Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

11

MANAGING FATIGUE BY DROWSINESS DETECTION: CAN TECHNOLOGICAL PROMISES BE REALIZED?

David F. Dinges, Ph.D. and Melissa M. Mallis
University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

INTRODUCTION

This chapter draws on previous reports (Dinges, 1995b, 1997), as well as work in progress (Mallis et al., in preparation) to review key issues, and present illustrative data regarding the validation of technologies claiming to provide on-line monitoring of operator alertness and vigilance. Throughout the chapter the term "fatigue" is used to refer to the effects on performance capability of sleep loss, night work, prolonged work, or inadequate recovery, alone or in combination.

Although scientific and applied initiatives to develop on-line measures of alertness/drowsiness and hypovigilance have a long history (O'Hanlon & Kelley, 1974; Dinges & Graeber, 1989; Brookhuis, 1995), this area has undergone renewed interest and intensified activity especially in the USA (Rau, 1996; Knipling, 1996), and in Europe (Brown, 1995, 1997) in the past 3 years (see also Dinges, 1995a,b, 1996, 1997). Reasons for this "investment" in technology to manage fatigue-related performance impairment can be found in all modes of transportation, but especially in commerical motor vehicle operations. A number of arguments have been put forward to justify development of drowsiness detection technologies.

REASONS FOR FATIGUE-DETECTION TECHNOLOGIES

Below we summarize the main reasons that fatigue-detection technologies have become attractive as one way to prevent drowsy-driving crashes. These reasons are not necessarily the opinions of the authors, although we believe that drowsy-driving technology development has merit, as long as validity and reliability are demonstrated. We do not concern ourselves in this chapter with the often-voiced admonition that technology development is questionable

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

because such drowsy-driving detection systems, even if valid and reliable, likely will encourage some drivers to continue driving in an impaired state due to the false sense of security the technology affords, or to coercion from employers, or to willful misuse of the technology (Evans, 1991; Brown, 1997). Misuse of a device or technology is a legitimate concern along with a range of other legal/ethical issues (see Dinges, 1997; Table 2 below), but a technology that can potentially enhance safety and save lives should not be prejudged based on speculations about users' ethics, whether realistic or exaggerated. (For example, most motor vehicles can be accelerated to dangerous speeds, but the fact that some drivers elect to speed is not used to justify placing acceleration governors on motor vehicles. Rather, education and laws exist to teach drivers how to properly use the acceleration potential of a motor vehicle.) Elsewhere, and below, we propose that education, legal, and policy standards for implementing fatigue-monitoring technologies be developed in parallel with the engineering and scientific development of the devices (Dinges, 1995b, 1997). This is essential, because the reasons for technology development in this area are compelling and indicate devices will be implemented whether or not their use is considered in public policy on road safety.

<u>Fatigue-related crashes are common and serious</u>. Since 1994, there has been growing evidence from industrialized countries that fatigue from varying combinations of sleep loss, night driving (i.e. circadian rhythms), and prolonged work time (i.e., wake time on task) contributes to substantial numbers of motor vehicle crashes. Although estimates and the methods they are based on vary widely, few dispute that the problem of drowsy driving is inadequately addressed. Fatigue has been estimated to be involved 2% to 23% of all crashes (cf., O'Hanlon, 1978; McDonald, 1984; Horne & Reyner, 1995; Knipling & Wang, 1995; Maycock, 1997); in 4% to 25% of single-vehicle crashes (cf., Wang & Knipling, 1994; Brown, 1995); in 10% to 40% of crashes on long motor ways (Shafer, 1993; Dinges 1995b); and in 15% of single-vehicle fatal truck crashes (Wang & Knipling, 1994). Fall asleep crashes are also very serious in terms of injury severity (Pack et al., 1995). In the USA, fatigue has been implicated as the most frequent contributor to crashes in which a truck driver was fatally injured (U.S. National Transportation Safety Board, 1990). Much of the focus on the putative role of sleepiness/drowsiness in traffic accidents has centered on single vehicle crashes, although there is no reason to believe that sleepiness is not also involved in multiple vehicle crashes. In the USA, single vehicle crashes in which no alcohol was involved account for more than a quarter of all fatal crashes, more than a quarter of all injury-only crashes, and more than a quarter of all property damage-only crashes (see Dinges, 1995b). While many factors can contribute to single vehicle non-alcohol-related crashes, the fact that they comprise 27% (i.e., 1.67 million crashes in 1993) of all motor vehicle crashes in the USA, suggests that fatigue may contribute to more of these crashes than current estimates allow.

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

Consequently, with so many persons driving fatigued, technologies that detect dangerous levels of sleepiness before a crash occurs are essential.

Subjective estimates of sleepiness are unreliable. Experiments have demonstrated that subjects cannot reliably predict when they are impaired to the point of having an uncontrolled sleep attack (i.e., microsleep) and/or a serious vigilance lapse (Dinges, 1989). Drivers know when they are experiencing sleepiness (Horne & Reyner, 1995), but they cannot necessarily translate those introspections into accurate predictions of how long their eyes are closed and whether they are missing signals, or when they will have an uncontrolled sleep onset while driving (Wylie et al., 1996; Brown, 1997). On the other hand, although self reports of sleepiness are highly influenced by contextual variables (Dinges, 1989, 1995b), drivers should know when they are experiencing heavy eyelids and head bobbing, which is likely past the point of impairment by drowsiness (Kribbs & Dinges, 1994). Hence, technology may offer the potential for an earlier and more reliable warning of performance-impairing sleepiness, before drowsiness leads to a catastrophic outcome.

Drowsiness-detection technology offers an alternative to proscriptive hours of service. Technology is viewed by some as a key component in a package of fatigue management options that can replace or at least put flexibility into federally-mandated proscriptive hours of service. For example, the current USA federal hours of service for commercial motor vehicle operators were written in 1939, and rely on work time as the primary determinant of fatigue (this is not unique to the trucking industry). It has been recognized for some time, however, that within limits, work duration accounts for only a modest proportion of accident risk (Hamelin, 1987). Thus, the current hours of service may not prevent many fatigue-related crashes, even when compliance is 100%. Fall-asleep crashes are more likely to occur during night driving and in sleep-deprived persons (e.g., Harris, 1977; Mitler et al., 1988; Pack et al., 1995). This is consistent with scientific studies in the past 30 years that have demonstrated that the level of waking alertness is regulated by two neurobiological forces that shape the time course of subjective fatigue and aspects of performance--the endogenous circadian rhythm and the need for sleep (Dinges, 1995b). When considered together and in combination with work hours, the product of these processes regulating fatigue and vigilance is nonlinear, temporally dynamic, and complex. This makes it complicated to derive regulatory schemes to prevent fatigue. Hence, technologies that monitor the driver's temporally dynamic state of alertness/drowsiness over time, are viewed as offering an advantage over proscriptive regulations. Such technologies may provide one of a number of ways to optimize safety through prevention of fatigue-related crashes, while permitting greater flexibility in work-rest scheduling to facilitate economic and related pragmatic goals, as well as drivers' personal choices.

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

Technological advances have made the goal feasible. Many technologies being developed for detecting drowsiness are miniaturized and unobtrusive (their durability and cost-effectiveness are less well established). Advances in electronics, optics, sensory arrays, data acquisition systems, algorithm development (e.g., neural nets), and other areas have made it far more likely that the goal of an affordable drowsiness detection system in a truck or automobile will be achieved and implemented in far less than the 10-20 years estimated by Brown (1995, 1997). In the USA, for example, there are currently many efforts underway at federal, industry, and entrepreneurial levels toward development of technologies for monitoring a driver's physiology or behavior in order to "manage" performance-impairment from fatigue in transportation. This marriage of technology and the human operator for drowsiness detection is part of a broader emphasis in the USA on development of "intelligent vehicle" and "driver condition warning" initiatives.

Operator-Centered Fatigue-Monitoring Technologies

Nearly all of the technologies currently being proposed to monitor on-line, the alertness or drowsiness or vigilance capability of a driver are in the prototypical development, validation testing, or early implementation stages. Their full effectiveness, practicality, and acceptance remain unproven scientifically and practically. Consequently, in this chapter we will not advocate for, endorse, or criticize any specific technology. However, we will demonstrate (below) the level of scientific standards that technologies must meet to warrant progression to the implementation stage. Reviewing the different categories of technologies being proposed for management of fatigue/sleepiness in transportation modes necessarily requires some categorizations. Technologies can be arbitrarily grouped by different criteria, but at least three broad categories of fatigue-related technologies (for detection and/or prevention) include: operator-centered technologies, system-centered technologies, and environmentally-oriented technologies (Dinges, 1997). Operator-centered technologies are the focus here.

Although many enthusiastic claims are made at meetings and in promotional materials regarding the merits of one type of operator-centered fatigue-monitoring technology or algorithm over another, such comparison data are only now coming to be generated (Mallis et al., in preparation). Moreover, the precise operational definition of what each technology is attempting to "detect" or the theoretical construct it purports to tap into, are often not well defined. In general, however, most technologies explicitly claim or imply detection of some aspect of either a heightened risk of operator error or outright impairment through one or more of the following hypothetical constructs: operator vigilance; operator

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

attention/inattention; operator alertness/drowsiness; operator microsleeps; operator hypovigilancce; operator performance variability; or operator vulnerability to error. It remains a matter for scientific inquiry to what extent these constructs overlap empirically. However, no distinctions will be made among them in this chapter. Generic categories of operator-centered technologies currently being developed or marketed include, but are not limited to, the following four classes. This chapter is particularly concerned with the fourth category.

1. Readiness-to-perform and fitness-for-duty technologies

The final concept in the above list of fatigue constructs--vulnerability to error--bridges the distinction between technologies that afford on-line monitoring of the driver and those that involve probed evaluation or temporally discrete sampling of biobehavioral variables and that fall in the category of "fitness for duty." Fitness-for-duty or readiness-to-perform approaches, which are becoming popular replacements for urine screens for drugs and alcohol, can involve sampling aspects of performance capability or physiological responses. Because these tests are increasingly becoming briefer and more portable, the developers are seeking to extend their use beyond prediction of functional capability at the start of a given work cycle (i.e., prediction of relative risk over many hours), to prediction of capability in shorter time frames (e.g., whether someone is safe to extend work time at the end of a shift or duty period).

These technologies are intended to provide some behavioral or biological estimate of an operator's functional capability for work yet to be performed, relative to a standard, such as the operator's idiosyncratic function when unimpaired, or relative to a group norm (Gilliland & Schlegel, 1993). While some biologically-based (primarily ocular and pupilometric) technologies for fitness for duty are currently available ("EPS-100" by Eye Dynamics, Inc.; "FIT" by PMI, Inc.; "PUPILSCAN" by Fairville Medical Optics, Inc.), most of the technologies in this area are performance-based (Gilliland & Schlegel, 1993; Daecher, 1996). There are a large number of performance test batteries touted as candidates for readiness-to-perform and/or fitness-for-duty testing. Unfortunately, many of them are aptitude- and language-skill sensitive, and many have rather dramatic learning curves, making them less than ideal candidates for repeated usage in a diverse population. In addition, many have not been validated to be sensitive to fatigue, are not predicated on a model of human performance failure due to fatigue, and do not provide criteria by which to determine when someone is dysfunctional. On the other hand, there are a few simple performance tests that have been deployed (e.g., Rosekind et al., 1994; Dawson & Reid, 1997) that do not appear to have some of the aptitude- and language-sensitive problems of other tests, and that have been validated

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

to be sensitive to sleep loss and circadian variation (Kribbs & Dinges, 1994; Dawson & Reid, 1997), suggesting that there is potential for this approach, both as a readiness-to-perform test and a way of probing functional capability while on the job.

2. Mathematical models of alertness dynamics joined with ambulatory technologies

This approach involves application of mathematical models that predict operator alertness/performance at different times based on interactions of sleep, circadian, and related temporal antecedents of fatigue (e.g., Akerstedt & Folkard, 1994; Belenky, 1997; Jewett, 1997). This is the subclass of operator-centered technologies that includes those devices that seek to monitor sources of fatigue, such as how much sleep an operator has obtained (via wrist actigraphy), and combine this information with a mathematical model that is designed to predict performance capability over a period of time and when future periods of increased fatigue/sleepiness will occur. However, like the other categories of technologies, precision and validation are critical criteria that must be met. A mathematical model that mis-estimates a cumulative performance decrement by only a few percentage points can lead to a gross miscalculation of alertness and performance capability over a work week. This is clearly a promising area, but much more work is needed.

3. Vehicle-based performance technologies

These technologies are directed at measuring the behavior of the transportation hardware systems under the control of the operator, such as truck lane deviation, or steering or speed variability, which are hypothesized to reflect identifiable alterations when a driver is fatigued (e.g., Wylie et al., 1996; Grace, 1996; King, 1996; Schwenk, 1996). The technologies are challenging to develop and implement owing to the complexity of driving behaviors under different conditions and the complexity of vehicle behavior relative to environmental conditions, but they offer the ultimate performance output--namely the behavior of the moving vehicle. They are less concerned with the condition of the operator than with the status of the vehicle. In addition to their face validity, they have many advantages (e.g., no wires, devices, or monitors on or aimed at an operator), but as with all technologies for preventing drowsy driving crashes, their scientific validity and cost-effectiveness remain to be demonstrated.

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

Table 1. Examples of biobehavioral measures used alone (or with other measures), as operator-centered technologies for on-line monitoring of alertness/vigilance.

Type of measurement	Examples of technologies
Video of the face (may include eyelid position, eye blinks, eye movements, pupillary activity, facial tone, direction of gaze, head movements)	• "PERCLOS" (Wierwille & Ellsworth, 1994; Wierwille, 1996) • Ford Motor Co. (UK) & HUSAT Res. Inst. (Richardson, 1995) • Nissan Research & Development, Inc. (Kaneda et al., 1994) • Toyota (Fukuda et al., 1995) • "Gaze Control System" (Stern, 1996)
Eye trackers	• "Eyegaze Systems" (LC Technologies; Lahoud, 1996) • "Eye Tracking System" (Applied Science Group, Inc.)
Wearable eyelid monitors	• "Alertness Monitor" (MTI Research, Inc.; MacLeod, 1996) • "Blinkometer" (IM Systems, Inc.) • "Nightcap" (Healthdyne Technologies; Stickgold et al., 1995) • "Eyelid activity measurement" (Leder et al., 1996) • "Stay-Awake Eye-Com Biosensor" (Torch, 1996)
Head movement detector	• "Proximity Array Sensing System" (Advanced Safety Concepts, Inc.; McIntosh, 1996)
EEG algorithms	• "Drowsiness detection" (Consolidated Research, Inc.) • "EEG algorithm adjusted by CTT" (Makeig & Jung, 1996) • "EEG spectral analysis" (Brookhuis, 1995) • "Quantitative EEG analysis" (Wylie et al., 1996)
ECG algorithms	• "MAP Process" (PALS Technology; Juszkiewicz, 1996)

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

STANDARDS FOR DROWSINESS DETECTION TECHNOLOGIES

If fatigue-monitoring technology development continues and is proposed as one piece of a programatic "fatigue management" alternative to proscriptive hours-of-service regulations, then technologies that allege to be effective must be shown to meet or exceed a range of criteria involving scientific, practical, and legal/ethical standards (Dinges, 1995b, 1996, 1997). A great deal of harm can be done if invalid and/or unreliable devices are quickly and uncritically implemented. In addition to the potential for increased crash risk, deployment of invalid and/or unreliable fatigue-detection technologies will result in wasted resources and provide only a false sense of security and fatigue management.

Table 2 summarizes the primary scientific and engineering criteria that technologies should meet to be maximally effective for monitoring operator vigilance (Dinges, 1997). The lack of answers to many of the scientific and engineering questions in Table 2 for most current technologies stems from the fact that many of them are being developed in a proprietary context, and their validation either is not attempted, or is not complete, or if complete is not available. There are many anecdotal claims and very few published validation studies. In addition to the lack of validation evidence, there is a dearth of information for many of the practical implementation questions in Table 2. Most technologies have not yet been systematically transitioned to operational test beds for evaluation. This is a critical second-stage hurdle to be overcome after validation is established. Practical questions of size, intrusiveness, cost, etc. can be addressed as the validity of a prototypical technology is tested.

The critical first step of establishing validity and reliability

The criteria outlined in Table 2 are achievable, but the progression should be from scientific / engineering validation, through the practical implementation phase, to the public-policy phase. Unfortunately, this has often not been the case. Excessive concern for selling technology to the user without first establishing the scientific validity and reliability of approaches is misguided and risky. For example, time, energy and resources spent on implementing a technology that is easily integrated into the work environment may be wasted if the device is later found not to detect vigilance errors, inattention, or drowsiness.

Once a technology has been developed to record a putative fatigue marker it must first be demonstrated to work. That is, it must actually record what it purports to record and do so consistently. Thus, an eyelid closure detection device must record eye lid closure and only

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized? In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

eye lid closure and must not miss eyelid closures. After these engineering criteria are met, then scientific validity relative to fatigue/drowsiness/hypovigilance detection must be established. For example, in terms of devices that purport to detect a fatigued operator, basic data must be provided on whether the device detects what it alleges to detect (i.e., fatigue/drowsiness/hypovigilance--this is the validity standard), and whether it can detect it repeatedly (this is the reliability standard). Even if a device is valid and reliable, to be practically useful, it must meet additional standards of high sensitivity and high specificity. Thus, the device must detect all (or nearly all) fatigue events and fatigued operators (i.e., high sensitivity standard), without too many false alarms (i.e., high specificity standard). A device that has high sensitivity but low specificity will detect fatigue, but may give too many false alarms to be useful. In contrast, a device with low sensitivity but high specificity will give few false alarms, but it may miss too many fatigue events to be useful.

Given the growing federal support for technology development in fatigue management (the facilitators), the entrepreneurial zeal currently overtaking technology companies in this area (the vendors), and the escalating attractiveness of fatigue management technologies to transportation industries (the buyers), there is a risk of a rush toward widespread use of technologies that do not reliably detect fatigue. At this time, more proactive and coordinated efforts are urgently needed among relevant governmental agencies, transportation industries, and the scientific and engineering communities to ensure that promising technologies for fatigue management meet minimum standards for the criteria outlined in Table 2. There is, also a need to scientifically validate claims of the efficacy of on-board fatigue alarms and countermeasures delivered through contingency with fatigue-detection technologies (e.g., claims that certain odors will reverse fatigue for an unspecified period of time).

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

Table 2. Scientific, practical, and legal criteria and questions regarding the development and use of technologies for monitoring operator vigilance or impairment.

Scientific / Engineering criteria	
Validity	Does it measure what it purports to measure, both, operationally (e.g., eye blinks) and conceptually (e.g., hypovigilance)?
Reliability	Does it measure the same thing consistently?
Generalizability	Does it measure the same event (operationally and conceptually) in everyone?
Sensitivity	What proportion of the persons (or times within a given person) does it detect when reduced vigilance is actually present? (Does it miss some hypovigilance or some hypovigilant persons?)
Specificity	What proportion of the persons (or times within a given person) does it correctly identify safe vigilance when it is actually present? (How often does it false alarm?)
Practical / Implementation criteria	
Ease of use	Can nearly everyone use it correctly?
Acceptance	Will the target population use the technology?
Unobtrusiveness	Is the technology "transparent" or convenient for the user?
Robustness	Can the technology withstand heavy use and/or abuse?
Economical	Is the technology cost-effective?
Implementation	Operationally how is the technology to be used? (For example, does it only detect reduced vigilance conditions? Does it also alert the operator? If it alerts the operator, what is the nature of the alert? Does it trigger a broader countermeasure response?)
Legal / Policy criteria	
Purpose	What is the goal of implementing the technology? Is the use of the technology mandatory? If so, who mandates it and for what purpose?
Privacy	Who has access to any data acquired by the technology?
Enforcement	Is the technology to be used for enforcement, compliance, or advancement/demotion? If so, how is this to be accomplished?
Misuse potential	Can use of the technology lead to misuse (1) by the person being

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

	monitored (e.g., continuing to operate while impaired); (2) by the mandating entity (e.g., requiring an operation to continue when impairment is present).
Liability	Who is liable if the technology fails to detect impairment or if it is misused in association with an adverse event?

Experimental design for establishing scientific validity

It is not possible to establish scientific validity and reliability for a device that was developed to detect fatigue/drowsiness/hypovigilance without testing the technology in a controlled context in which the antecedents of fatigue (sleep loss and/or circadian time) are explicitly manipulated in an experiment, and the effects of this manipulation are explicitly measured. This requires an experiment with precise control of the independent variable (i.e., sleep loss and/or circadian time to induce a range of alertness/drowsiness) and precise measurement of dependent variables. The latter should include an outcome variable that is well established to co-vary with alertness level to serve as a validation criterion, as well as the variable measured by the device. A validation variable (preferably a vigilance performance variable) is critical in the experiment to confirm that what the device recorded was a meaningful covariate of the adverse effects of the induced fatigue. The implications of these experimental design criteria for validating fatigue-detection technologies are clear. Studies in which there is no controlled or reliable manipulation of the independent variable (i.e., fatigue) cannot provide evidence of fatigue-detection validity. Similarly, if the validation criterion variable is either inherently unreliable itself (e.g., self report of alertness) or a physiological variable of uncertain relationship to actual performance (e.g., Richardson, 1995), then it cannot be determined with certainty whether the device's output variable was changing in relationship to a reliable criertion of impairment from fatigue. To the best of our knowledge, only a very few studies of fatigue-monitoring technologies have actually used a performance criterion variable in conjunction with controlled sleep deprivation to validate their drowsiness detection algorithm (e.g., Wierwille & Ellsworth, 1994).

The arguments that laboratory validation studies of fatigue-detection technology should be avoided due to their necessarily artificial activities (relative to actual driving) or because their results have been ignored in the past (cf., Brown, 1995) misses the point of scientific validation and the circumstances under which it must be demonstrated. Scientific validity is the most basic standard upon which fatigue-detection is predicated. The controlled laboratory context is precisely where validity should first be established to ensure that the device measures a meaningful fatigue-induced change, especially since fatigue associated with sleep

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

loss, circadian phase, and time-on-task is a brain-based phenomenon (Monk, 1991; Broughton & Ogilvie, 1992). Validation may be possible in some field-based controlled driving conditions, but only if error variance ("noise") from extraneous sources is controlled, and objective validation outcomes are recorded. In uncontrolled field driving trials, error variance can mask the validity of a technology, or create an apparent validity that is artificial and therefore not generalizable to other contexts. In contrast, if a device cannot be shown to be valid in well-designed laboratory experiments, where there is an optimal context for demonstrating that the independent variable (fatigue) will affect the device's dependent variable (alertness detection), there is little justification for transitioning the technology to field studies. If a device can be shown to be valid in well-designed laboratory experiments, then there is justification for transitioning it to field studies, where it will need to be evaluated to establish whether the added uncontrolled factors of a driving environment obscure its sensitivity or specificity. There are no shortcuts to this logical progression.

Experimental Validation Of Six Technologies To Detect Hypovigilance

In an effort to obtain measures of the scientific validity of a number of fatigue-detection technologies, we have recently completed a double-blind, controlled laboratory validation experiment on six of the technologies in Table 1. Psychomotor vigilance performance lapses were selected as the validation criterion variable for three reasons: (1) driving is fundamentally first and foremost a vigilance task requiring psychomotor reactions; (2) psychomotor vigilance has been validated to be very sensitive to fatigue from night work and sleep loss (Dinges et al., 1997); (3) hypovigilance while driving is the outcome most fatigue-detection technologies seek to identify. The technologies tested included a video-based scoring of eye closure by trained observers (PERCLOS; Wierwille & Ellsworth, 1994); two wearable eye-blink monitors (MTI Research, Inc.; IM Systems, Inc.); a head tracker device (Advanced Safety Concepts, Inc.); and two EEG algorithms (Consolidated Research, Inc.; Makeig & Jung, 1996). Fourteen healthy adult males remained awake in the lab for 42 hr, while working on a computerized test battery every 2 hr. Vigilance lapses each minute and every 20 minutes were used as the validation criteria. Each technology was time-locked to vigilance performance to test coherence between vigilance lapses and each technology's specific hypovigilance metric. Algorithms were applied to technology results by the respective vendors, who remained blind both to the lapse data and to time (i.e., the specific hour of continuous wakefulness at which each data file was acquired). These procedures were used to eliminate possible bias in drowsiness scores derived from the vendors.

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

The results of this trial are being published elsewhere, and therefore only a few key observations will be described here. While nearly all of the technologies were found to accurately predict lapses in at least one subject or a subset of subjects, only one technology consistently correlated at a high level with lapses within and between different subjects. Meeting the validation criterion both through high intra-subject and high inter-subject coherence is an important and highly promising outcome, since one of the more serious problems plaguing fatigue-detection and prevention is the large inter-subject differences in vulnerability to fatigue. For example, in the recent USA-Canada driver fatigue and alertness study, 14% of drivers accounted for 54% of all observed video-drowsiness episodes (Wylie et al., 1996). Some of the technologies proved to have inadequate predictive power or to be technically unreliable. While the technologies that performed poorly may not be able to predict performance outcomes in general, and fatigue-induced lapses in particular, there is also the possibility that their algorithms can be improved to enhance delectability. If such an retrospective enhancement were possible, a prospective re-validation trial would be necessary. More validation studies of this type are needed to sort out from the wide variety of biobehavioral fatigue monitors those that have the highest validity and reliability relative to actual performance capability.

The Need For Legal, Ethical, And Public Policy Standards

Finally, the legal and policy questions in Table 2 are largely not considered by the technology developers, beyond what is necessary to meet a given standard of safety in product development. However, these issues must be confronted if fatigue-detection technologies are to be located in the workplace. One of these issues concerns who has control over the detection technology. Another concerns the privacy rights of the individual being monitored, and the confidentiality issues of the information acquired. Related to the confidentiality question, are issues of enforcement and use of punitive contingencies when fatigue is detected. For example, from an enforcement perspective, should fatigue be viewed in the same way impairment from alcohol and drugs are viewed (Dawson & Reid, 1997)? What role should fatigue monitoring/detection have in assessing regulatory compliance versus providing feedback and/or education to an operator? What are the practices and policies for repeated detection of fatigue in an operator? If fatigue detection is involuntary and subject to enforcement contingencies, will operators accept it or seek to disable it? The answers to these and related questions are not obvious. Public policy discussion of these issues should begin now, while different technologies for fatigue detection in all modes of transportation are being developed, to allow legal policy to guide integration of the use of these technologies.

THE ULTIMATE CHALLENGE IS HUMAN

It appears that within the not-too-distant future some technologies will be deployed to prevent or limit certain catastrophic outcomes due to fatigued performance while driving. However, as reviewed above, there is much validation work yet to do to achieve this goal, and to determine how to use the most promising technologies in the face of large inter-individual differences in response to sleep loss and night work. While there is clearly potential for improving the safety margin with technologies, such a development should not be a substitute for setting standards for the functional capability of a transportation operator. Technologies may eventually prevent or limit certain catastrophic outcomes due to fatigued performance, but technologies are not substitutes for setting societal standards for the functional capability of an operator. On the other hand, technologies can help establish and maintain adherence to that standard if they are developed and used in a valid and responsible manner.

ACKNOWLEDGMENT

The substantive evaluation upon which this article was based was supported by contract DTNH22-93-D-07007 from the U.S. Department of Transportation, and in part by grant F49620-95-1-0388 from the U.S. Air Force Office of Scientific Research, by cooperative agreement NCC-2-599 from U.S. National Aeronautics and Space Administration, by grant NR04281 from the National Institutes of Health, U.S. Public Health Service, and by the Institute for Experimental Psychiatry.

REFERENCES

Akerstedt, T. and Folkard, S. (1994) Prediction of intentional and unintentional sleep onset in Ogilvie, R.D. and Harsh, J.R. (eds) Sleep Onset: Normal and Abnormal, American Psychological Association; Washington, D.C.

Belenky, G. (1997) Sustaining performance during continuous operation: The U.S. Army's sleep management system. International Conference Proceedings on Managing Fatigue in Transporation, Tampa, FL

Brookhuis, K.A. et al. (1986) EEG energy-density spectra and driving performance under the influence of some antidepressant drugs in O'Hanlon, J.F. and de Gier, J.J. (eds) Drugs and Driving, Taylor & Francis; London

Brookhuis, K. (1995) Driver impairment monitoring by physiological measures in Hartley, L. (ed) Fatigue & Driving: Driver Impairment, Driver Fatigue and Driving Simulation. Taylor & Francis; London

Broughton, R.J. and Ogilvie, R. (eds) (1992) Sleep, Arousal and Performance. Birkhauser-Boston, Inc.: Cambridge

Brown, I.D. (1995) Methodological issues in driver fatigue research in Hartley, L. (ed) Fatigue & Driving: Driver Impairment, Driver Fatigue and Driving Simulation. Taylor & Francis; London

Brown, I.D. (1997) Prospects for technological countermeasures against driver fatigue, Accident Analysis and Prevention, 29, 525-531

Daecher, C. (1996) Fitness for duty testing in the trucking environment in Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 31-34

Dawson D. and Reid K. (1997) Fatigue, alcohol and performance impairment. Nature, 388, 235

Dinges, D.F. and Graeber, R.C. (1989) Crew fatigue monitoring. Flight Safety Digest, 8, 65-75

Dinges, D.F. (1989) The nature of sleepiness: Causes, contexts and consequences in Stunkard, A., Baum, A. (Eds) Perspectives in Behavioral Medicine: Eating, Sleeping and Sex. Lawrence Erlbaum; Hillsdale

Dinges, D.F. (1995a) Technology / Scheduling Approaches in Managing Fatigue in Transportation: Promoting Safety and Productivity. Proceedings from the multi-modal symposium co-sponsored by the National Transportation Safety Board and NASA Ames Research Center, 53-58

Dinges, D.F. (1995b) An overview of sleepiness and accidents. Journal of Sleep Research, 4, Suppl. 2, 4-14

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized? In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

Dinges, D.F. (1996) Validation of psychophysiological monitors. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 35-41

Dinges, D.F. (1997) The promise and challenges of technologies for monitoring operator viiglance. Proceedings of the International Conference on Managing Fatigue in Transportation, American Trucking Associations Foundation, Tampa, FL, 77-86

Dinges, D.F. et al. (1997) Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep, 20, 267-277

Evans, L. (1991) Traffic Safety and the Driver. Van Nostrand Reinhold: New York

Fukuda, J. et al. (1995) Development of driver's drowsiness detection technology. Toyota Technical Review, 45, 34-40

Gilliland K. and Schlegel, R.E. (1993) Readiness to perform: A critical analysis of the concept and current practices. Office of Aviation Medicine, Federal Aviation Administration, NTIS No. AD-A269 379

Grace, R. (1996) Field testing a prototype drowsy driver detection and warning system in Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 27-30

Hamelin, P. (1987) Lorry drivers' time habits in work and their involvement in traffic accidents. Ergonomics, 30, 1323-33

Harris, W. (1977) Fatigue, circadian rhythm and truck accidents in Mackie, R.R. (ed) Vigilance: Theory, Operational Performance and Physiological Correlates. Plenum Press; New York

Horne, J.A. and Reyner, L.A. (1995) Sleep related vehicle accidents. British Medical Journal, 310, 565-567

Jewett, M.E. (1997) Models of circadian and homeostatic regulation of human performance and alertness. Doctoral Thesis, Harvard University, Cambridge, MA.

Juszkiewicz, K. (1996) Correlation of ultra-slow electrophysiological signals with performance. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 79

Kaneda, M. et al. (1994) Development of a Drowsiness Warning System. Proceedings of the 14th International Technical Conference on Enhanced Safety of Vehicles, Munich, Germany

King, D. (1996) Correlation of steering behavior with heavy truck driver fatigue. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 69

Knipling, R.R and Wang, J.S. (1995) Revised estimates of the US drowsy driver crash problem size based on general estimates system case reviews. 39th Annual Proceedings, Association for the Advancement of Automotive Medicine, Chicago

Knipling, R. (1996) The promise of technology for fatigue management: The Federal Highway Administration perspective. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 8-13

Kribbs, N.B. and Dinges, D.F. (1994) Vigilance decrement and sleepiness in Ogilvie, R. and Harsh, J. (eds) Sleep Onset: Normal and Abnormal Processes. American Psychological Association; Washington, D.C.

Lahoud, J.A. (1996) Technologies that have potential for driver vigilance studies. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 67

Leder, R.S. et al. (1996) Eyelid activity measurements: A new retroreflective sensor. Sleep Research, 25, 509

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

MacLeod, E. (1996) Personal Alertness Monitor. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 47-52

Makeig, S. and Jung, T-P. (1995) Changes in alertness are a principal component of variance in the EEG spectrum. NeuroReport, 7, 213-216

Maycock, G. (1997) Sleepiness and driving: The experience of U.K. drivers, Accident Analysis and Prevention, 29, 453-462

McDonald, N. (1984) Fatigue, Safety and the Truck Driver. Taylor & Francis: London

McIntosh, B. (1996) Capacitive Proximity Sensor. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 75

Mitler, M.A. et al. (1988) Catastrophes, sleep and public policy: Consensus report of a committee for the Association of Professional Sleep Societies. Sleep, 11, 100-109

Monk, T.H. (ed) (1991) Sleep, Sleepiness and Performance. John Wiley and Sons, Ltd.: Chichester

O'Hanlon, J.F. and Kelley, G.R. (1974) A psychophysiological evaluation of devices for preventing lane drift and run-off-road accidents, Federal Highway Administration, Report No. 1736-F

O'Hanlon, J.F. (1978) What is the extent of the driving fatigue problem? in Driving Fatigue in Road Traffic Accidents, Brussels: Commission of the European Communities Report No. EUR6065EN, 19-25

Pack, A.I. et al. (1995) Characteristics of crashes attributed to the driver having fallen asleep. Accident Analysis and Prevention, 27, 769-775

Rau, P. (1996) The National Highway Traffic Safety Administration's Drowsy Driver Technology Program. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 14-16

Rechtschaffen, A. and Kales, A. (eds) (1968) A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects. U.S. Department of Health, Education and Welfare, Public Health Service, Bethesda, MD

Richardson, J.H. (1995) The development of a driver alertness monitoring system in Hartley, L. (ed) Fatigue & Driving: Driver Impairment, Driver Fatigue and Driving Simulation. Taylor & Francis; London

Rosekind, M.R. et al. (1994) Crew factors in flight operations IX: Effects of planned cockpit rest on crew performance and alertness in long-haul operations, NASA Technical Memorandum 108839

Schwenk, S. (1996) TRW Driver Alertness Warning System (DAWS). Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 76-77

Shafer, J.H. (1993) The decline of fatigue related accidents on the NYS thruway. Proceedings of the Highway Safety Forum on Fatigue, Sleep Disorders and Traffic Safety, Albany, NY

Stickgold, R. et al. (1995) Nightcap detection of decreased vigilance. Sleep Research, 24, 500

Stern, J.A. (1996) Alertness and states of attention. Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 72-74

Torsvall, L. and Akerstedt, T. (1987) Sleepiness on the job: Continuously measured EEG in train drivers. Electroencephalography and Clinical Neurophysiology, 66, 502-511

Torch, W.C. (1996) The Stay-Awake Eye-Com Biosensor & Communicator in Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 68

U.S. National Transportation Safety Board (1990) Safety Study: Fatigue, Alcohol, Other Drugs, and Medical Factors in Fatal-to-the-Driver Heavy Truck Crashes, Vol. 1, Washington, D.C., NTSB/SS-90-01

Dinges, D.F. and Mallis, M.M: Managing fatigue by drowsiness detection: Can technological promises be realized?
In: Managing Fatigue in Transportation. Ed. Hartley, L., Pergamon, Kidlington: 209-229, (1998).

Wang, J.S. and Knipling, R.R. (1994) Single-Vehicle roadway departure crashes: Problem size assessment and statistical description. US Department of Transportation, National Highway Traffic Safety Administration

Wierwille, W.W. and Ellsworth, L.A. (1994) Evaluation of driver drowsiness by trained raters. Accident Analysis and Prevention, 26, 571-581

Wierwille, W.W. (1996) Driver status in Proceedings of the Technological Conference on Enhancing Commercial Motor Vehicle Driver Vigilance, American Trucking Associations Foundation, FHWA, NHTSA, McLean, VA, 17-26

Wylie, C.D. et al. (1996) Commercial Motor Vehicle Driver Fatigue and Alertness Study: Project Report, U.S. Department of Transportation Report No. FHWA-MC-97-002