-hour data collection periods of various durations (i.e. one or more days).

Procedure: The first validation was to assess agreement between human-scored and algorithmically-scored states of 1) Awake and moving about, 2) In bed—Awake or Asleep, or 3) Actigraph off wrist. Two data analysts visually scanned each five minute epoch of actigraphic data and independently classified each into one of the three categories. Wherever a discrepancy arose, a third analyst adjudicated. This manual scoring methodology was considered the reference standard.

Results: The first step in the algorithm involves the decision as to whether there is sufficient activity (i.e., zero crossing detected by the actigraph's internal accelerometer) to judge the wearer of the actigraph as "Awake and moving about." If not, a decision was required as to whether the wearer of the actigraph was lying in bed (Awake or Asleep) or whether the wearer had removed the actigraph from the wrist. OFF-WRIST data is likely to contain activity epochs of very low motion as is the case with quiet IN-BED data epochs, and these two states may easily be confused. Thus, it is important to substantiate that the initial classification step accurately differentiates among the three relevant categories. Table 1 shows the classification accuracy for the states WAKE, IN-BED, and OFF-WRIST as determined in the first stage of the sleep/wake algorithm using our proprietary combination of smoothing, thresholding, and dilation procedures.

**Table 1.** Identification accuracy of the three state model—the first step in the sleep/wake scoring

| AWAKE, up and about | IN BED, awake or asleep | OFF WRIST |
|---|---|---|
| 95% | 95% | 87% |

**Modeling WAKE and SLEEP states during IN-BED state (Study 2).**

Overview: Once the algorithm determined that the wearer of the actigraph was in fact lying in bed rather than moving about, the next requirement was to determine whether he/she was IN-BED Awake or IN-BED Asleep. The second validation study (reported here) assessed the accuracy with which the algorithm's automatic determination of these two states agreed with concurrent polysomnographic determination of these two states. As noted previously, polysomnography is the "gold standard" for making sleep/wake determinations.

Subjects: A total of 50 participants were tested. These individuals were patients who had been referred to the Kettering Health Network Sleep Disorders Center for a clinical sleep evaluation. Table 2 describes the diagnostic determinations of the participants, polysomnographically-derived mean wake time during sleep, polysomnographically-determined sleep efficiency (percentage of time in bed actually sleeping), and a self reported sleepiness rating. As can be seen below, the patients in the present comparison averaged over ninety minutes of wake time during their night of sleep. This provided adequate in-bed-awake (in comparison to in-bed-asleep data) for assessing the specificity of our sleep scoring algorithm.

Apparatus: For the conduct of the polysomnographic study, a standard sleep-study montage of sensors was used. For the scoring and staging of sleep, EEG sensors referenced to contra-lateral mastoids (A1 and A2) were placed at C3, C4, O1, and O2; submental EMG electrodes were attached under the chin; and EOG electrodes were attached to the outer canthus of left and right eyes. Although not utilized for the purposes of algorithmic validation, the patients were also outfitted with ECG electrodes, respiration transducer bands, nasal/oral airflow thermistor, leg EMG electrodes, and a pulse oximetry sensor to determine the presence, absence, and type of any existing sleep disorder . Electrodes were filled with an electrolyte cream and taped or collodioned to the skin. Data were recorded on a Sandman Sleep Diagnostic System for subsequent digital storage and visual inspection. The Sandman system digitizes data at a sampling rate suitable for Fast Fourier Analysis. For the recording of actigraphic data, the Fatigue Science ReadiBand actigraphs were used. As noted earlier, these actigraphs are the size of a typical wrist-watch and contain a 3D accelerometer, a storage chip, and a 1.5 V battery. A ReadiBand was strapped to the wrist and ankle of each participating patient via a wrist-watch style plastic band.

Procedure: Upon arrival to the Sleep Disorders Center, patients were outfitted with polysomnographic recording electrodes. First the electrode placement sites were determined via a standard measurement procedure. Second, the skin was cleansed with a mild abrasive solution and/or alcohol to ensure low recording impedance. Third, the electrodes were attached to the skin for the all-night sleep recording.

**Table 2**. Sleep Diagnostics for Patients - **Fifty patients participated however some patients had multiple diagnoses therefore the second column does not sum to fifty.

| Diagnosis | N (% of patients) | Age (yrs) | Sleep Efficiency Index (%) | Wake Time (min) | Epworth Sleepiness Scale (out of 24) |
|---|---|---|---|---|---|
| 327.21 Central Sleep Apnea (positional) | 11 (22%) | 53.82 +/- 13.71 | 80.15 +/- 13.52 | 79.42 +/- 49.01 | 10.00 +/- 5.55 |
| 327.23 Obstructive Sleep Apnea (positional) | 28 (56%) | 51.64 +/- 13.99 | 81.27 +/- 10.45 | 77.65 +/- 40.96 | 13.29 +/- 5.64 |
| 307.44 Insufficient Sleep Syndrome | 7 (14%) | 35.57 +/- 12.73 | 81.13 +/- 11.91 | 86.81 +/- 53.57 | 10.60 +/- 6.23 |
| 327.52 Periodic Limb Movement | 13 (26%) | 47.38 +/- 13.04 | 72.41 +/- 14.57 | 116.6 +/- 49.16 | 9.18 +/- 5.74 |
| 333.94 Restless Leg Syndrome | 4 (8%) | 49.5 +/- 14.82 | 76.70 +/- 12.8 | 97.98 +/- 65.19 | 13.25 +/- 7.85 |
| 327.12 Idiopathic Hypersomnia w/ Long Sleep Time | 2 (4%) | 41.00 +/- 12.73 | 89.35 +/- 10.25 | 56 +/- 54.45 | 16.50 +/- 0.71 |
| 327.14 Hypersomnia associated w/ Medical Condition Snoring | 1 (2%) | 36 | 76.7 | 107.5 | 10 |
| 327.54 Sleep Related Bruxism | 7 (14%) | 50.14 +/- 15.17 | 79.2 +/- 11.71 | 91.3 +/- 47.54 | 11.8 +/- 4.32 |
| 307.42 Psycho-physiological Insomnia | 1 (2%) | 36 | 43.1 | 217 | 3 |
| 347.11 Narcolepsy | 2 (4%) | 35.00 +/- 7.07 | 86.20 +/- 6.22 | 60.00 +/- 29.7 | 9.00 +/- 2.83 |
| 327.31 Delayed Sleep Phase | 3 (6%) | 33.67 +/- 3.21 | 54.53 +/- 23.72 | 172.1 +/- 78.9 | 6.00 +/- 4.36 |
| 788.36 Sleep Enuresis | 1 (2%) | 56 | 76.3 | 89 | N/A |
| V69.8 Inadequate Sleep Hygiene | 1 (2%) | 35 | 38.7 | 218.3 | 4 |
| 327.36 Shift Work Sleep Disorder | 3 (6%) | 47.00 +/- 1.00 | 76.93 +/- 8.53 | 103.47 +/- 57.3 | 15.33 +/- 4.04 |
| 327.53 Leg Cramps | 2 (4%) | 58.00 +/- 15.56 | 81.05 +/- 4.60 | 75.5 +/- 14.14 | 20.00 +/- 1.41 |
| 327.43 Sleep Paralysis | 1 (2%) | 48 | 68 | 167.6 | 16 |
| | | | | | |
| **Totals** | **50**** | **47.97 +/- 13.78** | **77.45 +/- 13.92** | **94.38 +/- 53.09** | **11.49 +/- 5.81** |

Before or during the standard polysomnographic preparations, the sleep technician asked the patient whether he/she was willing to wear two wrist actigraphs (one on the wrist and one on the ankle) in addition to the other sensors described above. Patients who agreed to this additional non-invasive procedure were provided with an informed consent agreement which they were asked to sign. Any questions about the study were answered at this time. Once informed consent was obtained, the actigraphs were strapped to the patient's wrist and ankle, and the patient was escorted to the bedroom in which he/she slept during the study. Recording electrodes were connected to the Sandman Sleep System, and the lights were turned out and recordings begun. The following morning, all electrodes and actigraphs were removed, and the patient was allowed to leave the Sleep Disorders Center. Once the patient departed, the actigraph data were downloaded for storage and transmission along with the digitized polysomnographic data and sleep summary.

Data analysis: The accuracy of the algorithmic IN-BED Sleep/Wake determinations was established as follows. First each patient's polysomnographically-determined hypnogram was segmented into consecutive 5-minute epochs which, although originally classified as one of several sleep stages versus awake or movement, was subsequently reclassified simply as either awake or asleep. Some epochs contained periods of both wake and sleep and were classified as the predominant state. Next these polysomnographically-derived hypnogram epochs were aligned with the algorithmically-scored five-minute data epochs recorded by the wrist actigraph. The ankle actigraph data was not used in this study. Five algorithms for sleep scoring were compared sample by sample using the same data and each of these was directly scored against the polysomnographic results. The algorithms compared were:
(1) Respironics Actiwatch® L20 algorithm, (2) Respironics Actiwatch® L40 algorithm, (3) Cole-Kripke algorithm (Cole et al., 1992), (4) Lötjönen algorithm (Lötjönen et al., 2003), and (5) our Fatigue Science algorithm. Finally, the accuracy of each algorithm was computed using the following statistical measures:

- **Sensitivity**: This is the true positive rate and determines how good the actigraph and algorithm is at detecting SLEEP when PSG also detects SLEEP.
- **Specificity**: This is the true negative rate and determines how good the actigraph and algorithm are at correctly excluding SLEEP when the subject is not asleep as determined by PSG.
- **Positive predictive value (PPV)**: This is the post-test probability of a positive test. If the actigraph and algorithm indicate the time period under consideration is SLEEP, this determines the probability that it really was SLEEP.
- **Negative predictive value (NPV)**: This is the post-test probability of a negative test. If the actigraph and algorithm indicate that the time period was not SLEEP, this determines the probability that it really was not SLEEP.
- **Overall accuracy**: This is the proportion of all actigraph and algorithm determinations that yielded correct results according to PSG.

Results: The data that had accurately been labeled as the IN-BED state using the algorithm validated in study 1 was subsequently classified into IN-BED Awake or IN-BED Asleep states. The results of the automated sleep wake scoring algorithms (described earlier) on the data from the 50 subjects in the PSG sleep study are shown in Table 3. Combining these results with the results of IN-BED classification, the estimated 24-hour sleep scoring accuracy of the combined model is 93%. This number is derived from a weighted combination using the overall sleep detection accuracy (sensitivity) of 88% from the sleep/wake model developed in Study 2 and IN-BED detection accuracy of 95% from Study 1 to provide an overall accuracy score for sleep periods during a 24-hour period (Other algorithms and actigraphs have been validated for sleep periods using PSG and not during normal wake periods).

**Table 3**: Classification accuracy measures for SLEEP/WAKE states.

| Algorithm | Total accuracy | Sensitivity | Specificity | PPV | NPV |
|---|---|---|---|---|---|
| Actiwatch®-L20 | 80% | 97% | 29% | 79% | 25% |
| Actiwatch®-L40 | 78% | 99% | 20% | 88% | 25% |
| Cole-Kripke | 78% | 99% | 19% | 91% | 23% |
| Lötjönen | 81% | 92% | 42% | 74% | 36% |
| Fatigue Science | 82% | 88% | 55% | 69% | 43% |

## Discussion

Fatigue is increasingly recognized as a safety hazard in transportation and industrial sectors, but fatigue management efforts have been hampered by an inability to quantify the extent and sources of fatigue. Unlike methods designed to screen for other work-place hazards such as on-the-job alcohol intoxication, there is no “Breathalyzer™” for fatigue and no universally-recognized physiological or biochemical marker for fatigue. However, thanks to many years of research on sleep and circadian rhythms coupled with recent advances in mathematical modeling of physiological processes, the impact of various sleep and duty schedules on performance effectiveness and fatigue risk can be determined. But it is first important to accurately track the sleep/wake behavior of individuals so that the fatigue-analysis model starts with an accurate understanding of individualized fatigue-producing factors. In a laboratory context, this tracking would be accomplished via polysomnography, but in day-to-day operations, wrist actigraphy is a better-suited alternative.

This paper established the validation of the Fatigue Science actigraph for determining sleep/wake periods, sleep quantity, and sleep quality for the purposes of subsequently feeding the SAFTE model for predicting fatigue risk. As described earlier, the SAFTE model already has been validated in several previous studies. Thus, the novel portion of this report centers upon the validation of the sleep/wake classification models used to convert wrist-actigraphy data from the Fatigue Science actigraph into usable fatigue-component inputs (sleep duration, sleep quality, and sleep/waking timing) for SAFTE.

The Fatigue Science actigraph was 93% accurate in determining sleep scoring when contrasted to results derived from sleep scoring using “gold-standard” polysomnography. With regard to scoring sensitivity alone, these results are consistent with studies using different actigraphs and scoring algorithms in prior PSG validation studies. Sedah, et.al. (1994) reported a 91-93% agreement with PSG using the AMI-32 (Ambulatory Monitoring Inc., Ardsley, NY) coupled with AMI’s automatic sleep scoring algorithm, and Cole, et. al. (1992) obtained an 88% classification accuracy using the Cole-Kripke algorithm and the Ambulatory Monitoring, Inc (AMI) device. However, while both of these algorithms were highly sensitive in detecting actigraphically-determined sleep, they yielded poor specificity in terms of classifying wake episodes that occurred during the sleep period. For instance the specificity of the Cole-Kripke algorithm as reported in the 1992 assessment was only 34%. In our direct comparison, the Cole-Kripke algorithm only detected 19% of the wake periods. Other algorithms performed better with 42% specificity for the Lötjönen algorithm; 29% for the Actiwatch®-L20; and 20% for the Actiwatch®-L40. Finally, the Fatigue Science approach scored wake episodes during sleep (specificity) with a 55% classification accuracy.

Specificity is important when judging the fatigue levels of operational personnel because recent sleep is a primary determinant of an individual's level of fatigue. If there are wakeup periods during sleep, it is important to correctly detect and record these periods. Otherwise, the total amount of daily sleep is overestimated and the individual's fatigue level is underestimated. In the present investigation, the Fatigue Science algorithm outperformed the other algorithms in this regard. All the other algorithms falsely reported that subjects were asleep during periods in which PSG indicated they were actually awake at a greater rate than was the case with Fatigue Science algorithm. In general terms, the Fatigue Science algorithm and actigraph showed at least a 13% improvement over these previous methods in detecting wake periods. This at first may seem like an inconsequential improvement, but it can have a rather serious impact on subsequent fatigue-risk predictions.

We can illustrate this impact using model comparisons based on the polysomnographic data used to validate the Fatigue Science model in this investigation. Recall that specificity was calculated for each of the models, resulting in the following wake accuracy classification values (ordered from best to worst): 55% for the Fatigue Science algorithm; 42% for the Lötjönen algorithm; 29% for Actiwatch®-L20; 20% for Actiwatch®-L40; and 19% for the Cole-Kripke algorithm. To illustrate the impact of these differences on fatigue-risk calculations, we considered a 7-day situation in which a hypothetical subject spent 8 hours in bed each night, but only obtained 5 hours of actual sleep (while remaining awake for the other 3 hours in bed). Using the different levels of specificity associated with each of the classification models examined in the present study (see figures 2-6), the 3 hours of actual in-bed awake periods would have been incorrectly scored as sleep for: 1 h 15 m by the Fatigue Science model; 1 h 45 m by Lötjönen; 2 h by Actiwatch®-L20; 2 h 30 m by Actiwatch-L40; and 2 h 45 m by Cole-Kripke. Processing these sleep results through the SAFTE model shows that these different amounts of sleep accumulated across a 7- day period would have resulted in the following performance-effectiveness scores at the end of the 7th day: Fatigue Science- 79%; Lötjönen-83%; Actiwatch®-L20-85%; Actiwatch®-L40-88%; and Cole-Kripke- 90%. Contrast these with Figure 7 which shows the performance effectiveness scores for correctly classifying 5 hours of sleep and 3 hours of wake periods in an 8 hour sleep period. The performance effectiveness using our hypothetical 'ground truth' is 68%. Note that since effectiveness levels less than 80 have been associated with substantially elevated fatigue-risk levels, the only modeling and classification result that would be actionable based on this hypothetical situation (where the subject in fact obtained only 5 hours of sleep per night) would have been the result obtained with the Fatigue Science approach. With every other classification model, the level of fatigue risk associated with 3- hours of chronic nightly sleep restriction would have been considered inconsequential (i.e., the final performance effectiveness scores would have been greater than 80%). Such a result is cause for concern in cases where an actigraphy/modeling-based fatigue-risk management system is being employed to safeguard operational safety in safety-sensitive occupations.

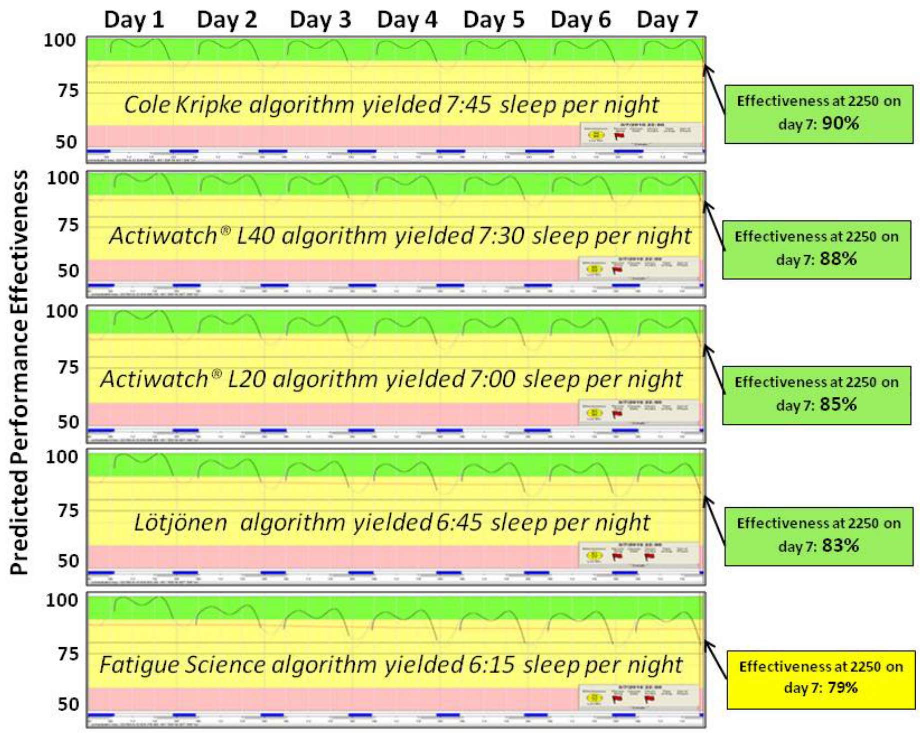



**Figure 2.** The figure contains a series of moment-by-moment performance effectiveness curves (the dark undulating lines in each panel) generated by the SAFTE model based on the sleep/wake calculations generated from the different actigraphy classification algorithms. A 7-day period is graphed for illustrative purposes, with the central point of interest being the effectiveness calculation at the end of the 7th day. Since sleep history and sleep/wake timing are the two primary inputs to the SAFTE model, overestimating the amount of sleep (due to algorithm inaccuracies) makes a difference in the effectiveness scores--yielding overly-optimistic estimates of performance capabilities. Comparisons among the algorithms showed that actigraphy data processed via the Cole Kripke approach was the least sensitive to in-bed wake-up episodes which led to an overestimation of nightly sleep and a consequent greater overestimation of end-of-the-week effectiveness. Conversely, since the Fatigue Science approach more accurately tracked sleep/wake episodes, the model is less likely to overestimate sleep, and more likely to generate a more conservative effectiveness score. Note that effectiveness scores below 80 have been correlated with impaired performance. Thus, although by a very narrow range, the Fatigue Science algorithm was the only algorithm that yielded an actionable score at the end of the simulated work week.

## Conclusion

The use of almost any actigraph-based sleep/wake assessment strategy is better for deriving the input data for fatigue-prediction models than reliance upon self-reported sleep histories (which are typically highly inaccurate). And the use of an objective and validated fatigue-prediction model overcomes the difficulties associated with self-assessments of fatigue (which also have been proven highly unreliable). However, it is essential that actigraphy classification accurately discriminates periods of in-bed awake from periods of in-bed sleep since the level of the sleep reservoir is one of the two primary influences on fatigue-risk predictions. Coupling the Fatigue Science actigraphy approach with an accurate fatigue-risk assessment model such as the Fatigue Science SAFTE, provides an operationally-useful fatigue- measurement strategy that will facilitate fatigue-risk management systems in transportation, industrial, and other contexts.

## REFERENCES


Akerstedt TE. Work hours, sleepiness and accidents. J Sleep Res 1995;4(2):15-22.

Akerstedt, T., and Folkard, S. (1996). Predicting duration of sleep from the three process model of regulation of alertness. Occupational and Environmental Medicine, 53(2): 136–141.

Akerstedt T, Haraldsson PO. International consensus meeting on fatigue and the risk of traffic accidents. The significance of fatigue for transportation safety is underestimated. Lakartidningen 2001;98:3014-17.

American Sleep Disorders Association (1995). Practice parameters for the use of actigraphy in the clinical assessment of sleep disorders. Sleep, 18(4):285-287.

Balkin, T.J., Rupp, T., Picchioni, D., and Wesensten, N.J. (2008). Sleep loss and sleepiness. Chest, 124(3):653-660.

Balkin T, Thorne D, Sing H, Redmond D, Wesensten N, Williams J et al. Effects of sleep schedules on commercial motor vehicle driver performance (DOT Report No. DOT-MC-00-133). Washington DC: Department of Transportation Federal Motor Carrier Administration 2000.

Banks S, & Dinges DF. Behavioral and physiological consequences of sleep restriction. J Clin Sleep Med 2007;3(5):519-528.

Barger LK, Ayas NT, Cade BE, Cronin JW, Rosner B, Speizer FE, & Czeisler CA. Impact of extended- duration shifts on medical errors, adverse events, and attentional failures. PLos Med 2006;3(12):e487.

Bonnet MH. Sleep deprivation. In Kryger M H, Roth T, & Dement W C (Eds.), Principles and practice of sleep medicine, 2nd ed. Philadelphia: W. B. Saunders Company 1994:50-67.

Caldwell, J. A., Caldwell, J. L., and Schmidt, R. M. (2008). Alertness management strategies for operational contexts, Sleep Medicine Reviews, 12:257-273, 2008.

Carskadon MA, & Roth T. Sleep restriction. In Monk TH (Ed.), Sleep, sleepiness and performance. New York: John Wiley & Sons 1991.

Carusa, C.C., Bushnell, T., Eggerth, D., Heitmann, A., Kojola, B., Newman, K., Rosa R.R., Sauter, S.L., Vila, B. (2006). Long work hours, health and safety: Toward a national research agenda. American Journal of Industrial Medicine, 49:930-942.

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. Sleep, 15(5), 461-469.

Dembe AE, Erickson JB, Delbos RG, & Banks SM. The impact of overtime and long work hours on occupational injuries and illnesses: New evidence from the United States. Occup Environ Med 2005;62(9):588-597.

Dinges DF. Probing the limits of functional capability: The effects of sleep loss on short-duration tasks. In Broughton RJ, & Ogilvie RD (Eds.), Sleep, arousal, and performance. Boston: Birkhauser 1992:177-188.

Dorrian J, Hussey F, & Dawson D. Train driving efficiency and safety: Examining the cost of fatigue. J Sleep Res 2007;16(1):1-11.Horne JA. Sleep loss and “divergent” thinking ability. Sleep 1988;11:528-536.

Horne JA. Human sleep, sleep loss and behaviour: Implications for prefrontal cortex and psychiatric disorder. Brit J Psychiat 1993;162:413-419.

Hursh, S.R., Raslear, T.G., Kaye, A.S., and Fanzone, J.F. (2006). Validation and calibration of a fatigue assessment tool for railroad work schedules, summary report. U.S. Department of Transportation Report No. DOT/FRA/ORD-06/21. Washington DC: U.S. Department of Transportation.

Hursh, S.R., Redmond, D.P., Johnson, M.OL., Thorne, D.R., Belenky, G., Balkin, T.J., Storm, W.F., Miller, J.C. and Eddy, D. (2004). Fatigue models for applied research in warfighting. Aviation, Space, and Environmental Medicine, 75(3 Suppl): A44-A53.

Jacobs, J.A., Gerson, K. (2004). The time divide: work, family, and gender inequality. Cambridge, MA: Harvard University Press.

Kirsch AD. Report on the statistical methods employed by the U.S. federal aviation administration in its cost/benefit analysis of the proposed flight crew member duty period limitations, flight time limitations and rest requirements. Washington, DC 1996: Report No.: Docket No. 28081.

Koslowsky M, & Babkoff H.. Meta-analysis of the relationship between total sleep deprivation and performance. Chronobiol Int 1992;9(2):132-136.

Lötjönen, J., Korhonen, I., Hirvonen, K., Eskelinen, S., Myllymäki, M., & Partinen, M. (2003). Automatic sleep-wake and nap analysis with a new wrist worn online activity monitoring device vivago WristCare. Sleep, 26(1), 86-90.

Mallis, M.M., Mejdal, S., Nguyen, T.T., and Dinges, D.F. (2004). Summary of the key features of seven biomathematical models of human fatigue and performance. Aviation, Space, and Environmental Medicine, 75(3 Suppl): A4-A14.

Mathis, J., and Hess, C.W. (2009). Sleepiness and vigilance tests. Swiss Medical Weekly, 139(15-16): 214-219.

Merilahti, J., Saarinen, A., Parkka, J., Antila, K., Mattila, E., and Ilkka, K. (2007). Long-term subjective and objective sleep analysis of total sleep time and sleep quality in real life settings. Proceedings of the 29th Annual International Conference of the IEEE EMBS, Lyon, France. August, 2007. Pp 5202-5205.

Morgenthaler, T., Alessi, C., Friedman, L., Owens, J., Kapur, V., Boehlecke, B., Brown, T., Chesson, A., Coleman, J., Lee-Chiong, T., Pancer, J., Swick, T.J. (2007). Practice Parameters for the Use of Actigraphy in the Assessment of Sleep and Sleep Disorders: An Update for 2007. Sleep, 30(4), 2007.

Naitoh P. Sleep deprivation in humans. In Venables PH, & Christie MJ (Eds.), Research in Psychophysiology. London: John Wiley 1975.

National Highway Traffic Safety Administration. NHTSA vehicle safety rulemaking and supporting research priorities 2005-2009. 2005; http://www.nhtsa.gov/cars/rules/rulings/PriorityPlan-2005.html.

National Sleep Foundation (2008). Sleep in America poll, summary of findings. Washington, DC: National Sleep Foundation.

Ozminkowski, R.J., Wang, S., and Walsh, J.K. (2007). The direct and indirect costs of untreated insomnia in adults in the United States. Sleep, 30(3):263-273.

Paquet, J., Kawinska, A., & Carrier, J. (2007). Wake Detection Capacity of Actigraphy During Sleep. Sleep, 30(10), 1362-1369

Pollak, C. P., Tryon, W. W., Nagaraja, H., & Dzwonczyk, R. (2001). How accurately does wrist actigraphy identify the states of sleep and wakefulness? Sleep, 24(8), 957-965.

Rosekind MR. Underestimating the societal costs of impaired alertness: Safety, health and productivity risks. Sleep Med 2005;6(suppl 1):S21-S25.

Roth, T. (2004). Measuring treatment efficacy in insomnia. Journal of Clinical Psychiatry, 65 (suppl 8): 8- 12.

Roth, T. (2007). A physiologic basis for the evolution of pharmacotherapy for insomnia. Journal of Clinical Psychiatry, 68(Suppl 5):13-18.

Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: an empirical test of methodological issues. Sleep, 17(3), 201-207

Sadeh, A., and Acebo, C. (2002). The role of actigraphy in sleep medicine. Sleep Medicine Reviews, 6(2):113-124.

Stoller, M.K. (1994). Economic effects of insomnia. Clinical Therapeutics, 16(5):873-897.

Thase, M.E. (2005). Correlates and consequences of chronic insomnia. General Hospital Psychiatry, 27:100-112.

Van Dongen, H. (2004). Comparison of mathematical model predictions to experimental data of fatigue and performance. Aviation, Space, and Environmental Medicine, 75(3 Suppl): A15-A36.

Walsh, J., Ustun, B. (1999). Prevalence and health consequences of insomnia. Sleep, 22 (suppl 3): S427- S436.