Impaired Performance in Commercial Drivers
Role of Sleep Apnea and Short Sleep Duration

Allan I. Pack, Greg Maislin, Bethany Staley, Frances M. Pack, William C. Rogers*, Charles F. P. George, and David F. Dinges

Center for Sleep and Respiratory Neurobiology; Division of Sleep Medicine; Division of Pulmonary Allergy and Critical Care, Department of Medicine; Division of Sleep and Chronobiology, Department of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania; Federal Aviation Administration, Washington, DC; and London Health Sciences Centre, University of Western Ontario, London, Ontario, Canada

Sleepiness plays an important role in major crashes of commercial vehicles. Because determinants are likely to include inadequate sleep and sleep apnea, we evaluated the role of short sleep durations over 1 wk at home and sleep apnea in subjective sleepiness (Epworth Sleepiness Scale), objective sleepiness (reduced sleep latency as determined by the Multiple Sleep Latency Test), and neurobehavioral functioning (lapses in performance, tracking error in Divided Attention Driving Task) in commercial drivers. Studies were conducted in 247 of 551 drivers at higher risk for apnea and in 159 of 778 drivers at lower risk. A multivariate linear association between the sets of outcomes and risk factors was confirmed (p < 0.0001). Increases in subjective sleepiness were associated with shorter sleep durations but not with increases in severity of apnea. Increases in objective sleepiness and performance lapses, as well as poorer lane tracking, were associated with shorter sleep durations. Associations with sleep apnea severity were not as robust and not strictly monotonic. A significant linear association with sleep apnea was demonstrated only for reduced sleep latency. The effects of severe apnea (apnea–hypopnea index, at least 30 episodes/h), which occurred in 4.7%, and of sleep duration less than 5 h/night, which occurred in 13.5%, were similar in terms of their impact on objective sleepiness. Thus, addressing impairment in commercial drivers requires addressing both insufficient sleep and sleep apnea, the former being more common.

Keywords: commercial drivers; excessive sleepiness; obesity; obstructive sleep apnea; sleep duration

In the United States, approximately 5,600 people are killed annually in crashes involving commercial trucks (1). Falling asleep while driving is an important factor in serious crashes involving commercial vehicles (2, 3), prompting the question, why? There are likely to be at least two mechanisms. The first is chronically insufficient sleep, that is, excessive sleepiness and performance impairments from accumulating sleep debt when individuals curtail the duration of their sleep day after day (4–6). Short sleep durations have previously been described in commercial drivers (7). The second mechanism is obstructive sleep apnea, which has been found to be common in commercial drivers (8, 9). This study was conducted to assess the relative role of these risk factors for impairments in commercial driver's license holders. Daytime sleepiness and performance were assessed using neurobehavioral tests relevant to the driving task. These were evaluated relative to laboratory measures of sleep-disordered breathing and objective estimates of sleep duration at home. There are currently no published data to guide policy and decision makers regarding performance assessment, or regarding factors associated with impaired performance in drivers of commercial vehicles. The results presented in this article have previously been reported in abstract form (10, 11). There has also been a publication, using data collected from this study, that focused on screening for apnea in this population (12).

METHODS

Sample Studied

Stratified sampling (13) was employed to enrich the sample with respect to the presence of sleep apnea (see Figure E1 in the online supplement). Government agencies provided contact information for n = 4,826 randomly selected holders of commercial driver's licenses living within 50 miles of the Center for Sleep and Respiratory Neurobiology (University of Pennsylvania, Philadelphia, PA). Individuals received, by mail, the Multivariable Apnea Prediction questionnaire (14) to facilitate oversampling of drivers at higher risk for sleep apnea. Information from this questionnaire included age, sex, body mass index (BMI), and an apnea symptom–frequency index determined from three individual questions and was used to determine the relative likelihood of apnea on a scale from 0 to 1. Complete responses were obtained for 1,329 of 4,410 (30.1%) valid contacts. By design we recruited n = 247 of 551 (44.8%) highest risk drivers and then, in randomized order, n = 159 of 778 (20.4%) lower risk drivers. Rationale and additional details for this design including a schematic diagram (see Figure E1) are provided in the online supplement. The study was approved by the Institutional Review Board of the University of Pennsylvania and all subjects gave written, informed consent.

Impairment Outcomes

The Epworth Sleepiness Scale (15–18) was used to assess subjective sleepiness (19). The Multiple Sleep Latency Test (MSLT) (20) was used to objectively determine propensity to fall asleep. A standardized Psychomotor Vigilance Task (PVT) assessed behavioral alertness (21, 22), defining vigilance lapses as reaction times of 500 ms or more (22). More than three lapses per 10-min trial (four trials) defined impairment, based on a dose–response study of the effects of alcohol (23). The mean (four trials) average tracking error over 30 min on the Divided Attention Driving Task (DADT) was used to assess driving performance (24, 25). Impairment was defined as a mean absolute tracking error of greater than 250 cm, based on studies of patients with untreated obstructive sleep apnea and healthy control subjects ingesting alcohol (25). Additional details regarding this task are provided in the online supplement.

Risk Factors for Impairment

Average sleep duration. Mean sleep duration at home was assessed over 1 wk by actigraphy (Ambulatory Monitoring, Ardsley, NY). Daily sleep diary data were collected and used in the interpretation of actigraphy data but were not systematically assessed on their own. Mean sleep

duration was estimated as the mean cumulative duration of relative inactivity during the main sleep bout. Sleep duration among apneics might be underestimated because some movements associated with apnea events during sleep might be interpreted as wakefulness (26). Therefore, Bland-Altman (27) agreement analyses were performed, comparing actigraphy-based sleep estimates collected during in-laboratory polysomnography with simultaneously recorded EEG-determined sleep. The mean bias was 0.31 h (95% confidence interval, 0.16–0.47 h) and the standard deviation of differences was 1.30 h. Additional details of the validation analyses are provided in the online supplement, including an agreement analysis for an alternative measure of sleep duration based on a bias-adjusted length of the main sleep bout (*see* Table E2). Primary analyses assessed mean sleep duration as a continuous risk factor. In supporting analyses, mean sleep duration was assessed in simple categories (< 5, 5 to < 6, 6 to < 7, 7–8, and > 8 h). Multivariable analyses using the alternative measure of sleep duration are also provided (*see* Table E3).

Sleep apnea. Full overnight in-laboratory polysomnography using multichannel recordings of sleep, based on EEG/EMG recordings and breathing, was performed in all subjects with airflow measured with thermistors (12). The apnea–hypopnea index (AHI; number of apneas plus hypopneas per hour of sleep) was defined on the basis of current criteria (12, 28) and assessed as a continuous risk factor in primary analyses. In secondary and complementary analyses, the effects of sleep apnea severity on sleepiness and performance outcomes were assessed in terms of suggested categories (28): mild (AHI, at least 5 and < 15 events/h), moderate (≥ 15 but < 30 events/h), and severe (≥ 30 events/h). The online supplement includes details regarding variables measured during in-laboratory polysomnography.

Statistical Approaches

Weighted analyses designed to account for stratified random sampling (29) were used only when describing the pooled sample of higher and lower risk drivers to provide summaries relevant to the target population. Because the sampling frame was constructed to ensure that the sample consisted of participants with a range of apnea, and the target population for inferential statistics is not the higher and lower risk strata, multivariate and multivariable regression analyses did not incorporate sample weights. An omnibus test of the overall study hypothesis that there is an association between at least one of the impairment outcomes with either sleep duration or sleep apnea severity was based on canonical correlation analysis ($\alpha = 0.05$). Vigilance lapses and mean tracking error were transformed ($\log_e[1 + x]$) to better fit parametric assumptions in all analyses with results presented on the original scales to improve clinical interpretability. The AHI was transformed in the multivariate canonical correlation analysis but not when used as a predictor variable in regression models. Primary multivariable regression models for each outcome expressed risk factors as continuous variables to maximize statistical power to detect hypothesized monotonic associations. Statistical interactions between the two continuous risk factors were assessed by adding a cross-product term to each model. None of these interactions was statistically significant and so results are not further discussed. To assess nonlinear associations and for simplicity of exposition, sleep apnea severity and average sleep duration at home were also expressed as categorical variables in secondary multivariable models. All regression models simultaneously included factors for apnea severity and sleep duration and also simultaneously controlled for age, sex, and obesity in terms of BMI as covariates. Statistical comparisons across apnea severity categories and sleep duration categories were performed, controlling for this same set of variables. Nominal significance levels (unadjusted for multiple comparisons) were reported only for planned contrasts versus reference categories (< 5 events/h for AHI and 7–8 h for mean sleep duration). A sleep duration reference of 7–8 h rather than longer durations was chosen because of the relative small number of drivers with mean duration exceeding 8 h (n = 17), and also out of concern that mean durations exceeding 8 h might be associated with poorer rather than better health status as observed by Kripke and coworkers (30). Significance levels were also determined for both planned contrasts as well as the remaining pair-wise (unplanned) contrasts, adjusting for multiple comparisons by using the simulation method of Edward and Berry (31). This method has greater statistical power than alternative methods such as the Bonferroni adjustment.

Least-squares adjusted mean values were determined by setting each covariate to its respective sample mean.

RESULTS

Description of Sample Demographics

Table E1 summarizes the sample; 93.3% of the sample were male and the mean age was 45.4 yr. In addition, 81.6% were employed as drivers of a commercial vehicle at the time of study enrollment. Of these, 9.3% reported exclusively long-distance driving, 65.5% reported local driving, and 25.2% reported both. Primary analyses were based on the total sample because those not working as commercial drivers at the time of the study had the capacity to do so. Similar results were obtained when analyses were repeated for the 81.6% currently working as commercial drivers. These results are not presented here but are summarized in the online supplement (*see* Table E4).

Description of Sleep Apnea and Sleep Durations

The percentages of drivers with mild, moderate, or severe sleep apnea are summarized in Table 1. It was found that 28.2% had at least mild sleep apnea by conventional criteria, and 4.7% had severe sleep apnea. The percentages of drivers with different durations of mean cumulative duration of relative inactivity are also shown in Table 1. The (weighted) group mean was 6.33 h (SD, 1.34). Table 2 provides a summary of the outcomes for Epworth Sleepiness Scale score and objective impairments. Additional details including summaries by apnea risk strata are included in Table E1.

Test of Overall Study Hypothesis

The multivariate linear association between outcomes and risk factors was significantly different (Wilks's lambda F[8, 594] = 5.7;

TABLE 1. SUMMARY OF SLEEP APNEA SEVERITY AND SLEEP DURATION

	n	Median (*IQR*)[†]	Percentiles (*5th, 95th*)
AHI, events/h	406	1.6 (0.4, 6.1)	0, 29.4
Actigraphy-estimated sleep duration, h	340	6.6 (5.6, 7.3)	3.9, 8.1

	n*	Percentage of Weighted Sample[‡]	(*95% CI*)[§]
Apnea severity category			
AHI, ≥ 30	28	4.7	3.5, 5.9
AHI, 15 to < 30	32	5.8	4.2, 7.3
AHI, 5 to < 15	86	17.7	14.7, 20.7
AHI, ≤ 5 (reference category)	260	71.8	68.6, 75.1
Actigraphy: Mean cumulative duration of relative inactivity			
< 5 h	55	13.5	10.5, 16.4
5 to < 6 h	75	21.8	17.8, 25.8
6 to < 7 h	104	29.8	25.3, 34.2
7 to 8 h (reference category)	89	28.7	24.1, 33.2
> 8 h	17	6.2	3.6, 8.9

Definition of abbreviations: AHI = apnea–hypopnea index; CI = confidence interval; IQR = interquartile range.

All estimates are adjusted for stratified sampling by risk group.

* n = number of subjects within the specified sleep apnea severity category or mean cumulative duration of relative inactivity.

[†] (25th percentile, 75th percentile). The IQR is defined as the difference between the 25th and 75th percentile values.

[‡] The weighted percentage is computed as (0.415 × higher risk [%]) + (0.585 × lower risk [%]).

[§] Numbers represent lower and upper bounds of 95% confidence intervals for percent values, accounting for stratified sampling.

TABLE 2. SUMMARY OF IMPAIRMENT OUTCOMES

	n	Median (IQR)*	Percentiles (5th, 95th)
Epworth Sleepiness Scale, total score	386	8.0 (6.0, 12.0)	2.0, 17.0
Multiple Sleep Latency Test, min	406	7.5 (4.9, 11.1)	2.1, 15.2
Psychomotor Vigilance Task, lapses/trial	402	1.8 (0.8, 3.6)	0.0, 6.2
Divided Attention Driving Task, mean tracking error	379	213.5 (164.3, 292.0)	102.2, 528.2
	n	Percent†	95% CI‡
Psychomotor Vigilance Task, lapses/trial			
> 3 lapses	111	29.2	25.1, 33.4
⩽ 3 lapses	291	70.8	66.6, 74.9
Divided Attention Driving Task, mean tracking error			
> 250 cm	142	36.9	32.5, 41.4
⩽ 250 cm	237	63.1	58.6, 67.5
Multiple Sleep Latency Test			
< 5 min	111	25.7	21.9, 29.4
⩾ 5 min	295	74.3	70.6, 78.1
Number of impairments			
3	22	5.3	3.3, 7.3
2	80	19.6	16.1, 23.2
1	138	34.6	30.3, 38.9
0	166	40.5	36.1, 44.9

Definition of abbreviations: n = number of subjects with complete data for specified outcome.

All estimates are adjusted for stratified sampling by risk group.

* The IQR is defined as the difference between the 25th and 75th percentile values.

† (25th percentile, 75th percentile). The weighted percentage is computed as (0.415 × higher risk [%]) + (0.585 × lower risk [%]).

‡ Lower and upper bounds, respectively, of 95% confidence intervals for percent values, accounting for stratified sampling.

p < 0.0001), revealing 14.2% shared variance between the two sets of variables.

Results from Primary Continuous Risk Factor Models

Details of the primary continuous risk factor models for the associations between sleep apnea severity and mean sleep duration and each of the four measures of sleepiness and performance are summarized in Table 3. Details of comparisons between unadjusted results and those presented in Table 3 are provided in the online supplement. Before covariate adjustment, increases in sleep apnea severity were significantly associated with increased objective sleepiness (MSLT) (estimated slope, –0.88; p = 0.0002) and tracking error on the DATD (estimated slope, 0.10; p = 0.001). After adjusting for covariates only the association with MSLT remained significant (Table 3). Each increase of 11.5 events/h (the standard deviation for this variable) in the AHI was associated with a 0.54 (SE, 0.27)-min reduction in the expected value of the MSLT (p = 0.05), controlling for mean sleep duration at home, age, sex, and BMI.

In contrast, reduced sleep durations were significantly associated with increases, with or without adjustment, in subjective sleepiness (p = 0.01), increases in objective sleepiness (p = 0.008), increases in vigilance lapses (p = 0.0001), and increases in tracking error (p = 0.0001) (all p values after covariate adjustment). For each 1.34-h decrease in estimated mean sleep time (the standard deviation for this variable), there was a 0.72 (SE, 0.28)-point increase in the predicted Epworth Sleepiness Scale total score, a 0.75 (SE, 0.28)-min reduction in the predicted value of MSLT, a 0.18 increase in the predicted value of the log of 1 plus the number of vigilance lapses per 10-min trial, and a 0.11 increase in the predicted value of log of 1 plus mean absolute tracking error on the DADT.

Assessment of Risk Factors in Categories for Different Outcomes

Subjective sleepiness. Results of the categorical risk model are also summarized in Table 3. Least-squares estimates of mean Epworth Sleepiness Scale scores by sleep apnea severity category were as follows: no apnea, 8.9; mild apnea, 8.5; moderate apnea, 9.4; and severe apnea, 9.3. Differences among adjusted means were not statistically significant (F[3, 313] = 0.3; p = 0.81).

In contrast, adjusted mean values for Epworth Sleepiness Scale scores among sleep duration categories were as follows: 10.0, 9.9, 8.6, 7.9, and 8.7 for mean durations of less than 5, 5 to less than 6, 6 to less than 7, 7 to 8, and more than 8 h, respectively, which did vary significantly (F[4, 313] = 2.5; p = 0.05). Planned contrasts versus the reference category of 7 to 8 h achieved nominal statistical significance and also appeared clinically significant in magnitude (∼ 2 Epworth points) relative to less than 5 h and 5 to less than 6 h of sleep. However, after adjustment for multiple comparisons, none of the pair-wise comparisons among sleep duration categories was statistically significant.

Objective sleepiness. Least-squares estimates for mean MSLT (minutes) by sleep apnea severity category were as follows: no apnea, 8.8; mild apnea, 9.3; moderate apnea, 7.9; and severe apnea, 6.3 min (F[3, 329] = 2.4; p = 0.07). Category differences are illustrated in Figure 1A. Nominal p values for planned contrasts versus reference category of fewer than 5 events/h are provided in Table 3. The difference in expected MSLT values between those with severe sleep apnea and those with no sleep apnea was –2.46 min (SE, 1.15; p = 0.03), controlling for all other variables in the model. However, after adjustment for multiple comparisons, only the difference between moderate and severe apnea categories remained statistically significant (p < 0.05).

Adjusted mean values for the MSLT among sleep duration categories were as follows: 6.3, 7.9, 7.8, 8.6, and 9.9 for mean durations of less than 5, 5 to less than 6, 6 to less than 7, 7 to 8, and more than 8 h, respectively (F[4, 329] = 2.5; p = 0.04). These are illustrated in Figure 1B. Nominal significance relative to reference (7–8 h) was found for less than 5 h (p = 0.007). After controlling for multiple comparisons, the differences in means between those with less than 5 h versus the reference

TABLE 3. SUMMARY OF MULTIPLE LINEAR REGRESSION MODEL RESULTS

Variable	Epworth Sleepiness Scale			Multiple Sleep Latency Test			Performance Lapses[†]			DADT Tracking Error[†]		
	Slope	SE	p Value	Slope	SE	p Value	Slope	SE	p Value	Slope	SE	p Value
Continuous risk factor model												
Apnea–hypopnea index												
(SD = 11.5/h)	0.01	0.27	0.96	**−0.54**	**0.27**	**0.05**	0.03	0.04	0.45	0.06	0.03	0.08
Sleep duration (SD = 1.34 h)	**−0.72**	**0.28**	**0.01**	**0.75**	**0.28**	**0.008**	**−0.18**	**0.04**	**0.0001**	**−0.11**	**0.03**	**0.0001**
Categorical risk factor model												
Apnea–hypopnea index category*			0.81			0.07			**0.007**			0.10
> 30 vs. < 5	0.39	1.19	0.74	**−2.46**	**1.15**	**0.03**	0.22	0.16	0.18	**0.27**	**0.13**	**0.04**
15–30 vs. < 5	0.46	0.94	0.62	−0.88	0.96	0.36	−0.07	0.14	0.64	0.09	0.10	0.38
5 to < 15 vs. < 5	−0.40	0.66	0.54	0.58	0.66	0.38	**−0.26**	**0.09**	**0.005**	−0.05	0.07	0.46
Sleep duration category*			**0.05**			**0.04**			**0.005**			**0.05**
< 5 vs. 7–8 h	**2.05**	**0.85**	**0.02**	**−2.29**	**0.85**	**0.007**	**0.39**	**0.12**	**0.001**	**0.26**	**0.09**	**0.003**
5 to < 6 vs. 7–8 h	**1.95**	**0.74**	**0.009**	−0.66	0.74	0.38	0.18	0.10	0.08	0.10	0.08	0.20
6 to < 7 vs. 7–8 h	0.65	0.66	0.33	−0.80	0.67	0.23	0.11	0.09	0.23	0.10	0.07	0.13
> 8 vs. 7–8 h	0.72	1.21	0.55	1.31	1.25	0.29	−0.23	0.17	0.18	−0.02	0.12	0.88

Definition of abbreviations: DADT = Divided Attention Driving Task; SE = standard error of the associated estimate.

Note: Slope refers to estimated model parameters. Slopes or sets of slopes as indicated (p < 0.05) are set boldface to highlight patterns of statistical significance. All multiple linear regression models included age, sex, and body mass index as well as either categorical factors for apnea–hypopnea index (AHI) category and sleep duration category or continuous variables for AHI and sleep duration. Sleep duration was estimated as the mean cumulative duration of relative inactivity as determined by actigraphy over 1 wk. Continuous variables were divided by estimated population standard deviation values so that linear slopes reflected the expected effects of a 1-SD increase. The SD values used were 11.5 for AHI and 1.34 for mean sleep duration as determined by actigraphy over 1 wk at home as the cumulative duration of relative inactivity.

* p Value is for the overall categorical factor (degrees of freedom = 3 for apnea severity category and degrees of freedom = 4 for sleep duration category).

[†] Psychomotor vigilance lapses (performance lapses) and DADT mean tracking error were transformed according to the function $\log_e(x + 1)$ before analysis.

category of 7 to 8 h remained significant (p = 0.05). The multiplicity-adjusted p value comparing less than 5 h with greater than 8 h was p = 0.06. None of the other category differences adjusted for multiple comparisons was statistically significant. There was an estimated –2.29 (SE, 0.85)-min reduction in sleep latency among those with sleep durations less than 5 h compared with the reference group with 7 to 8 h. Thus, the magnitude of the group difference between less than 5 h of average sleep compared with 7 to 8 h was similar in magnitude to that for severe sleep apnea compared with no apnea.

Although sleep latency was reduced in those with severe sleep apnea compared with no sleep apnea, evidence of differential effects of sleep apnea on objective sleepiness was observed among individuals. When we categorized subjects into those with pathologic sleepiness (MSLT, < 5 min), in the "gray zone" (MSLT, at least 5 and < 10 min), or normal (≥ 10 min), we found that 77% of those with mild sleep apnea (5 to < 15 events/h), and 56% with moderate sleep apnea (15 to < 30 events/h) did

not have pathologic sleepiness. Approximately 4% of those with severe apnea (≥ 30 events/h) had completely normal MSLT values of 10 min or more (Figure 2). There was also evidence of substantial differential susceptibility to the effects of short sleep duration on MSLT values (Figure 3).

Behavioral alertness. There were statistically significant differences among apnea severity categories for performance lapses (F[3, 325] = 4.0; p = 0.007). Comparisons among categories are illustrated in Figure 4A after back-transformation to the original scale. Pair-wise contrasts are expressed in terms of the ratio of expected values as a consequence of the log transformation. The regression-adjusted ratios comparing those with 30 or more, 15 to fewer than 30, and 5 to fewer than 15 events/h with those with fewer than 5 events/h were 1.24 (1.17), 0.98 (1.15), and 0.76 (1.09; mean [SE]), respectively. The only significant planned contrast versus the reference category of less than 5 events/h was for AHI at 5 to fewer than 15 events/h (p = 0.005). Unexpectedly, as seen in Figure 4A, those with mild apnea (AHI, 5 to
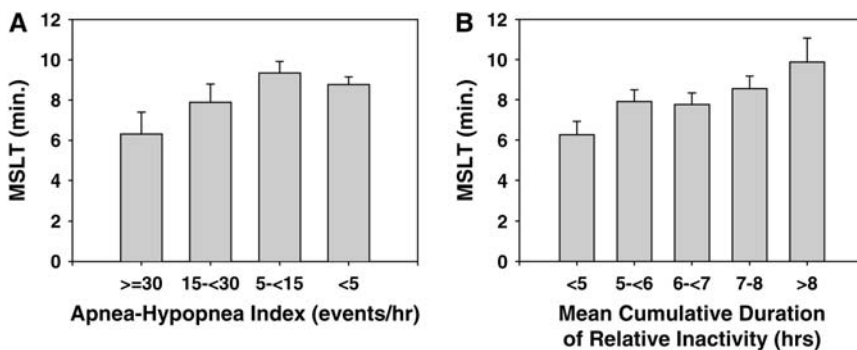

Figure 1. Least-squares estimates of mean Multiple Sleep Latency Test (MSLT) with standard errors controlling for apnea severity category, sleep duration category, age, sex, and body mass index. (*A*) Apnea–hypopnea index category. (*B*) Mean length of cumulative duration of relative inactivity.
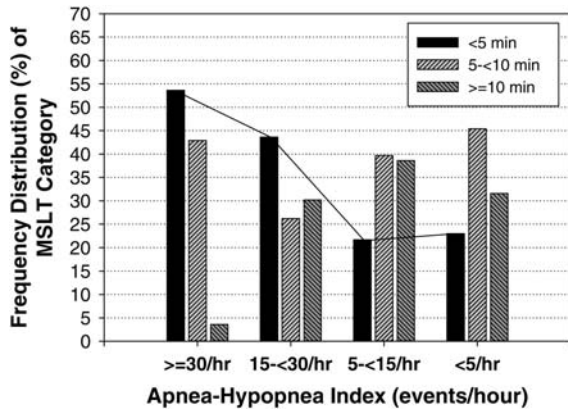
Figure 2. Percentages of subjects with MSLT values of less than 5 min, 5 to less than 10 min, and 10 min or more in different categories of severity of sleep apnea. Percentages were determined on the basis of weighted data. The *line* connects the percentages with pathologic sleepiness, that is, MSLT < 5 min.

fewer than 15 events/h) performed better than those with no apnea. Comprehensive analyses resulted in no evidence of confounding as a cause of this finding. These analyses are briefly described below, with further detail in the online supplement. After adjustment for multiplicity, significant pair-wise contrasts were observed between those with at least 30 events/h versus 5 to fewer than 15 events/h (p = 0.02) and between those with 5 to fewer than 15 events/h versus those with fewer than 5 events/h (p = 0.02).

Least-squares adjusted mean values back-transformed to the original scale are shown across sleep duration categories in Figure 4B. Sleep duration category differences were highly statistically significant (F[4, 325] = 3.8; p = 0.005). There was a significant difference between those with less than 5 h mean sleep duration relative to the reference category of 7 to 8 h (p = 0.0001). After adjustment for multiple comparisons, a significant contrast remained between those with less than 5 h mean sleep duration and those with 7 to 8 h (p < 0.01), as was also observed on comparing those with less than 5 h and those with more than 8 h (p = 0.009).
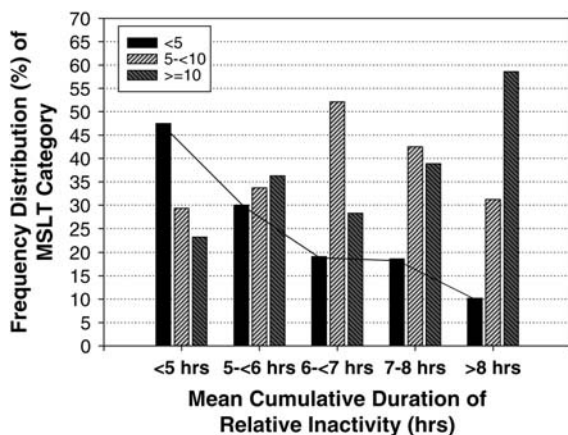

Figure 3. Percentages of subjects with MSLT values of less than 5 min, 5 to less than 10 min, and 10 min or more in different categories of cumulative duration of relative inactivity (hours). The *line* connects the percentages in each group with MSLT < 5 min.

DADT. Least-squares adjusted mean absolute DADT tracking errors back-transformed to the original scale are compared across sleep apnea severity categories in Figure 5A and among sleep duration categories in Figure 5B. Sleep apnea category differences were not statistically significant (F[3, 309] = 2.1; p = 0.10), although the planned contrast between those with severe apnea and no apnea was significant (p = 0.04). None of the pair-wise contrasts was statistically significant after adjustment for multiplicity (p > 0.05).

In contrast, sleep duration category differences were statistically significant (F[4, 309] = 2.4; p = 0.05). The planned contrast between less than 5 h relative to the 7- to 8-h reference group also was significant (p = 0.003). After multiplicity adjustment, the significance of this contrast remained (p = 0.03). No other pair-wise comparison was significant.

Investigation of Nonmonotonicity between Outcome Measures and Sleep Apnea Severity

There was a consistent nonmonotonicity in the associations between sleep apnea severity and the sleepiness and performance outcomes. We examined the following factors and found no evidence of confounding that could explain this nonmonotonicity: sleep duration, lifetime and current reports of specific medical conditions, medication use, age, BMI, systolic and diastolic blood pressure, and self-reported sleep quality. However, it is possible that confounding arose because of factors not assessed in this study. Additional details about these analyses are provided in the online supplement (*see* Tables E6 and E7).

Likelihood of Impairment

Objective criteria were used to dichotomize the outcomes into impaired or nonimpaired; for psychomotor vigilance lapses we used more than three lapses per 10-min trial and, for tracking error on the DADT, a mean absolute deviation greater than 250 cm per 30-min trial. These criteria were derived from previous studies evaluating impairments produced by alcohol use (23, 25). The impairment criterion for MSLT was the commonly used clinical threshold (< 5 min) (20). Results are summarized in Table 2. The weighted percentages with these impairments were 29.2, 36.9, and 25.7%, respectively. Only 5.3% of drivers had all three impairments. If impairment is defined as having at least one of these, then 59.5% of drivers would be considered impaired. The risk of multiple impairments was larger for those with severe sleep apnea. Among these drivers 60.7% had two or three impairments. Multiple impairments were also common among those drivers with the shortest sleep duration. The percentages of drivers with two or three impairments among those with less than 5 h was 49.5% compared with 24.3, 22.4, 17.2, and 10.2% in those drivers with 5 to less than 6, 6 to less than 7, 7 to 8, and more than 8 h, respectively.

DISCUSSION

In this study, we showed that both subjective and objective sleepiness, as well as performance impairments, are common in our sample of commercial driver's license holders. Our analyses reveal that chronic short sleep duration is a risk factor for subjective sleepiness, objectively measured sleepiness, and performance impairments. The results for sleep apnea are less clear. Although increases in sleep-disordered breathing were associated with reduced sleep latency, corresponding associations with subjective sleepiness and performance lapses were not confirmed. Increased tracking error was statistically significant only without covariate adjustment. In complementary analyses expressing risk factors in categories, severe sleep apnea relative to fewer than 5 events/h and short sleep durations of less than
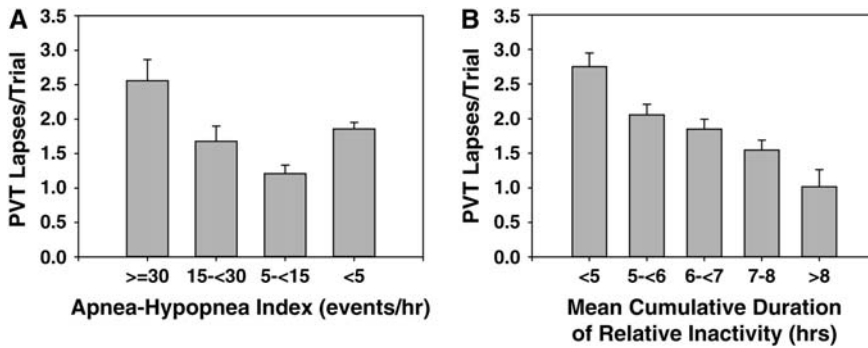
Figure 4. Predicted mean Psychomotor Vigilance Task (PVT) lapses per trial, controlling for apnea severity category, sleep duration category, age, sex, and body mass index. Statistical comparisons were performed on transformed data. Predicted values are reexpressed on the original scale to enhance interpretability. The magnitudes of the standard error bars are proportional to the ratio of the standard error to predicted value on the transformed scale to provide an accurate assessment of statistical precision. (*A*) Categories of severity of sleep apnea, as measured by apnea–hypopnea index. (*B*) Categories for sleep duration at home (mean cumulative duration of relative inactivity).

5 h relative to 7 to 8 h had similar magnitudes of effect for objective sleepiness as measured by the MSLT (average reductions of 2.5 and 2.3 min, respectively). Durations of sleep of less than 5 h were more common than severe sleep apnea (weighted percentages, 13.5 compared with 4.7%, respectively).

Our study design involved a random sample of persons with commercial driver's licenses, avoiding bias that might have occurred if drivers had been selected from only one company or had been selected on the basis of some other common characteristic. Moreover, the stratified sampling that we employed improved statistical precision of estimates, whereas sampling in clusters across several companies has the potential for greatly reducing statistical precision (32). A disadvantage of our approach was the expected low response rate, which strictly limits statistical inference to a "theoretical" population of drivers that would respond to surveys such as ours. However, the age, sex, and zip code distributions of respondents and nonrespondents were essentially identical, providing some evidence of minimal response bias. The respondents to our mail questionnaire knew that the study was about sleep but not the specific hypothesis being tested. Although the response rate was relatively low, response bias would have affected estimated relationships between risk factors and impairment outcomes only had there been differential recruitment of individuals with respect to one of the risk factors, that is, sleep apnea and short sleep duration and if the relationship between that risk and the impairment outcome differed between those responding and not responding. This seems unlikely. The issue of potential response bias means, however, that our estimates of the percentages of drivers with sleep apnea and short sleep durations need to be interpreted with caution. The percentage of subjects that we found with any sleep apnea (28.2%) is similar to that in other general population

studies (33–35) but less than that in other studies of commercial drivers (8, 9). This difference could be related to one or all of the following: (*1*) differences in response rate in different studies, with our study having the lowest response rate; (*2*) different sampling strategies—a population-based random sample in our study and a cluster-based sampling scheme across many companies in the study by Howard and coworkers (8); a nonrandom sample of drivers from a single company in the study by Stoohs and coworkers (9); (*3*) different definitions of respiratory events; and (*4*) different technologies used to estimate the degree of sleep-disordered breathing, with a simplified method (MESAM) used in the study by Stoohs and coworkers (9). The differences in estimated percentages with apnea are particularly large among studies in those with mild to moderate apnea, whereas the percentages with severe sleep apnea are more similar, that is, 4.7% in our study, 10.6% in the study by Howard and coworkers (8), and 10.0% in the study by Stoohs and coworkers (9).

In our study, we measured sleep durations at home by actigraphy and demonstrated associations between sleep duration and self-report sleepiness; physiologic sleepiness by MSLT, with vigilance lapses indicated by PVT; and tracking error by the DADT. We considered two potential metrics of sleep duration derived from actigraphy. The first was the cumulative duration of relative inactivity during the sleep bout, thereby removing periods of movement as part of sleep. The second was the overall duration from the beginning to the end of the main sleep bout and continuing to count periods of movement as sleep. To validate these measurements, we performed simultaneous analyses of actigraphy during full sleep studies with sleep duration being assessed by EEG in a large convenience sample of drivers being studied (n = 277); 69.2% of the higher risk group and 63.9% of the lower risk group. Validation analysis, as described in the online
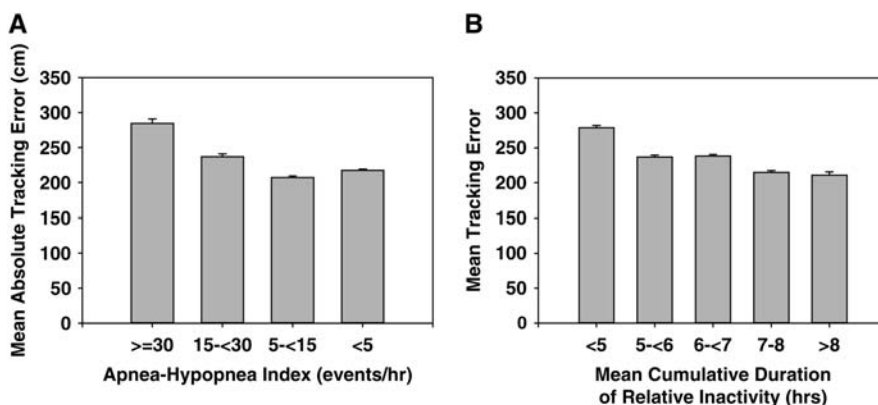

Figure 5. Predicted mean absolute tracking errors on the Divided Attention Driving Task controlling for apnea severity category, sleep duration category, age, sex, and body mass index, and expressed on the original scale (centimeters). Statistical comparisons were performed on transformed data. Predicted values are reexpressed on the original scale to enhance interpretability. Magnitudes of the standard error bars are proportional to the ratio of the standard error to predicted value on the transformed scale to provide an accurate assessment of statistical precision. (*A*) Categories of severity of sleep apnea as judged by apnea–hypopnea index. (*B*) Categories of sleep duration at home (mean cumulative duration of relative inactivity).

supplement, revealed that the duration of cumulative inactivity gave a closer approximation to actual sleep time (average difference, 0.31 h) than the total duration of the sleep bout, which overestimated sleep by 1.33 h, on average. However, this latter metric had a reduced standard deviation relative to the former, indicating a potential for improved agreement with the "gold standard" EEG in this population. We therefore repeated analyses, using this alternative metric after removing constant bias. The associations of sleep apnea with performance lapses and with tracking error become statistically significant (p = 0.05 and p = 0.005, respectively) and that with MSLT retained significance (p = 0.02) in the multivariable continuous risk factor models. We attribute the strengthening of these associations to better separation between estimates of sleep apnea severity and sleep duration, because activity due to termination of apneic events is eliminated from sleep duration estimates. These results are described in the online supplement (*see* Table E3).

In our sample, only 81.6% of the drivers studied were working as commercial drivers at the time of this study although all had the capacity to do so. Analyses were performed by restricting attention to currently employed drivers. These results are presented in Tables E4 and E5. The associations between outcomes and the primary measure of sleep duration were all somewhat stronger when restricting attention to drivers currently employed at the time of our study. However, the AHI linear slope for the MSLT was reduced in absolute value from –0.54 (p = 0.05, all drivers) to –0.44 (p = 0.16, currently employed drivers).

To identify important nonmonotonic associations, sleep apnea severity and average sleep duration at home were expressed as categorical variables in secondary multivariable models. Planned contrasts revealed significantly increased objective sleepiness (MSLT; p = 0.03) and tracking error (DADT; p = 0.04) for those with severe sleep apnea (AHI, at least 30 events/h) relative to no sleep apnea (< 5 events/h), but not for PVT performance lapses. Although sleep apnea severity category was significantly associated with performance lapses (p = 0.007), the relationship was nonmonotonic and greatly influenced by drivers with AHI values less than 5 events/h performing worse than those with values between 5 and 15 events/h. Although particularly marked for performance lapses, this nonmonotonic relationship with apnea severity was present for all four outcome variables. Therefore, we investigated potential causes of confounding to explain this. We examined differences in sleep duration, presence of medical conditions, and use of medications that could cause sleepiness and did not find any obvious reason. Thus, confounding due to group difference in medical status did not explain this nonmonotonicity. Although we cannot rule out confounding from other unobserved factors, random sampling variation is as likely an explanation as is confounding for these results. Details are provided in the online supplement (*see* Tables E6 and E7).

Although sleep duration was found not to be a confounding factor for sleep apnea severity, there remains the potential for sleep duration to serve as an effect modifier, that is, the effect of increasing sleep apnea severity may result in different levels of impairment depending on mean sleep duration. We tested for interaction but found no statistically significant interaction in any of the multivariable models. However, the statistical power for detection of an interaction effect between sleep duration and apnea severity is low because of the low power of interaction tests generally, and specifically because of the relatively small sample sizes in some of the cells in the cross-tabulation of sleep duration and apnea severity categories.

Another factor that needs to be considered in interpretation of our results is obesity. Obesity has been shown to be associated with subjective and objective sleepiness independent of sleep apnea (36). In our study, there were no statistically significant associations between outcomes and BMI. But because only 11.3% of our sample had a BMI less than 25 kg/m^2 our study is limited in its ability to detect associations between obesity and performance.

Other aspects of this study are also worthy of comment. We used thermistors to assess airflow. This study started before results conclusively showing the superiority of nasal pressure measurements to assess flow had been presented (37). In our study, performance tests were done on the day after the overnight sleep study in the laboratory. Because the duration of sleep during the overnight polysomnography was not limited it is possible that some subjects obtained more sleep than was their usual habit. Consequently, our findings on objective sleepiness and performance in the laboratory will likely underestimate the effect of reduced sleep duration.

We found that a high proportion of subjects (32.6%) had an Epworth Sleepiness Scale score in the range compatible with self-report excessive sleepiness. Thus, in our sample, self-report sleepiness was extremely common. It is arguable that this may be the result of response bias, that is, that individuals with self-report sleepiness were more likely to respond to our survey. But, as noted, the proportion of subjects in our study with sleep apnea is less than in other studies of commercial drivers (8, 9). Moreover, self-report excessive sleepiness was also extremely common and similar (29.0%) in the only other study of commercial drivers to assess this (8). We found in multivariable modeling that when assessed as a continuous variable, reductions in sleep duration did increase the expected level of self-reported sleepiness. A 1-SD decrease in sleep duration (1.34 h) was associated with an expected increase of 0.75 points (p < 0.01) on the Epworth Sleepiness Scale. However, sleep apnea was not associated with self-report sleepiness. This is unlike the results of Howard and coworkers (8), who found a weak association between Epworth Sleepiness Scale scores and sleep apnea severity in a sample of commercial drivers. The basis for this lack of association in our study is unknown but appears to limit the usefulness in commercial drivers of self-report sleepiness in determining who is likely to have sleep apnea. Whereas sleep apnea was not associated with self-report sleepiness, it was associated with objective sleepiness as measured by the MSLT in the primary continuous risk factor model (p = 0.05). In our categorical model it was found that this association was particularly evident among individuals with severe apnea relative to those no apnea (reduction in mean equal to 2.5) and also relative to those with mild apnea (5 to 15 events/h; reduction in mean, 3.0 min). (Previous studies of commercial drivers [8, 9] have not assessed objective tests of sleepiness or performance.) Simultaneously, average sleep duration at home was associated with objective sleepiness, with sleep durations less than 5 h associated with the most severe sleepiness (reduction in mean equal to 2.3 min). Thus, multivariable modeling revealed that severe sleep apnea and sleep duration at home of less than 5 h had similar effects on objective sleepiness as determined by MSLT. However, the percentage of subjects in this sample with durations less than 5 h was nearly triple that of subjects with severe sleep apnea.

Our data are consistent with studies showing that performance is differentially susceptible to the effects of sleep deprivation (38, 39). The performance of some subjects is quite impacted by sleep deprivation whereas other subjects are relatively resistant (38, 39). We see evidence to support similar differential susceptibility to the effects of both chronic short sleep duration at home and the degree of sleep apnea (*see* Figures 2 and 3). However, it is possible

that measurement error regarding apnea severity and mean sleep duration has contributed to our observations.

When we examined definitions of impairment for PVT performance lapses and DADT tracking error based on data comparing results with those produced by alcohol intoxication, we found that 29.2% (for lapses) and 36.9% (for mean tracking error) had performance decrements comparable to that induced in control subjects, albeit in different populations, after alcohol intoxication. Moreover, 25.7% had a mean sleep latency in the range (< 5 min) considered pathologic. This is an issue of concern. However, we found considerable discordance among these objective impairment outcomes. The proportion of individuals with such performance impairments in our society is unknown, and there are no data that currently show an association between impairments in the tests we performed and crash risk in commercial drivers. Also, there are currently limited data on the association between, for example, the presence of sleep apnea and crash rates among commercial drivers. In the study by Stoohs and coworkers (40), no association was found between sleep apnea severity and crash risk, as was the case for the group described in the study by Howard and coworkers (8), who performed full sleep studies. The latter result is likely to be related to the small sample size (n = 161) because, in the larger questionnaire sample (n = 2,343) in this study, the multivariable apnea prediction index, a surrogate for apnea (14), was weakly related to the risk of a single vehicle accident but not to the total crash history (8). This topic has been more extensively studied, however, in drivers of passenger cars, where a clear association has been shown between crashes and obstructive sleep apnea (41–48). There is a need for more in-depth studies of crashes involving commercial drivers and the role of sleep apnea and insufficient sleep. In particular, there is a need for a case-control study focusing on the role of sleep apnea, and potentially sleep duration, in major crashes of commercial vehicles.

In conclusion, there are daytime neurobehavioral performance impairments that are found commonly in commercial drivers, and these are more likely among those with durations of average sleep less than 5 h/night and those with severe obstructive sleep apnea. These results suggest that strategies employed by the Federal Motor Carrier Safety Administration to reduce sleepiness, and potentially crash risk, in commercial drivers include plans to (*1*) develop and implement approaches to identify "impaired" drivers by objective testing, (*2*) implement and ensure quality programs to identify and treat individuals with severe sleep apnea as well as monitor adherence to therapy, and (*3*) introduce approaches to assess and promote increased sleep durations among commercial drivers.

References

1. Center for National Trucking Statistics. THA facts 1999. Ann Arbor, MI: University of Michigan Transportation Research Institute; 2000.
2. National Transportation Safety Board. Factors that affect fatigue in heavy truck accidents. Vol. 1: Analysis safety study. Washington, DC: National Transportation Safety Board; 1995.
3. AAA Foundation for Traffic Safety. A report on the determination and evaluation of the role of fatigue in heavy truck accidents. Washington, DC: AAA Foundation for Traffic Safety; 1995.
4. Dinges DF, Pack F, Williams K, Gillen KA, Powell JW, Ott GE, Aptowicz C, Pack AI. 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.
5. Belenky G, Wesensten NJ, Thorne DR, Thomas ML, Sing HC, Redmond DP, Russo MB, Balkin TJ. Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose–response study. *J Sleep Res* 2003;12:1–12.
6. Van Dongen HP, Maislin G, Mullington JM, Dinges DF. The cumulative cost of additional wakefulness: dose–response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* 2003;26:117–126.
7. Mitler MM, Miller JC, Lipsitz JJ, Walsh JK, Wylie CD. The sleep of long-haul truck drivers. *N Engl J Med* 1997;337:755–761.
8. Howard ME, Desai AV, Grunstein RR, Hukins C, Armstrong JG, Joffe D, Swann P, Campbell DA, Pierce RJ. Sleepiness, sleep-disordered breathing, and accident risk factors in commercial vehicle drivers. *Am J Respir Crit Care Med* 2004;170:1014–1021.
9. Stoohs RA, Bingham LA, Itoi A, Guilleminault C, Dement WC. Sleep and sleep-disordered breathing in commercial long-haul truck drivers. *Chest* 1995;107:1275–1282.
10. Pack AI, Maislin G, Staley B, George C, Pack FM, Dinges DF. Factors associated with daytime sleepiness and performance in a sample of commercial drivers [abstract]. *Sleep* 2001;24:A427.
11. Maislin G, Hachadoorian R, Dinges DF, Pack AI. Apnea and sleep behavior co-determine sleepiness and alertness in commercial truck drivers [abstract]. *Sleep* 2002;25:A119.
12. Gurubhagavatula I, Maislin G, Nkwuo JE, Pack AI. Occupational screening for obstructive sleep apnea in commercial drivers. *Am J Respir Crit Care Med* 2004;170:371–376.
13. Gislason T, Almqvist M, Eriksson G, Taube A, Boman G. Prevalence of sleep apnea syndrome among Swedish men: an epidemiological study. *J Clin Epidemiol* 1988;41:571–576.
14. Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, Schwab RJ, Dinges DF. A survey screen for prediction of apnea. *Sleep* 1995;18:158–166.
15. Johns MW. A new method for measuring daytime sleepiness: the Epworth Sleepiness Scale. *Sleep* 1991;14:540–545.
16. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;15:376–381.
17. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea: the Epworth Sleepiness Scale. *Chest* 1993;103:30–36.
18. Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17:703–710.
19. Gottlieb DJ, Whitney CW, Bonekat WH, Iber C, James GD, Lebowitz M, Nieto FJ, Rosenberg CE. Relation of sleepiness to respiratory disturbance index: the Sleep Heart Health Study. *Am J Respir Crit Care Med* 1999;159:502–507.
20. Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. *Sleep* 1986;9:519–524.
21. Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav Res Methods Instrum Comput* 1985;17:652–655.
22. Doran SM, Van Dongen HP, Dinges DF. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch Ital Biol* 2001;139:253–267.
23. Powell NB, Riley RW, Schechtman KB, Blumen MB, Dinges DF, Guilleminault C. A comparative model: reaction time performance in sleep-disordered breathing versus alcohol-impaired controls. *Laryngoscope* 1999;109:1648–1654.
24. Moskowitz H, Burns M. The effects of alcohol and valium, singly and in combination, upon driving-related skills performance. In: Huelke DF, editor. Proceedings of the 21st Conference of the American Association for Automotive Medicine. Morton Grove, IL: American Association for Automotive Medicine; 1997. pp. 226–240.

25. George CF, Boudreau AC, Smiley A. Simulated driving performance in patients with obstructive sleep apnea. *Am J Respir Crit Care Med* 1996;154:175–181.

26. Middelkoop HA, Knuistingh Neven A, van Hilten JJ, Ruwhof CW, Kamphuisen HA. Wrist actigraphic assessment of sleep in 116 community based subjects suspected of obstructive sleep apnoea syndrome. *Thorax* 1995;50:284–289.

27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–310.

28. American Academy of Sleep Medicine. Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research: the Report of an American Academy of Sleep Medicine Task Force. *Sleep* 1999;22:667–689.

29. Cochran WG. Survey sampling. New York: John Wiley & Sons; 1977.

30. Kripke DF, Garfinkel L, Wingard DL, Klauber MR, Marler MR. Mortality associated with sleep duration and insomnia. *Arch Gen Psychiatry* 2002;59:131–136.

31. Edwards D, Berry JJ. The efficiency of simulation-based multiple comparisons. *Biometrics* 1987;43:913–928.

32. Kish L. Survey sampling. New York: John Wiley & Sons; 1965.

33. Young T, Palta M, Dempsey J, Skatrud J, Weber S, Badr S. The occurrence of sleep-disordered breathing among middle-aged adults. *N Engl J Med* 1993;328:1230–1235.

34. Bearpark H, Elliott L, Grunstein R, Cullen S, Schneider H, Althaus W, Sullivan C. Snoring and sleep apnea: a population study in Australian men. *Am J Respir Crit Care Med* 1995;151:1459–1465.

35. Bixler EO, Vgontzas AN, Ten Have T, Tyson K, Kales A. Effects of age on sleep apnea in men. I. Prevalence and severity. *Am J Respir Crit Care Med* 1998;157:144–148.

36. Vgontzas AN, Bixler EO, Tan TL, Kantner D, Martin LF, Kales A. Obesity without sleep apnea is associated with daytime sleepiness. *Arch Intern Med* 1998;158:1333–1337.

37. Heitman SJ, Atkar RS, Hajduk EA, Wanner RA, Flemons WW. Valida-tion of nasal pressure for the identification of apneas/hypopneas during sleep. *Am J Respir Crit Care Med* 2002;166:386–391.

38. Van Dongen HP, Baynard MD, Maislin G, Dinges DF. Systematic inter-individual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep* 2004;27:423–433.

39. Leproult R, Colecchia EF, Berardi AM, Stickgold R, Kosslyn SM, Van Cauter E. Individual differences in subjective and objective alertness during sleep deprivation are stable and unrelated. *Am J Physiol Regul Integr Comp Physiol* 2003;284:R280–R290.

40. Stoohs RA, Guilleminault C, Itoi A, Dement WC. Traffic accidents in commercial long-haul truck drivers: the influence of sleep-disordered breathing and obesity. *Sleep* 1994;17:619–623.

41. George CF, Nickerson PW, Hanly PJ, Millar TW, Kryger MH. Sleep apnoea patients have more automobile accidents. *Lancet* 1987;2:447.

42. Findley LJ, Unverzagt ME, Suratt PM. Automobile accidents involving patients with obstructive sleep apnea. *Am Rev Respir Dis* 1988;138:337–340.

43. Young T, Blustein J, Finn L, Palta M. Sleep-disordered breathing and motor vehicle accidents in a population-based sample of employed adults. *Sleep* 1997;20:608–613.

44. George CF, Smiley A. Sleep apnea and automobile crashes. *Sleep* 1999;22:790–795.

45. Teran-Santos J, Jimenez-Gomez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. *N Engl J Med* 1999;340:847–851.

46. Horstmann S, Hess CW, Bassetti C, Gugger M, Mathis J. Sleepiness-related accidents in sleep apnea patients. *Sleep* 2000;23:383–389.

47. Connor J, Whitlock G, Norton R, Jackson R. The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. *Accid Anal Prev* 2001;33:31–41.

48. Sassani A, Findley LJ, Kryger M, Goldlust E, George C, Davidson TM. Reducing motor-vehicle collisions, costs, and fatalities by treating obstructive sleep apnea syndrome. *Sleep* 2004;27:453–458.