Ocular parameters as an objective tool for the assessment of truck drivers fatigue

Yair Morad a,∗, Yaniv Barkana a, David Zadok a, Morris Hartstein a, Eran Pras a, Yosefa Bar-Dayan b

a Department of Ophthalmology, Assaf Harofeh Medical Center, Tel Aviv University, Zrifin 73000, Israel
b Israel Defense Forces Medical Corps, Israel

1. Introduction

In today’s technologically driven society, fatigue has become a major problem in occupational activities involving late-evening and night shifts, which require vigilance and attention over an extended period of time. Truck drivers are involved in many traffic accidents that cost high to the health and insurance systems (McCall and Horwitz, 2005). Fatigue and sleep deprivation have been identified as one of the major causes for traffic accidents. We assessed the use of these parameters as an objective screening tool for a driver’s fitness for duty. Pupillary diameter, pupil reaction to light and saccadic velocity were measured in 29 army truck drivers every morning for two months and compared to baseline measurements taken while the subjects were alert. An index which expressed the difference between study and baseline measurements was calculated, and drivers with significant deviation from baseline were disqualified and interviewed. Non-disqualified drivers served as controls. Twenty-nine percent of disqualified drivers reported sleeping less than the minimum of 7 h required by army regulations compared with 8% of control drivers (p = 0.01). Disqualified drivers had worse sleep quality the night before the test (Groningen Sleep Quality Scale, p = 0.03) and incurred more accidents per driving day during their service (0.023 vs. 0.015 accidents/day, p = 0.03). Two disqualified drivers admitted to using alcohol or sleeping pills. Thus, these ocular parameters may serve as a screening tool for drivers that are at high risk for driving. Drivers who were disqualified even once, tend to be involved in more motor vehicle accidents than their peers.

© 2009 Elsevier Ltd. All rights reserved.
to sleep deprivation by many others (Morad et al., 2000; Russo et al., 2005; Wilhelm et al., 2001; Wilhelm et al., 1996a; Wilhelm et al., 1996b). It was shown that several measurable parameters of the pupillary activity such as pupil diameter, the latency of pupillary reaction to light and the amplitude of constriction in reaction to light were all influenced by sleep deprivation. Velocity of saccades was also shown to be influenced by sleep deprivation (Morris and Miller, 1996; Porcu et al., 1998; Rowland et al., 2005; Russo et al., 2003; Zils et al., 2005).

In this study we examined the efficiency of measuring these two ocular parameters, pupillary activity and saccade velocity, as a screening tool for truck driver’s fatigue.

Several issues were addressed in the study:

1. The impact of performing the test on the unit’s daily routine—did it cause any delays or inconvenience?
2. The efficacy of the screening method in detecting drivers unfit for duty.
3. The effect of disqualifying several drivers each morning on the entire unit’s performance—did the system disqualify too many drivers?
4. The characteristics of the disqualified drivers.

2. Methods

This study was performed in the Israel Defense Forces Transportation Center during March and April 2006 and approved by the Israel Medical Corps Research Ethics Board. All truck drivers of a single transportation unit were included in the study. In this army unit every accident, with or without casualties, is reported to the driver’s safety officer. The exact details of the accidents are recorded and placed in the driver’s personal file. Since the vehicles are regularly inspected for damage, it is unlikely that a significant accident would go unnoticed and not be reported.

2.1. Apparatus

The FIT 2500 Fatigue Analyzer (FIT, Pulse Medical Instruments, Inc., Rockville, MD) was used for objective assessment of fatigue. This is a self-contained, fully automated, computer-controlled, commercially available optical tracking and recording system, which contains an infra-red pupillometry device together with an eye movements tracking system.

Examination was carried out as follows: subjects focused with their dominant eye on a low brightness light visible through a round opening in the device. When ready, the subject pressed a button to start the sequence. After 1 s the center light extinguished while left and right side lights were alternately illuminated, and the subject was instructed to shift the gaze between them (total visual angle 26.80). Saccadic velocity (SV) was measured at a sampling rate of 750 Hz with a calculated resolution of 0.6°. Subjects then focused their gaze on a center fixation light and pupillary diameter (PD) was measured at a rate of 60 samples per second. A flash of high-intensity bright light (0.2 s duration) then stimulated the pupillary light response, and pupillary constriction latency (CL) was measured (time from flash to onset of pupil constriction). Pupillary constriction amplitude (CA) was derived from the difference between the pre-flash pupillary diameter and the smallest after-flash diameter. If eye responses were not captured, as would occur with excessive blinking, head movement, or failure to direct gaze appropriately, the FIT issued an error message. The time for each test session was approximately 30 s.

Eighteen separate baseline measurement sets of the above parameters were recorded on at least five separate days for each subject during morning hours following a full night’s sleep. The four data variables (SV, PD, CA, CL) measured for each subject were stored in the system’s data bank. Any measurement taken during the study was compared to these measurements using a chi-square distributed index:

\[
\text{FIT index} = \frac{(PD – \mu_{PD}}{\sigma_{PD}}^2 + \frac{(CA – \mu_{CA}}{\sigma_{CA}}^2 + \frac{(CL – \mu_{CL}}{\sigma_{CL}}^2 + \frac{(SV – \mu_{SV}}{\sigma_{SV}}^2)
\]

\(\mu = \text{baseline average; } \sigma = \text{baseline standard deviation}\). Since all four parameters have normal Gaussian distribution, index value above 18.467 is indicative of a 99.9% chance that all the parameters are outside the normal distribution of the baseline data (Dudewicz, 1976; Perry, 1988). When this occurred, the system beeped and a “high risk” warning was given. If the FIT index value was equal or below 18.467, the current data set was added to the driver’s existing data set and used for further calculations. Thus, more and more normal data samples of each driver were added to the baseline data bank, allowing more accurate detection of deviated measurements.

2.2. Data acquisition

2.2.1. Baseline measurements

Baseline measurements of ocular parameters were obtained from all drivers on at least five non-consecutive mornings. Only drivers that slept in the army barracks were examined as these drivers night’s sleep was monitored. Drivers who stated that they slept less than 7 h were not examined. In addition, before taking the ocular test, each subject was asked to complete two questionnaires:

1. The Stanford Sleepiness Scale (Appendix A) (Hoddes et al., 1973) This is a self-rating questionnaire, which contains seven statements describing a gradually increasing feeling of sleepiness ranging from “Feeling active and vital; alert; wide-awake” (score 1) to “Almost in reverie; sleep onset soon; lost struggle to remain awake” (score 7). Drivers who graded their feeling as equivalent to score 3 or higher did not take the test during that day.

2. The “The Groningen Sleep Quality Scale” (Mulder-Hajonides van der Meulen et al., 1980). This scale ranges from 0 to 14, with a higher score indicating a lower subjective quality of sleep (see Appendix A). The scale was originally constructed to study sleep problems in patients with depression. In a validation study of 80 inpatients with depression, the mean (S.D.) score on the scale was 6.0 (4.2) and Cronbach’s alpha for internal consistency 0.88 (Leppamaki et al., 2003). The scale has previously been used in studies of quality of sleep (de Weerd et al., 2004; Hoekema et al., 2004; Jafari et al., 2007; Leppamaki et al., 2003; Vermeulen et al., 2004). In general, if sleep is unrestricted and undisturbed, subjects score between 0 and 2 points. A higher score indicates disturbed sleep. Subjects that scored their sleep quality as 3 or higher did not take the test during that day.

2.2.2. Study data acquisition

Every morning, before commencing their regular tasks each driver was asked to grade his feeling of sleepiness using the Stanford Sleepiness Scale and grade their sleep quality the night before with “The Groningen Sleep Quality Scale”. Afterwards, ocular parameters were measured by the Fatigue Analyzer. If the FIT index value was ≤ 18.467, the driver carried on with his regular duties, and the data set was added to the stored baseline measurements. If the index was above 18.467 in two consecutive measurements the driver was defined as a “high risk suspect” and was referred to an interview with one of the researchers. During the interview each subject was asked how many hours he slept on the previous night, and if he
had used any medications or chemical substances during the 24 h period before the test. The collected data was then presented to the commanding officer, who decided whether or not the driver would proceed with his duties or not. In addition, two drivers who were not disqualified by the system (FIT index ≤ 18.467) were paired with each disqualified driver and were also interviewed similarly. Their data was analyzed as a control group. The results of the FIT index of each subject were masked to the interviewing researcher.

At the end of the working day, the drivers were asked to fill the questionnaires and perform the ocular test once more.

2.3. Statistical analysis: Student’s t-test, Fisher’s exact test and Mann–Whitney test were used for statistical analysis

In order to assess the effect of doing the test on the unit’s daily routine, we asked the commander of the unit to answer two questions:

1. Did the unit perform all its tasks during the study on time and without delays relevant to the study?
2. Was the number of disqualified drivers within the units capacity to find replacement or did the unit had to take excessive measures such as calling in drivers who were off duty or cancel tasks?

3. Results

3.1. Ocular tests

All 29 drivers of one transportation unit participated in the study. The average FIT index values that were measured during morning hours, before the drivers started their driving tasks was 5.1 ± 0.2, which was significantly lower than the average value measured in the evening hours at the end of the working day (11.1 ± 0.2, student’s t-test p = 0.02). This difference was also seen in the Stanford Sleepiness Score: average values changed from 2.1 ± 0.2 during morning hours to 4.0 ± 0.4 in the evening (p = 0.04), however, these parameters were not correlated (r = 0.2, p = 0.3).

The study was carried out over a period of two months. During the first month, 127 examinations were performed. On 12 (9.4%) examinations, FIT index values were >18.467, resulting in driver’s disqualification. During the second month, when the existing normal database on each driver was considerably larger, 180 examinations were performed but only five examinations were labeled “high risk” leading to driver’s disqualification (2.7%). The difference in disqualification rate between the first and second month was statistically significant (Fisher’s exact test, p = 0.02).

The results are presented in the Table 1. Of the 17 examination sets that were labeled “high risk”, on five occasions the drivers had used any medications or chemical substances. The use of pupillary activity to detect driver fatigue was first suggested by Yoss in the late sixties (Yoss, 1969), and later by others (Russo et al., 2003; Russo et al., 2005; Wilhelm, 2008). Other ocular parameters such as blink rate and duration and saccade velocity were also suggested (Papadelis et al., 2007; Volkov and Mashkova, 1984).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>“High Risk” test score (disqualified)</th>
<th>“Low Risk” test score (control)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sleep less than 7 h</td>
<td>5/17 (29%)</td>
<td>1/32 (3%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Average sleep time (h)</td>
<td>6.6 ± 1.4</td>
<td>7.4 ± 1.7</td>
<td>0.08</td>
</tr>
<tr>
<td>Sleepiness score</td>
<td>3.2 ± 1.715</td>
<td>2.344 ± 1.310</td>
<td>0.12</td>
</tr>
<tr>
<td>Sleep Quality Scale</td>
<td>2.29 ± 0.35</td>
<td>1.37 ± 0.2</td>
<td>0.03</td>
</tr>
<tr>
<td>Consumed alcohol or sedatives</td>
<td>2/17 (12%)</td>
<td>0/32 (0%)</td>
<td>0.11</td>
</tr>
</tbody>
</table>

3.2. Questionnaires

The results of the Stanford Sleepiness Scores and the Groningen Sleep Quality Scale of the disqualified and control drivers are also presented in the Table 1. Disqualified drivers had higher scores in both scales however the difference reached statistical significance only for the Groningen Scale.

3.3. Accident history

Motor vehicle accident history of each driver was collected from the unit’s safety officer files. Twelve drivers who were disqualified at least once were involved in 11 accidents, and 17 drivers who were not disqualified were involved in 11 accidents. We normalized the data by the duration of service in the unit. Disqualified drivers performed 0.023 accidents per driving day, compared with 0.015 accidents performed by non-disqualified drivers (student’s t-test, p = 0.03).

With regard to the effect of the study on the unit’s daily routine: the study did not cause any delays in the unit’s performance, nor resulted in the need to cancel a driver’s day off duty or assigned tasks.

4. Discussion

Driver fatigue has long been identified as a major cause for road accidents due to reduced driving performance (Connor et al., 2001; Horne and Reyner, 1999). This problem also affects professional truck drivers: in one study, over a quarter of long distance truck drivers reported falling asleep at the wheel at some time during the previous 12 months of driving (McCartt et al., 2000). Hence, finding an efficient screening tool that will determine those drivers who are too sleepy to drive or who are incapable of driving due to chemical substances use is important.

Such a screening tool should have several characteristics: the test should be rapid and easy to perform, it should not be influenced by the subject’s motivation, and the rate of false positive results should be low—as it is not practical to disqualify a large number of subjects every morning.

Several autonomic parameters have been suggested in the past for objective evaluation of fatigue and sleep deprivation such as electrocardiogram, electroencephalogram (Dussault et al., 2005; Zhong et al., 2005), rectal temperature (Monk et al., 1997; von Zerssen et al., 1987), and blood pressure (Zhong et al., 2005), etc. However, none of these assessments methods has become popular due to drawbacks such as inconvenience, difficulties in data interpretation, impractical use of field settings and other causes which eventually rendered these methods non-cost effective.

The use of pupillary activity to detect driver fatigue was first suggest by Yoss in the late sixties (Yoss, 1969), and later by others (Russo et al., 2003; Russo et al., 2005; Wilhelm, 2008). Other ocular parameters such as blink rate and duration and saccade velocity were also suggested (Papadelis et al., 2007; Volkov and Mashkova, 1984).

As seen in our study, the use of ocular parameters for such a purpose looks promising. The test was easy to perform, and took no more than 30 s per subject. Performing the test did not affect the daily routine of the transportation unit, and caused no delays. In this case, the FIT index was an efficient screening tool that successfully identified those drivers who were too sleepy to drive or who were incapable of driving due to chemical substances use.
study, we examined a population of drivers during their compulsory army service. These drivers had different levels of motivation to perform their work, and one of the concerns raised by the commanders before starting the study was whether they will use the test to “skip duties”. It was important, therefore, that the test examine parameters that could not be influenced subjectively—as was seen in our study.

The FIT index values were shown to be influenced by driver's fatigue: morning values were lower than evening values. A similar observation was reported by LeDuc et al. who measured ocular parameters in Apache Pilots before and after missions (LeDuc et al., 2005). Our results are also supported by another study using the same PMI apparatus. Russo et al. examined the relationship between the same ocular parameters examined in our study and truck driver's performance on a driving simulator. They found that drivers who slept only 3–5 h per night had significantly different ocular parameters than those who slept 8–9 h per night. These changes were significantly correlated with objective measurement of vigilance such as the sleep latency test, and objective parameters such as the Stanford Sleepiness Score. Moreover, several parameters such as saccade velocity and initial papillary diameter were found to be correlated with the number of accidents involving drivers who slept only 3–5 h per night (Russo et al., 2003).

Although the ocular parameters examined in this study were shown to be influenced by sleep deprivation in multiple studies (Morad et al., 2000; Rowland et al., 2005; Russo et al., 2003), they were also shown to have high intra- and inter-subject variability (Schmidt et al., 1979). It has been previously suggested that using the individual as his or her own control gives greater sensitivity and specificity than do comparisons to a population norm (Miller, 1996).

In our study, the importance of a large database of ocular parameters measurements for each individual was clearly seen. During the first month of the study, a relatively large proportion of the drivers were disqualified – 9.7%. This number decreased daily, as the system acquired more and more normal data of the ocular parameters of the subjects. This “self-learning” process led to a significant decrease in the number of disqualified subjects during the second month to 2.7%, and could theoretically lead to even lower number as the process continued. A large baseline data set reduces the rate of false positive results, however theoretically it may also increase the number of false negative ones. How this phenomenon should affect our cut-off criteria for disqualification should be further studied.

Pupillary activity is also known to be influenced by ingestion of chemical substances or alcohol, and ocular parameters were suggested as a screening tool for used by police officers and medical staff to detect such usage (Kosnoski et al., 1998; Tennant, 1988). For example, detecting HGN, or horizontal gaze nystagmus, has been part of the SFST (Standardized Field Sobriety Tests) that US police agencies use as a roadside test for alcohol use during driving. Ocular saccade velocity, which is one of the ocular parameters measured in this study, was shown to be influenced by alcohol intoxication (Crowdy and Marple-Horvat, 2004; Vorstius et al., 2008). Pupil size and activity was also shown to be influenced by alcohol abuse (Grunberger et al., 1998) and chemical abuse (Kosnoski et al., 1998; Richman et al., 2004). In this study, two subjects who used sleeping medications and alcohol were identified by the system, adding further merit to the screening process.

Subjects who were disqualified in the study had a significantly higher score in the Groningen Sleep Quality Scale compared to the weak correlation between the subjective assessment of fatigue and the objective measurements of ocular parameters graded by the Stanford Sleepiness Score. A possible explanation for this result is that the Groningen Scale has only yes/no questions which report events that actually happened the night before. On the other hand, providing a scale to the subjective feeling of fatigue may be influenced by many factors such as mood, motivation, habituation, etc. Similar results were found in an earlier study made by our group (Morad et al., 2000) and also with other objective assessments of fatigue, like the Multiple Sleep Latency Test (Schmidt and Schreie, 1989).

Our study had several limitations: the actual sleeping time before duty was not monitored and leaving us to rely on the subject's statement. Army regulations require 7 h of sleep on the night before a driving assignment which may mean that some of our subject slept less but were hesitant to admit it. However, the FIT index values of subjects who admitted to sleep less then 7 h were significantly higher than that of those who stated they slept more, indicating that our results are reliable. Other reports also support the assumption that self-reported data is usually accurate (Arthur et al., 2001).

Another inherent problem is the arbitrary use of the cut-off point of 18.467 as the FIT index value above which subjects were considered “high risk” for driving. A positive correlation between the FIT index value and the rate of accidents performed by the drivers was demonstrated in several previous studies (Rowland et al., 2005; Russo et al., 2003). Establishing an exact cut-off point for which all subjects may be impractical. We decided on this value since statistically there is only a 0.01% chance that these ocular parameters are not significantly different than the baseline databank obtained for each driver. The relatively good efficiency of revealing subjects that violated army regulations (29% of disqualified subjects—by voluntary statement alone) may mean that this cut-off point is indeed practical.

Another interesting observation of this study is that drivers who were disqualified even once were more likely to be involved in accidents than their peers. This may indicate that these drivers suffer from chronic fatigue. Moreover, failure to comply with regulations, such as sleeping time regulations, was previously shown to be correlated with a high accident record (Chipman and Morgan, 1975). Late-sleeping drivers, who were involved in more motor vehicle accidents, may be characterized by personality traits such as sensation-seeking (Schwebel et al., 2006).

5. Conclusions

This study examined the use of ocular parameters measured by a commercially available product for fatigue screening of truck drivers. We have found this tool to be practical and reliable with the limitations mentioned above. Ocular parameters can serve as an efficient screening tool for fitness for duty examinations in a variety of disciplines.

Appendix A.

A.1. The Groningen Sleep Quality Scale

1. I had a deep sleep last night
2. I feel that I slept poorly last night
3. It took me more than half an hour to fall asleep last night
4. I woke up several times last night
5. I felt tired after waking up this morning
6. I feel that I did not get enough sleep last night
7. I got up in the middle of the night
8. I felt rested after waking up this morning
9. I feel that I only had a couple of hours’ sleep last night
10. I feel that I slept well last night
11. I did not sleep a wink last night
12. I did not have trouble falling asleep last night
13. After I woke up last night, I had trouble falling asleep again.
14. I tossed and turned all night last night.
15. I did not get more than 5 h sleep last night.

All items are scored true/false.

The first question does not count for the total score.

One point if answer is ‘true’: questions 2, 3, 4, 5, 6, 7, 9, 11, 13, 14, 15.

One point if answer is ‘false’: questions 8, 10, 12.

Maximum score 14 points, indicating poor sleep the night before.

A.2. Stanford Sleepiness Scale

<table>
<thead>
<tr>
<th>Score</th>
<th>Subjective feeling of alertness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Feeling active and vital; alert; wide-awake.</td>
</tr>
<tr>
<td>2</td>
<td>Functioning at a high level, but not at peak; able to concentrate.</td>
</tr>
<tr>
<td>3</td>
<td>Relaxed; awake; not at full alertness; responsive.</td>
</tr>
<tr>
<td>4</td>
<td>A little foggy; not at peak; let down.</td>
</tr>
<tr>
<td>5</td>
<td>Fogginess; beginning to lose interest in remaining awake; slowed down.</td>
</tr>
<tr>
<td>6</td>
<td>Sleepiness; prefer to be lying down; fighting sleep; woozy.</td>
</tr>
<tr>
<td>7</td>
<td>Almost in reverse; sleep onset soon; lost struggle to remain awake.</td>
</tr>
</tbody>
</table>

References