The Laryngoscope
Lippincott Williams & Wilkins, Inc., Philadelphia

The Road to Danger: The Comparative Risks of Driving While Sleepy

Nelson B. Powell, MD; Kenneth B. Schechtman, PhD; Robert W. Riley, MD; Kasey Li, MD;
Robert Troell, MD; Christian Guilleminault, MD

Objectives/Hypothesis: **A large sector of the population of the United States has sleep deprivation directly leading to excessive daytime sleepiness. The prevalence of excessive daytime sleepiness in this population ranges from 0.3% to 13.3%. The consequences of even 1 to 2 hours of sleep loss nightly may result in decrements in daytime functions resulting in human error, accidents, and catastrophic events. The magnitude of risks in the workplace or on the highways resulting from sleepiness is not fully understood or appreciated by the general population. Hence, to more clearly emphasize the magnitude of these risks, we question whether mild sleep deprivation may have the same effect as alcohol on reaction times and driving performance.** ***Study Design:*** **Nonrandomized prospective cohort investigation.** ***Methods:*** **Sixteen healthy matched adult subjects (50% women) were stratified into two groups, sleep deprived and alcohol challenged. The sleep-deprived group was further subdivided into acute (one night without sleep) and chronic (2 h less sleep nightly for 7 d) sleep deprivation. Each group underwent baseline reaction time testing and then drove on a closed course set up to test performance. Seven days later, the group repeated this sequence after either sleep deprivation or alcohol intake.** ***Results:*** **There were no significant between-group differences (sleep deprivation or alcohol challenged) in the changes before and after intervention for all 11 reaction time test metrics. Moreover, with few exceptions, the magnitude of change was nearly identical in the two groups, despite a mean blood alcohol concentration of 0.089 g/dL in the alcohol-challenged group. On-track driving performances were similar (*P* = .724) when change scores (hits and errors) between groups were compared (baseline minus final driving trial).** ***Conclusion:*** **This comparative model suggests that the potential risks of driving while sleepy are at least as dangerous as the risks of driving illegally under the influence of alcohol.**
Key Words: **Sleep deprived, alcohol challenged, psychomotor vigilance test, driving performance.**

Laryngoscope, **111:887–893, 2001**

INTRODUCTION

A large sector of the American population has acute or chronic sleep deprivation, or both.[1] The prevalence of excessive daytime sleepiness (EDS) in this group varies from 0.3% to 13.3%.[2] It is well known that a cumulative reduction in the normal nightly sleep period results in decrements in daytime function and quality of life.[3–5] Sleep disruption or sleep loss causes EDS, whether secondary to a sleep disorder or simply to sleep deprivation. Regardless of how sleep is disrupted, disruption may have a profound effect on daytime performance resulting in human error, accidents, or even catastrophic events.[1,6–8] It is all too common, because of the pressures of our 24-hour society, to lose 1 to 2 hours of sleep nightly during the work week, to just keep up or to occasionally stay up all night to finish a project or engage in personal activities. Even these few hours of sleep loss can cause marked disruption in daytime performance. *The magnitude of the effects and risks in the workplace or on the highways as a result of this sleep loss is not fully understood or appreciated by the general population.* We questioned whether acute or chronic sleep deprivation, when compared with alcohol intake alone, had the same impact on reaction times and driving performance and whether subjects were aware of their decrements in alertness.

MATERIAL AND METHODS

Study Design

This nonrandomized prospective cohort investigation compares reaction times and actual driving performance in healthy unpaid volunteer subjects who underwent acute and chronic sleep deprivation versus a matched (age, gender, body mass index, neck circumference, and education) adult group challenged with alcohol (21 y of age or older) who underwent the same protocol requirements as the sleepy subjects.

Presented at the Annual Meeting of the American Academy of Otolaryngology—Head and Neck Surgery, Committee Session on Sleep Disorders, Washington, DC, September 27, 2000.

From the Stanford Sleep Disorders and Research Center, Department of Psychiatry and Behavioral Science (N.B.P., R.W.R., K.L., R.T., C.G.), Stanford University School of Medicine, Stanford, California, and the Department of Biostatistics (K.B.S.), Washington University School of Medicine, St. Louis, Missouri.

Supported in part by the Sleep Education and Research Foundation (a nonprofit public foundation).

Editor's Note: This Manuscript was accepted for publication January 31, 2001.

Send Correspondence to Nelson B. Powell, MD, 750 Welch Road, Suite 317, Palo Alto, CA 94304, U.S.A.

Setting

General Motors (GM) proving ground, Mesa, Arizona.

Test Subjects

Sixteen healthy adult volunteers (50% women) participated. Despite the small study size, an effort was made to represent an average cross-section of the population. All subjects were recruited by NBC Dateline News and gave written informed consent.

Inclusion Criteria

These investigations limited the study to an adult cohort. Daily alcohol intake could not exceed one to two drinks (moderate drinker).

Exclusion Criteria

Subjects who used alcohol on a more extensive basis (three drinks or more daily and drinking on the weekends) were excluded. Shift work excluded the subject. Driving records and licenses were checked, and documentation of a conviction for driving under the influence of alcohol or reckless driving was an exclusion.

Requirements and Evaluations

All medications, food, or drink that might interfere (increase or decrease) with alertness were to be avoided for a period commencing 48 hours before the start of the study and continuing until the study was completed. The following were conducted: a medical history, baseline vital signs, sleep habits questionnaire, sleep logs, visual analogue scale (VAS) testing, actigraphy, and psychomotor vigilance testing (PVT). Urine beta human chorionic gonadotropin studies were required for all fertile women.

Sleep-Deprived Group

Eight subjects (50% women) were directed to the sleep deprivation group, which was further subdivided into two groups. The acute sleep-deprived (ASD) subgroup ($n = 4$) was sleep deprived for a single night's sleep. The chronic sleep-deprived (CSD) subgroup ($n = 4$) was limited to 5 hours of sleep nightly for a 7-day period or a maximum of 2 hours less than the normal sleep time for each individual.

Alcohol-Challenged Group

The alcohol-challenged group consisted of eight matched subjects (50% women) who were not sleep deprived and who were challenged with alcohol.

Description of Study Methodology

Test subjects underwent a period of either sleep deprivation or alcohol challenge. Before sleep deprivation or alcohol challenge each subject underwent reaction time (RT) tests in a fully rested and nonsleepy state and then drove a test vehicle on a closed course set up to evaluate driving performance. This constituted the baseline performance on which comparisons of performance were made. After each group completed its respective sleep deprivation or alcohol challenge, the group members repeated the RT tests and driving challenge. This occurred exactly 1 week after the baseline testing. All testing, questionnaires, logs, and driving sessions (baseline and final) were performed at the same time of day to minimize circadian rhythm variations.

Driving Performance Course

The driving performance course was the GM proving ground at Mesa, Arizona. The main course was 1.4 miles of turns and coned-off pathways. The coned-off pathways were used to simulate a single-lane route incorporating occasional turns (some of decreasing radius), narrow segments, intersections, intermittent stops, and planned lane choices with the use of flashing yellow arrows. A non-coned loop of 0.60 miles connected to the main test pad was driven in both directions and was used only for the sleep-deprived group. A stoplight and yellow flashing arrows were each introduced once during both trials. *Accuracy, not speed, was emphasized as the primary goal of all the test drives*. To minimize any advantage or knowledge of the hazards, the driving subjects were not allowed to watch each other drive, and after driving they were separated from their other group members. The Dateline NBC News camera crew documented the driving performance of all subjects with 13 cameras, five of which were in or on the test vehicle. A face camera helped captured mood and behavior while driving. Any question of scoring error on the course could be reviewed and confirmed on these tapes.

Scoring and Hazards

All drivers were advised that the standard GM test course driving protocol counted any cone that was hit or moved as a driving infraction equal to one point. Off-track excursions outside the cones were also counted as one point. In addition, a stoplight and a flashing yellow arrow were used as challenges, and if there was a failure to respond appropriately, the drivers were charged with a point. Multiple (three) hazards as might be experienced in real life were also introduced. These included a rolling beach ball, a fake dog on wheels crossing in front of the test car, and a pop-up car (constructed of soft material). The subjects in all groups were *not* told that the final driving session would include hazards. Each hazard was introduced only once during the *final* test driving session. This was the protocol for both the sleepy and alcohol-challenged groups. All errors, hits, or improper evasive actions were scored on track by the primary investigator and the on-track designated observers, as well as from the chase car.

Driver Safety

To ensure safety, the traffic safety department of GM constructed the course to limit driving speed. All subjects drove a 1999 Pontiac Grand Prix GTP that was equipped with a special emergency braking mechanism. Each test subject was accompanied by a GM test-driving safety instructor who manned the safety-control braking button. The safety companion was instructed not to communicate with the subject for any reason *except if in his or her judgment it was necessary to terminate the test because of a clear and present danger.*

Sleep Logs

All subjects completed a sleep log for a 5-day period before the baseline driving test. This was repeated for the 7 days before the final driving test.

Actigraphy

Members of the CSD subgroup were also required to wear a wrist device that evaluates the sleep-wake cycle by motor activity (actigraphy).[9] This Actiwatch (model AW64, Mini Mitter Company, Inc., Sunriver, OR), was worn on the nondominant wrist for 7 days before testing. The data (activity) from the Actiwatch and concurrent nightly sleep logs were recorded for the 7 days and nights during sleep deprivation and before final testing to document the sleep limitation period before the final test date.

Visual Analogue Scales

Visual analogue scales[10–12] using the standard 10-cm scale were required for the CSD subgroup. The anchors "not sleepy to very sleepy" were used to assess the study variable of daytime sleepiness during the daily sleep-deprivation period. Anchors of "excellent sleep to poor sleep" were used to assess the quality of the previous night's sleep during the test period.

Epworth Sleepiness Scale

The Epworth Sleepiness Scale (ESS)[13,14] was completed by all subjects immediately before the first (baseline) and final driving tests (range, 0–24).

Breath Ethanol Testing

The eight alcohol-challenged, non–sleep-deprived subjects underwent alcohol dosing. Breath alcohol testing and data collection were supervised by two officers from the Phoenix, Arizona, police department. Planning and coordination of this portion of the study were conducted by the director of the Forensic Science Laboratory (Phoenix, AZ). An individualized dose of alcohol was given according to body weight to raise the blood alcohol levels to just below or near levels considered illegal when driving in the state of Arizona (blood alcohol content [BAC] equal to or greater than 0.10 g/dL). Oral rinses (H_20) were used after drinking to eliminate the effects of mouth alcohol. Before performance testing, each subject underwent multiple breath analyzer testing during the rise of the BAC. When their respective BAC values rose to just below legal levels, they immediately repeated the PVT and then drove the test course. An Intoxilyzer model 5000 (CMI Inc., Owensboro, KY) using infrared alcohol detection technology (U.S. Measurements, U.S. Department of Transportation–approved) was used for all measurements of breath alcohol concentration (BrAC). A standard was used to calibrate against a known BrAC control value (0.100 g/210 L of breath). The BrAc value is considered by the Department of Transportation as an equivalent to the BAC value. Results of BAC testing were reported as grams of alcohol per deciliter of blood.

Reaction Time Testing

All subjects performed simple visual RT testing using a PVT instrument (Ambulatory Monitoring, Inc., Ardsley, NY). This handheld, portable, microprocessor-controlled instrument has been previously validated to be highly sensitive to the sustained attentional process that is fundamental to normal, alert neurobehavioral functioning.[15–17] The PVT device allowed the collection, storage, and subsequent evaluation of all 11 RT metrics. A test trial consisted of responding to a bright red light (light-emitting diode—digital counter) by pressing a response button as soon as the light stimulus is seen. The response stops the light-emitting diode (1 s) and displays the reaction time in milliseconds. Data were maintained by solid-state storage and later downloaded to an IBM PC. A minimum of two PVT demonstration trials were given for practice, and to help minimize any learning effect. These outcome measures of analysis were all summary RT statistics that reflected the time it took for subjects to respond to 80 to 90 stimuli during each 10-minute testing period. Following the suggestion of Dinges,[18] reciprocal transforms were used because the reciprocals were more normally distributed than the variables themselves. Each subject completed baseline PVT twice before intervention: once before and once after driving. A mean of these PVT scores was calculated. Each subject also completed PVT twice after intervention (sleep deprived or alcohol challenged) in the same sequence, and the mean score was again calculated. Each PVT sequence (before and after driving) was separated by the time it took to complete the respective test drive. The mean scores before and after intervention were then compared.

Data Analysis and Statistical Methods

Between-group comparisons of continuous variables were performed using unpaired Student *t* tests; χ^2 tests were used to compare dichotomous variables. However, when the conditions for an unpaired Student *t* test were not satisfied, Wilcoxon's test was employed as a nonparametric alternative. For each subject with sequential data from 7 days of completing VAS testing, a regression slope was computed using simple linear regression. The null hypothesis that the resulting slopes were equal to zero was then tested using paired Student *t* tests. A *P* value of less than .05 was considered significant. An SAS statistical program (version 6.12, Cary, NC) was used to analyze these data.

RESULTS

Demographics

A total of 28 candidates were interviewed (Table I). The 12 individuals who were not selected were similar in all respects to the final study group. They were not selected because of an inability to avoid engaging in safety-sensitive activities during the sleep-deprivation phase or an inability to commit to both testing sessions. Sixteen subjects started the study, and all completed the study. The group consisted of three Hispanic, one black, and 12 white volunteers. The mean number of years of education beyond high school (mean ± SD) was 3.4 ± 1.7 years (range, 0–6 y).

Reaction Time Results

Table II shows the between-group characteristics before and after intervention when the treatment group is either sleep deprived or alcohol challenged. These two

TABLE I.
Demographic and Clinical Characteristics of Alcohol-Challenged Subjects Compared With Sleep-Deprived Subjects.

Variable	Entire Group (N = 16)	Alcohol-Challenged Group (N = 8)	Sleep-Deprived Group (N = 8)	*P* Value
Age	29.0 ± 6.1	31.1 ± 7.0	26.9 ± 4.5	.168
Female gender	50%	50%	50%	1.00
BMI (kg/m^2)	25.8 ± 4.5	24.2 ± 2.6	27.5 ± 5.5	.153
Systolic BP	126 ± 16	122 ± 9	130 ± 21	.306
Diastolic BP	76.9 ± 9.4	74.5 ± 8.6	79.7 ± 10.2	.302
Pulse	79.5 ± 9.8	76.2 ± 6.2	83.1 ± 12.2	.182
Neck circumference*	36.3 ± 4.7	35.2 ± 4.6	37.4 ± 5.0	.383

All results are based on *t* tests.
* Centimeters.
BMI = body mass index; BP = blood pressure.

TABLE II.
Reaction Time Results (Between-Groups Analysis).

Variable*	Alcohol-Challenged Group 2/19/00 Saturday Pre-intervention	Alcohol-Challenged Group 2/27/00 Sunday Post-intervention	Alcohol-Challenged Group Difference (Post-Pre)	Sleep-Deprived Group 2/19/00 Saturday Pre-intervention	Sleep-Deprived Group 2/26/00 Saturday Post-intervention	Sleep-Deprived Group Difference (Post-Pre)	*P* Value
Mean cones/hazards hit	1.37 ± 1.2	3.62 ± 3.2	2.25 ± 3	0.71 ± 1.0	3.00 ± 4.5	2.29 ± 5	.724
Mean RT in milliseconds	234 ± 24	294 ± 87	60.7 ± 77	239 ± 18	300 ± 62	60.4 ± 56	.891
SD of RT[1]	39.6 ± 12	119 ± 141	79.2 ± 143	58.0 ± 25	106 ± 44	48 ± 60	.948
Reciprocal of RT[2]	4.4 ± 0.4	3.9 ± 0.8	−0.56 ± 0.6	4.4 ± 0.3	3.7 ± 0.7	−0.64 ± 0.5	.837
Median of RT	228 ± 23	266 ± 54	38.6 ± 39	227 ± 18	272 ± 49	44.6 ± 38	.725
Maximum RT: fastest 10%[3]	187 ± 11	206 ± 32	19.9 ± 25	186 ± 11	206 ± 28	19.6 ± 29	.992
Minimum RT: slowest 10%[4]	317 ± 42	548 ± 389	231 ± 388	358 ± 53	534 ± 162	176 ± 183	.961
Reciprocal of RT: slowest 10%[5]	3.3 ± 0.5	2.5 ± 0.9	−0.78 ± 0.8	3.0 ± 0.3	2.3 ± 0.7	−0.74 ± 0.8	.941
SQRT (lapses) + SQRT (lapses + 1)[6]	1.18 ± 0.3	2.86 ± 2.5	1.69 ± 2.5	1.75 ± 0.7	3.72 ± 2.1	1.98 ± 2.2	.420
False starts[7]	1.31 ± 1.5	3.0 ± 4.6	1.69 ± 3.6	0.75 ± 0.6	0.9 ± 0.9	−0.19 ± 1.3	.458
No. of lapses (≥ 500 ms) over 10 min test	0.12 ± 0.2	3.2 ± 5.4	3.1 ± 5.5	0.62 ± 0.6	4.4 ± 4.5	3.7 ± 4.7	.500
Dwell time: total time of lapses over 10 min[8]	0.07 ± 0.1	3.0 ± 5.8	2.9 ± 5.8	0.47 ± 0.5	3.2 ± 3.2	2.7 ± 3.4	.581

Pre- and post-intervention characteristics by treatment group, in which treatment group is either sleep-deprived or alcohol-challenged pre- and post-intervention. Pre-intervention is under normal conditions and post-intervention is either alcohol-challenged or sleep-deprived. For the sleep-deprived group, pre-intervention data are quantified as the mean of the (pre-driving) and (post-driving) PVT tests, whereas post-intervention data are computed also as the mean of the (pre-driving) and (post-driving) PVT tests. In the alcohol-challenged group, means of the PVT tests are computed identically as in the sleep-deprived group for pre-intervention and post-intervention data in which the mean BAC was 0.089 g/dL during the post-intervention. *P* values are based on analyses of covariance which compare post-intervention values in the two groups after adjusting for pre-intervention values.

* Reaction time (RT) is in milliseconds (ms) unless otherwise noted.

[1] Standard deviation of the mean reaction time. The SD is computed for each subject, and tabulations reflect the average SD for all subjects.

[2] Mean of the reciprocal of RT computed by first translating each RT into seconds by dividing the number of milliseconds by 1000 (RT/1000 ms) for each 10-minute PVT trial. The mean of the reciprocal of all reaction times (measured in seconds) is computed to generate a single mean value for the given PVT. The tabulated mean of the reciprocal of RT/1000 is then computed as the mean of these values for all subjects.

[3] Maximum RT (fastest 10%): The fastest 10% in ms of the 80–90 responses over 10 minutes.

[4] Minimum RT (slowest 10%): The slowest 10% in ms of the 80–90 responses over 10 minutes.

[5] Reciprocal of the RT of the slowest 10% in ms. Computed precisely as was described above (#2), except that only the slowest 10% RT for a given PVT are included in the computation.

It is important to keep in mind that when reciprocals are used, slower reaction times (RT) are associated with lower values.

[6] Square Root Transforms of lapses [SQRT(Lapses)+SQRT(Lapses+1)] Lapse defined as an RT ≥ 500 ms.

[7] False starts are no stimulus hit correct key or no stimulus hit incorrect key.

[8] Dwell time is the cumulative time in seconds that is ≥ 500 ms during a 10-minute PVT trial. It is calculated for each minute (minute stats). A mean of the total 10 minute stats for each subject is tabulated and the dwell time is computed as the mean of these values for all subjects.

SD = standard deviation; PVT = psychomotor vigilance testing; BAC = blood alcohol concentration.

groups were not significantly different, and with few exceptions the magnitude of change was nearly identical.

Sleep Periods

Because of its relevance to the sleep-deprived subjects in this study, we carefully studied measures of sleepiness and sleep in the eight sleep-deprived subjects. Sleep-time data were measured by each subject using nightly sleep logs during two periods, the "prestudy" period, which included the 5 nights before the study began, and the "study" period, which included 7 nights after commencement of the study.

In the ASD subgroup, the prestudy period was concluded immediately before the night of sleep deprivation and therefore consisted of normal nights of sleep. In the CSD group (deprived of 2 hours of sleep per night), the self-reported number of hours of sleep during the prestudy period (mean ± SD) was 8.1 ± 2.0 hours in all eight subjects. In the ASD subgroup, mean sleep time was reported as 9.0 ± 2.5 hours during the prestudy period and 8.3 ± 0.2 hours during the study period, a small decrease of 0.7 hours per night, which could be explained entirely by one subject (truck driver) whose self-reported sleep time was 12.5 hours during the prestudy period. In the CSD subgroup, mean sleep time decreased from 7.2 ± 0.7 hours during the prestudy period to 4.9 ± 0.3 hours during the study period, a decrease of 2.3 hours per night, which reflected the sleep deprivation that was the essence of the study design in these subjects. Comparison of the four CSD sleep-deprived subjects' total sleep time estimated by *actigraphy* and *sleep logs* revealed no significant differences by nonparametric sign tests.

Epworth Sleepiness Scale

In the entire group of eight CSD and ASD subjects, mean ± SD ESS score increased from 9.9 ± 4.4 when measured in the morning before sleep deprivation to 14.4 ± 5.4 when measured in the morning (7 d later) after either chronic or acute sleep deprivation. The increase (indicative of a worse performance) in the mean ESS score of 4.5 ± 3.8 was statistically significantly (P = .012). Although the deterioration in the ESS score was somewhat greater in the ASD subgroup (5.5 ± 4.4) than in the CSD subgroup (3.5 ± 3.5), this did not approach statistical significance (P = .502). The mean ESS score at baseline in the eight subjects in the alcohol-challenged group was 8.3 ± 4.1; the final mean ESS score in this group was 8.1 ± 3.4.

Visual Analogue Scale Scores

The four CSD subjects were asked to complete VAS testing, which measured on a daily basis the quality of their sleep during the previous night and how sleepy they were feeling during the day. These scales were completed during the 7 days of sleep deprivation. Table III indicates that with minor exceptions there was a steady increase in the VAS score over time. The night-time score increased from 2.05 ± 2.2 (mean ± SD) on day 1 to 4.40 ± 3.8 on day 7. The daytime score began at a much higher 5.85 ± 2.8; nonetheless, there was a steady increase in daytime tiredness, which reached a score of 7.07 ± 1.3 by day 7. The tabulated slopes indicated that the mean rate of increase in the night-time score was 0.37 ± 0.36 VAS units per day, whereas the daytime increase was 0.18 ± 0.32 units per day. Although sleep quality and sleepiness deteriorated steadily as measured by the VAS testing, neither of the changes was statistically significant (P = .131 for the night-time score and P = .324 for the daytime score).

In interpreting all of the data just given, it is important to note that the small sample size in the study means the statistical power to detect associations was severely limited. Thus, large true differences might be reported as "not significant." In the case of the regression slopes just discussed, the steady increases over time suggest that chronic sleep deprivation may well be associated with a deterioration in self-reported tiredness and/or sleep quality as measured by the VAS testing. However, the sample size is too small to yield statistical support for that conclusion.

Driving Performance

The tally of performance was measured by the number of cones hit, episodes off track (error = outside of cone field), and reactions to a stoplight and a yellow flashing arrow. In addition, hazards consisting of a medium-sized beach ball thrown across the road in front of the car, a simulated toy dog pulled across the car's pathway, and a pop-up car were included only in the final driving test. Contact with these hazards was also calculated as an error. The driving performance of each group separately, as well as between-group analysis, is shown before and after intervention in Table II. The change scores between groups were compared (P = .724). A further breakdown of the individual group scores follows and can be seen in Table IV.

Sleep-Deprived Group

During the first trial (baseline) the total number of cones hit was zero. The number of cones hit up to and including the flashing yellow arrow was six. In the final trial (sleep deprived), the total number of cones hit was 20. The number of cones hit up to and including the flashing yellow arrow was 21. The number of cones hit up to the flashing yellow arrow and including hazards (ball to car) was 27. Comparing within-group changes, from the first trial to the final trial in "number of cones hit only" there was a 20-fold increase; from number of cones hit up to the flashing yellow arrow, a 3.5-fold increase; and from the number of cones hit up to the car hazard (final hazard), a 4.5-fold increase.

Alcohol Group

On the first trial (baseline) the total number of cones hit was 3. The number of cones hit up to and including the flashing yellow arrow was 11. On the final trial (alcohol challenged), in which mean BAC at the time of driving was 0.089 g/dL (basically, too intoxicated to drive legally in many states in the United States), total number of cones hit was 24. Number of cones hit up to and including the flashing yellow arrow was 32. Number of cones hit up to the flashing yellow arrow including hazards (ball to car) was 38. Comparing within-group changes, from the first trial to the final trial in "number of cones hit only" there was an 8-fold increase; in number of cones hit up to the flashing yellow arrow, a 2.9 fold increase; and in number of cones hit up to the car hazard (final hazard), a 3.4 fold increase.

Mood and Behavior

There were differences in subject behavior (mood and appearance), both observed and filmed, among all three groups (ASD, CSD, and alcohol-challenged). Overall, the sleep-deprived group struggled to focus or stay awake and in effect exhibited "*fitful*" behavior. Subjects in the ASD subgroup found it far harder to stay awake and was considerably more irritable and disheveled than subjects in the CSD subgroup. In contrast, the alcohol-challenged drivers appeared "*frivolous or carefree*" and uninhibited during their driving task. Subjects in all three groups generally did not anticipate the effect intervention would

TABLE III.
Mean Visual Analog Scale (VAS) Scores for the Chronic Sleep-Deprived.

Day	Previous Night VAS Score	Daytime VAS Score
1	2.05 ± 2.2	5.85 ± 2.8
2	2.37 ± 2.3	5.87 ± 0.6
3	2.55 ± 2.8	6.27 ± 1.0
4	4.60 ± 3.5	6.45 ± 1.1
5	3.22 ± 2.2	5.52 ± 2.2
6	3.45 ± 2.04	7.32 ± 1.5
7	4.40 ± 3.8	7.07 ± 1.3
Mean slope	0.37 ± 0.36	0.18 ± 0.32
P value for slope*	0.131	0.324

* Tests the hypothesis that the slope = 0.

VAS scores measured the quality of sleep at night and the degree of tiredness on the following day.

Scores range from 0 to 10 with higher numbers reflecting increased tiredness or worse sleep. Data are tabulated for the four subjects in the chronic sleep deprivation (CSD) subgroup who were deprived of 2 hours of sleep for 7 consecutive nights. Tabulations are for each of those 7 *nights* and for the succeeding *days*.

TABLE IV.
Closed Course Driving Performance.

Subjects and Trial	Cones # Hit	Off Track	Stop Light	Yellow Arrows	Ball†	Dog†	Car†	Trial 1 Cones to Arrows	Trial 2 Cones to Arrow	Trial 2 Cones to Car
Sleepy drivers										
ASD + CSD (n = 8)										
Totals trial 1	0	2	3	1	–	–	–	6		
Totals trial 2*	20	0	1	0	2	2	2		21	27
Alcohol drivers										
AC (n = 8)										
Totals trial 1	3	5	1	2	–	–	–	11		
Totals trial 2	24	4	3	1	3	2	1		32	38

* Subject 7 in the acute group started the final driving session, but as a result of severe sleepiness and an inability to safely maintain wakefulness and vigilance, voluntarily pulled over late in the session and did not complete the driving course.
† Hazards only presented on trial 2.
ASD = acute sleep-deprived; CSD = chronic sleep-deprived; AC = alcohol-challenged.

have on their driving ability, nor did they accurately assess their driving performance after intervention.

DISCUSSION

This study was prospectively planned to be presented as a Dateline NBC News production and as a Stanford University Sleep Research investigation in cooperation with General Motors. It was designed as a feasibility study for us and was, by the Dateline NBC News protocol, limited in the number of subjects. We recognize and acknowledge that the study was limited in numbers and thus in statistical power. We further note that previous studies of driving performance and sleepiness were for the most part performed in the laboratory with driving simulators.[19–21] These studies also had relatively small sample sizes, yet reported important results. Hence, this study was undertaken as a preliminary first-step investigation because, to our knowledge, this type of cooperative study had not been previously performed. The stratification of the sleep-deprived group into ASD and CSD subgroups, along with a precise limit of sleep deprivation, was performed to present data on sleep deprivation as is common in our society.

Chronic sleep deprivation is rarely investigated; hence, this limited our ability to predict possible outcomes before the study was conducted. We suspected that the CSD subgroup would perform slightly worse than the ASD subgroup. However, objective outcomes measures between the ASD and CSD subgroups were the same *except* for mood and behavior. Because of their equivalent objective test outcomes, both subgroups (ASD and CSD) were combined and their cumulative data were used for comparison with the alcohol-challenged group. Results in previous studies of laboratory trials in sleepy subjects were consistent with the altered performance we report in the present study.[22,23]

The CSD subgroup had a different behavioral course than the ASD subgroup, although the final performance outcomes were the same. Examination of the CSD group's behavior, witnessed by the investigators and documented by NBC taped interviews and written transcripts, showed that at day 2 to 3, all CSD subjects were severely compromised by their then-cumulative 4 to 6 hours of total sleep loss (2 h nightly). However, by day 7, each subject symptomatically appeared to have improved and thought he or she would do well in the final trial. We suggest that this occurred either by compensatory effort or adaptation, or both. In clinical practice we have seen this adjustment process in patients with sleep-disordered breathing. Interestingly, subjects in all groups (CSD and ASD subgroups and alcohol-challenged group) demonstrated that they were, in fact, basically unaware of the magnitude of their decrements of alertness. This is consistent with the findings reported by Dinges[24] for sleepy drivers. He suggested that drivers know they may be sleepy but cannot reliably predict when they are impaired.

A statistical comparison of the driving performance between these two groups (sleep-deprived subjects vs. alcohol-challenged subjects) had limitations because of the distances that each group completed, even though both groups drove the same coned course, with minor exceptions, on the coned-off test pad. Attempts were made to adjust or normalize the distances by deleting the 1.2-mile back section, which was not coned (driven only by the sleep-deprived group), along with equalizing the distance by a percentage factor. Because of the small number of subjects in each group, statistical power was limited and none of these adjusted measures could be relied on with confidence. However, driving performance decrements were demonstrated in each group from baseline through the final driving trial; change scores are listed in Table II. For example, using only the cone point count there are 20-fold (sleep-deprived group) and 32-fold (alcohol-challenged group) increases in hits (Table IV). This suggests that there was a marked deterioration in vigilance within each group. Because there certainly was some advantage on driving the second trial as a result of a "practice effect" during the first trial, the scores for the second trial could have been slightly worse than recorded. The methodology for this type of feasibility study was established in a situation in which there were no previous sleepiness studies available that addressed design and safety for closed-course driving performance.

We found that the between-group RT metrics were, in all respects, statistically the same (baseline through final values after intervention). A closer examination of the dwell-time RT metric, which represents periods of "microsleeps," emphasizes the risk of driving while sleepy. Table II reports these dwell-time values (cumulative seconds equal to or greater than 500 ms) for all groups. Values for all groups were prolonged over baseline. These 500-ms microsleeps are usually associated with eyelid closure and herald the onset of inattention, sleepiness, or fatigue. The change scores from baseline to the period after intervention are revealing. The alcohol-challenged group had a cumulative mean score of 2.9 ± 5.8 seconds, and the sleep-deprived group had a cumulative mean score of 2.7 ± 3.4 seconds ($P = .581$ between groups). This was equal to a mean total lapse over a 10-minute test period of 2.8 seconds (2800 ms) or a mean number of repetitive lapses (equal to or greater than 500 ms) of 5.6. At 60 mph (88 feet/s), the subjects would have traveled a total of 246.4 feet while being essentially *unaware* of their surroundings and unable to respond.

CONCLUSION

To more clearly demonstrate the potential risks resulting from sleepiness, we investigated time-on-task RT and closed-course driving performances in a healthy cohort challenged with either sleep loss or alcohol consumption and then contrasted these outcomes as a comparative model. All 11 RT outcome metrics clearly showed that the sleepy drivers were the same as the alcohol-challenged drivers on this time-on-task performance. The driving performance for both groups (sleep deprived vs. alcohol challenged) were similar, but their behavior and mood were different ("fitful" vs. "frivolous"). Yet, both groups thought they did well on performance. The objective data show otherwise. This comparative model suggests that driving while sleepy should be recognized as potentially dangerous or as at least as dangerous as driving illegally under the influence of alcohol. It is important that individuals accept the responsibility to avoid engaging in safety-sensitive activities while sleepy.

Acknowledgments

This investigation was conducted in cooperation with Stanford Sleep Disorders and Research Center of Stanford University, Dateline NBC News, and General Motors.

BIBLIOGRAPHY

1. National Commission on Sleep Disorders Research. *Report of the National Commission on Sleep Disorders Research*. Washington, DC: Superintendent of Documents, US Government Printing Office, 1992. Department of Health and Human Services Publication 92.
2. D'Alessandro R, Rinaldi R, Cristina E, Gamberni G, Lugaresi E. Prevalence of daytime sleepiness: an open epidemiological problem. *Sleep* 1995;18:389–391.
3. Dinges DF, Pack F, Williams K, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night. *Sleep* 1997;20:267–277.
4. Kribbs N, Getsy J, Dinges D. Investigation and management of daytime sleepiness in sleep apnea. In: Saunders N, Sullivan C, eds. *Sleep and Breathing.* New York: Mercel Dekker, 1994:575–604.
5. Weaver T, Laizner A, Evans L, et al. An instrument to measure functional status outcomes for disorders of excessive sleepiness. *Sleep* 1997;20:835–843.
6. Mitler M, Carskadon M, Czeisler C, Dement W, Dinges D, Graeber R. Catastrophes, sleep, and public policy: consensus report. *Sleep* 1988;11:100–109.
7. Findley L, Unverzagt M, Suratt P. Automobile accidents involving patients with obstructive sleep apnea. *Am Rev Respir Dis* 1988;138:337–340.
8. Aldrich M. Automobile accidents in patients with sleep disorders. *Sleep* 1989;12:487–494.
9. Chang AW, Kushida C, Palombini L, et al. Comparison study of actigraphic, polysomnographic, and subjective perception of sleep parameters [abstract H071.N]. *Sleep* 1999; 22(Suppl):S43.
10. Jette AM. The functional status index: reliability and validity of a self-report functional disability measure. *J Rheumatol* 1987;(Suppl 15):15–21.
11. Streiner D, Norman G. Scaling responses. In: *Health Measurement Scales: A Practical Guide to Their Development and Use*, ed 2. Oxford: Oxford University Press, 1995: 31–53.
12. Babul N, Darke A, Johnson D, et al. Using memory for pain in analgesic research. *Ann Pharmacother* 1993;27:9–12.
13. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–545.
14. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;13:376–381.
15. Dinges D, Powell J. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments and Computers* 1985;17:652–655.
16. Kribbs NB, Dinges DF. Vigilance decrement and sleepiness. In: Harsh JR, Ogilvie RD, eds. *Sleep Onset Mechanisms.* Washington, DC: American Psychological Association, 1994:113–125.
17. Kribbs NB, Pack AI, Kline LR, et al. The effects of one night without CPAP treatment on sleep and sleepiness in patients with obstructive sleep apnea syndrome. *Am Rev Respir Dis* 1993;147:1162–1168.
18. 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.
19. Findley LJ, Suratt PM, Dinges DF. Time-on-task decrements in 'steer clear' performance of patients with sleep apnea and narcolepsy. *Sleep* 1999;22:804–809.
20. George CFFP, Boudreau AC, Smiley A. Simulated driving performance in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 1996;154:175–181.
21. Risser MR, Ware C, Freeman FG. Driving simulation with EEG monitoring in normal and obstructive sleep apnea patients. *Sleep* 2000;23:393–398.
22. Powell NB, Riley RW, Schechtman KB, et al. A comparative model: reaction time performance in sleep-disordered breathing versus alcohol-impaired controls. *Laryngoscope* 1999;109:1648–1654.
23. Williamson AM, Feyer AM. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occup Environ Med* 2000;57:649–655.
24. Dinges DF. The nature of sleepiness: causes, contexts, and consequences. In: Stunkard AJ, Baum A, eds. *Eating, Sleeping, and Sex*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1989:147–179.