A Study of Prevalence of Sleep Apnea Among Commercial Truck Drivers

DEPARTMENT OF TRANSPORTATION
UNITED STATES OF AMERICA

FMCSA
Federal Motor Carrier Safety Administration

May 2002

Technical Report Documentation Page

1. Report No. FMCSA-RT-02-030	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle A Study of Prevalence of Sleep Apnea Among Commercial Truck Drivers		5. Report Date 12-2-1	
		6. Performing Organization Code	
7. Author(s) Allan I. Pack, M.B., Ch.B., PhD.;David Dinges, PhD.;Greg Maislin, M.S., M.A		8. Performing Organization Report No.	
9. Performing Organization Name and Address Trucking Research Institute American Trucking Associations Foundation 2200 Mill Road Alexandria, VA 22314		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. DTFH61-93-R-00088	
12. Sponsoring Agency Name and Address Federal Motor Carrier Safety Administration Office of Research and Technology 400 Virginia Avenue, SW, Suite 600 Washington, DC 20024		13. Type of Report and Period Covered Final Report December 1993-December 2001	
		14. Sponsoring Agency Code	

15. Supplementary Notes

The Contracting Officer's Technical Representative was Albert Alvarez, FMCSA Office of Research and Technology. This study was performed by the University of Pennsylvania Center for Sleep & Respiratory Neurobiology, in cooperation with the Trucking Research Institute of the American Trucking Associations Foundation.

16. Abstract

The purpose of this study is: (1) to estimate the prevalence of sleep apnea among a high-risk sample of commercial drivers; (2) to examine the relationship in the high-risk sample between severity of sleep apnea and decrements in function related to driving tasks, and; (3) to develop a profile of the overall sample of commercial drivers with regard to their sleep apnea-related characteristics and risks.

This study, involving overnight laboratory testing of more than 400 subjects, was one of the largest sleep apnea studies ever conducted on any population. It revealed that sleep apnea is common among commercial drivers, but not as prevalent as indicated in some previous reports. Mild sleep apnea occurs in 17.6% of holders of commercial drivers licenses; moderate sleep apnea in 5.8%; and severe sleep apnea in 4.7%. Sleep apnea prevalence increases with age and with degree of obesity as measured by body mass index. Also, shorter average durations of nightly sleep are associated with higher prevalence of apnea. Subjects with sleep apnea and short duration sleep did not report higher levels of daytime sleepiness, but they did show deficits on all objective tests of alertness performance used in this study. Decrements in performance and excessive sleepiness were found in individuals with severe sleep apnea (apnea/hypopnea index >30 episodes/hour) and those sleeping less than 6 hours/night. In those tests affected by both factors, the effect on performance measures of these abnormalities is similar in magnitude.

Overall, this study revealed sleep apnea to be a major concern for commercial motor vehicle driver health and alertness. Accordingly, FMCSA is planning programs to educate industry and drivers on sleep apnea risks.

17. Key Words Commercial Drivers; Sleep Apnea; Daytime Sleepiness; Driving Tasks; Altertness Performance; Apnea/Hypopnea Index; Performance Measures; Determinants of Prevalence of Sleep Apnea; Sleep Duration.		18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 290	22. Price

Form DOT F 1700.7 (8-72)   Reproduction of completed page authorized

This form was electronically produced by Elite Federal Forms, Inc.

FINAL REPORT

A Study of Prevalence of Sleep Apnea Among Commercial Truck Drivers

DTFH61-93-R-00088

Principal Investigator

Allan I. Pack, M.B., Ch.B., Ph.D.
Professor of Medicine and Director
Center for Sleep and Respiratory Neurobiology
Chief, Division of Sleep Medicine/Department of Medicine
University of Pennsylvania
Philadelphia, Pennsylvania

Co-Principal Investigators

David F. Dinges, Ph.D.
Professor of Psychology in Psychiatry
Chief, Division of Sleep and Chronobiology
Department of Psychiatry
University of Pennsylvania
Philadelphia, Pennsylvania

Greg Maislin, M.S., M.A.
Adjunct Assistant Professor of Biostatistics in Medicine
Division of Sleep Medicine/Department of Medicine
Principal Biostatistician, Center for Sleep and Respiratory Neurobiology
University of Pennsylvania
Philadelphia, Pennsylvania

December 2, 2001

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This is the largest, most comprehensive study of prevalence of sleep apnea, sleep duration at home and the resulting effect on daytime performance that has ever been performed. It was done in a highly relevant population that is challenging to study—commercial drivers.

While our data lead to several interesting results, we summarize here what we consider to be the most important.

A major goal of our study was to determine the prevalence of sleep apnea in commercial drivers. We chose as our sampling frame a random sample of holders of commercial drivers licenses (CDLs) living in Pennsylvania within 50 miles of the University of Pennsylvania. Using this sample, we found that 17.6% of holders of CDLs (95% confidence interval (CI): 7.9-27.3%) had mild sleep apnea, i.e., apnea/hypopnea index (AHI) ≥ 5 and <15 episodes/hour; 5.8% (95% CI: 0.0-16.1%) had moderate sleep apnea (AHI≥ 15 and <30 episodes/hour); while 4.7% (95% CI: 0.0-9.6%) had severe sleep apnea (AHI≥ 30 episodes/hour).

These prevalence rates are similar to those reported in other prevalence studies of sleep apnea in general populations [Bearpark, 1995; Bixler et al, 1998; Young et al, 1993] but not close to the extremely high prevalence values reported previously for commercial drivers by Stoohs et al [1995].

As in other populations, we report that prevalence depends on age and degree of obesity. The prevalence increases with age and with increasing obesity. There is a multiplicative relationship between these factors, such that the effects of increasing obesity on apnea prevalence are magnified as age increases. Moreover, the effects of age are more pronounced at higher levels of obesity. These relationships allow us to provide the commercial driving industry with prediction equations useful for estimating prevalence of sleep apnea in any population of drivers provided only with the age and body mass index values of the drivers in the target population.

The prevalence of sleep apnea depends on the likelihood of sleep apnea as measured by the multivariable apnea prediction (MAP) [Maislin et al, 1995]. This simple tool is based on age, gender, body mass index (a measure of obesity) and report of the frequency of symptoms of sleep apnea. It gives a relative likelihood of apnea between zero and one. In our population of CDL holders, MAP values below 0.47 provide a high negative predictive value (99.4%) for AHI≥ 30 events/hours. Therefore, MAP values less than 0.47 can be used to virtually exclude severe sleep apnea. Severe sleep apnea is of primary concern as the results of our daytime performance testing indicate. This MAP value (less than 0.47) is found in 57.2% in our overall population of CDL holders.

An important result from our study, and one never previously assessed or reported in any epidemiological study of prevalence of sleep apnea, is that sleep apnea prevalence depends on average duration of sleep as measured over consecutive nights at home. Shorter sleep durations

are associated with increasing prevalence of sleep apnea. This might be the result of what has been called cumulative partial sleep deprivation (repeated too short sleep durations) exacerbating the degree of sleep-disordered breathing. Alternatively, the presence of sleep apnea might lead to individuals only being able to sleep for shorter durations. Our data do not allow us to differentiate between these possibilities.

Short sleep duration is itself an issue of concern. Measurement of sleep duration at home by measuring absence of wrist movements as a surrogate of sleep is problematic in individuals with sleep apnea. This is because such individuals move at the end of the apneic episodes even though they are still asleep. To deal with this issue, we defined two different measures of sleep duration. The first is the bout length of the major period of inactivity. This will likely overestimate sleep since individuals could be awake during this interval. The second is the cumulative duration of inactivity during the major sleep bout. This will likely underestimate sleep, particularly in those with sleep apnea who can be moving even while asleep. We found that functional impairments were associated with durations of the major bout of inactivity of <6 hours. This occurred in 9.8% of subjects. Likewise for the other variable—cumulative duration of inactivity—performance impairments were associated with duration of this of <5 hours. This occurred in 13.5% of subjects. Thus, both approaches give similar estimates of the prevalence of chronic partial sleep deprivation, i.e., subjects who on average have too short durations of sleep. We also found that both methods for estimating sleep duration in our study led to the conclusion that sleep duration was affected by the time at which individuals terminated their sleep. A surprisingly high percentage (35.6%) of holders of CDLs terminate sleep before 6:00 am. Such individuals had significantly shorter sleep durations. Thus, in considering causes of impairment in commercial drivers, there are issues both related to a proportion of drivers not getting enough sleep and to some drivers, particularly those who are most obese, having severe sleep apnea.

Sleep apnea and shorter sleep durations would be expected to result in excessive sleepiness during the day. We found that this was so for objectively measured sleepiness but there was no relationship between sleep apnea severity, sleep duration and degree of self-reported sleepiness. A large proportion of drivers did have self-reported sleepiness as assessed by standard instruments. For example, 32.6% of our sample of CDL holders had an Epworth Sleepiness Score (a self-report measure of sleepiness) in the sleepiness range (above 10). This range is associated with a three- to four-fold increased risk of fall-asleep crashes in drivers of passenger cars [Stutts et al, 1999]. While a substantial proportion of drivers reported sleepiness, the presence and severity of sleep apnea was not a determinant of the degree of self-reported sleepiness. The reason for this unexpected lack of relationship is unclear, but it means that we cannot rely on self-reports of sleepiness to identify drivers likely to have sleep apnea.

Objective tests did, however, reveal that the presence and degree of sleep apnea was a determinant of objectively measured sleepiness as well as of performance in tests sensitive to the effects of sleep loss. We found this result for all tests we performed: (a) multiple sleep latency test, a measure of the physiological pressure for sleep; (b) psychomotor vigilance reaction time test, measuring reaction time to stimuli as well as unwanted lapses in performance; (c) tracking errors in the divided attention task, a task designed to simulate the cognitive load of driving. While these are separate tests, the results of our various analyses were remarkably similar. In general, we find that severe sleep apnea (AHI≥30 episodes/hour) is associated with marked

excessive sleepiness and resulting performance decrements. Severe sleep apnea occurs in 4.6% of CDL holders. The relationship between sleep durations as measured at home and the various tests of performance was more complex. This is partly related to the difficulty in actually measuring sleep duration in individuals with sleep apnea (as discussed above). We found for almost all tests an association with the cumulative duration of inactivity, as described above. Individuals with a cumulative duration of inactivity of less than 5 hours showed an impact on performance on almost all variables of performance that we assessed. In contrast, for the duration of the main bout of relative inactivity, we did not find an association for all of the measures we used. We did find such an association for some of our key primary end-points, i.e., multiple sleep latency (a measure of physiological pressure for sleep); performance lapses on the psychomotor vigilance reaction time task (PVT); and decrement in performance on lane tracking on divided attention task. For these various measures, changes were found in individuals with an average duration of relative inactivity in the major sleep bout of <6 hours. This occurred in 9.8% of individuals. For these specific tests the impact of severe sleep apnea (AHI>30 episodes/hour) and short duration of the major bout of relative inactivity (<6 hours) were approximately equivalent.

Since both sleep apnea and shortened sleep were prevalent in our population of CDL holders, the latter being somewhat more common, any effort to deal with impaired performance, as a result of excessive sleepiness, needs to address both issues. These issues are, moreover, interrelated since prevalence of severe apnea is larger in CDL holders with shorter sleep durations. This interaction produces complexity in defining a precise level of sleep apnea, at which impairment occurs. The effects of sleep apnea on daytime performance depend not only on severity of sleep apnea, but also the average sleep duration of the subject. Our various tests of performance such as, reaction time and degree of sleepiness, etc., allow us to conclude that on average individuals with severe sleep apnea, i.e., apnea/hypopnea index ≥30 episodes/hour, show evidence of impairment. The challenge for the future will be developing cost-effective ways to identify such drivers and ensuring correction of the abnormality.

These results lead us to the following recommendations:

1. There is a need for educational programs for all components of the commercial driving industry about sleep apnea and chronic cumulative partial sleep deprivation.

2. Research is needed to assess the role of sleep apnea in crash causation for commercial vehicles.

3. Research is needed as to the efficacy of different strategies to identify commercial drivers with severe sleep apnea and/or chronic cumulative partial sleep deprivation (<6 hours/night), and to assess whether once identified in drivers these abnormalities can be reversed.

4. There is a need for a conference of experts from different backgrounds with different expertise to review the data from this report and to develop recommendations as to public policy in this area. The report contains specific policy questions that need to be addressed.

CHAPTER ONE

Summary of Study for Non-Scientists

1.1 Background to Study

Obstructive sleep apnea is a common condition. In its symptomatic form, i.e., with complaints of excessive sleepiness, it affects 4% of middle-aged males and 2% of middle-aged females [Young et al, 1993]. These estimates are derived from a study of state employees in Wisconsin. It is a condition that results from repeated closure of the upper airway, i.e., at the base of the tongue and/or behind soft palate, and/or repeated narrowing with decrements in breathing. Complete closures are called apneas (cessation of breathing) while decrements are called hypopneas. These events occur during sleep because of loss of tone of the muscles surrounding the airway which keep the airway open during wakefulness. Severity of disease is judged by counting the total number of apneas and hypopneas which occur during sleep. The average number, which occur per hour during sleep, is called the Respiratory Disturbance Index (RDI) or apnea/hypopnea index (AHI). (A list of abbreviations used throughout this report and their definitions is given in Appendix A.) An apnea/hypopnea index of less than 5 episodes/hour is considered normal; 5-15 episodes/hour is said to be mild sleep apnea; 15-30 episodes/hour, moderate sleep apnea; and 30 or greater episodes/hour, severe sleep apnea [American Academy of Sleep Medicine, 1999].

There are a number of important risk factors for sleep apnea [Young et al, 1993, Bearpark et al, 1995; Bixler et al, 1998]. It is more common in men than women. Its prevalence (i.e., how common it is in the population being studied) increases with age and also with degree of obesity (how overweight individuals are). There are also certain data that indicate that sleep apnea is more common in African Americans than in Caucasians [Kripke et al, 1997; Redline et al, 1997b]. (For summary of previous epidemiological studies in this area, see Appendix B.)

With these episodes of cessation or marked reduction in breathing, the oxygen level in the blood falls and the carbon dioxide level rises. These changes are sensed by the brain and the subject is awakened, possibly to complete wakefulness, but usually to a lighter stage of sleep (this sudden "awakening" is called an arousal). With awakening, the upper airway muscles are activated, the airway is opened, and breathing restarts. These repetitive arousals (awakenings) interrupt sleep and hence the quality of sleep in subjects with sleep apnea is reduced and sleep is not as refreshing. They are, moreover, excessively sleepy during the day; hence, they are at risk to fall asleep inappropriately and have impaired performance on tasks such as driving that require sustained vigilance and attention.

While sleep apnea is a risk factor for sleepiness, not all individuals with sleep apnea are excessively sleepy. This is acknowledged and the concept of the sleep apnea/hypopnea syndrome has been introduced [American Academy of Sleep Medicine, 1999]. The syndrome is said to be present when the individual has not only breathing abnormalities during sleep, but is also complaining of excessive sleepiness during the day. Currently the syndrome definition is based on self-reported sleepiness and not objectively measured sleepiness. This is an issue since

not all individuals who are found to be excessively sleepy on testing realize it and complain of it. As it transpires, this is a particular issue for commercial drivers in whom our study finds there is no relationship between presence and severity of sleep apnea and degree of self-reported sleepiness. It seems likely in the future that the syndrome definition will be reconsidered.

Sleep apnea is not only a risk factor for sleepiness, but there is also growing evidence that it is a risk factor for cardiovascular disease. Recent data from the Sleep Heart Health Study [Nieto et al, 2000] show that sleep apnea is an independent risk factor for hypertension, i.e., beyond the effects of obesity on both hypertension and apnea. When the apnea/hypopnea index is in the severe range, i.e., ≥30 episodes/hour, sleep apnea is an independent factor for hypertension. Moreover, hypertension arises more frequently over time in individuals with sleep apnea that is untreated [Peppard et al, 2000]. The major likely mechanism for hypertension is the recurrent falls in oxygen. There are some data that sleep apnea is also a risk factor for heart attack [Hung et al, 1990] and stroke [Dyken et al, 1996; Wessendorf et al, 2000; Bassetti and Aldrich, 1999] but the data are less complete than for hypertension.

Since excessive sleepiness can be a consequence of apnea, subjects with sleep apnea can have degraded performance. On tests such as the Divided Attention Task, which was designed to mimic the cognitive load of driving [Moskowitz and Burns, 1977], subjects with sleep apnea perform, on average, as poorly as individuals over the legal limit of blood alcohol concentration [George et al, 1996].

As a result, subjects with sleep apnea have an increased risk of vehicular crashes. This has been shown in different types of studies: (a) in patients attending sleep centers for evaluation and treatment [e.g., George and Smiley, 1999]; (b) in subjects in the community who were found in a research study to have sleep apnea and who were followed while untreated over time [Young et al, 1997b]; and (c) in subjects who had a major crash on rural highways where alcohol was not implicated [Teran-Santos et al, 1999]. Almost all of these studies show that individuals with sleep apnea have of the order of a 3-7 times increased risk of crashes. The studies have been criticized on methodological grounds in a recent review on the topic [Connor et al, 2000]. Nevertheless, the evidence does point to an increased risk of crashes in subjects with sleep apnea. There is some controversy about the level of severity of sleep apnea at which this increased risk occurs. Some studies show an increased risk of crashes in individuals with severe sleep apnea, while others find an increased risk even in those with mild disease. This topic is discussed more fully in Chapter Two where those studies are reviewed in more depth. A summary of studies conducted to date on the topic is given in Appendix C.

Given these data, and that commercial drivers as a group are more obese [Rather et al, 1981] (a major risk factor for sleep apnea), it is not surprising that the question of sleep apnea in commercial drivers was raised [Stoohs et al, 1994, 1995]. A study of the prevalence of sleep apnea was conducted in a single trucking company in a somewhat non-standard way, i.e., simplified sleep studies done at the company hub [Stoohs et al, 1995]. The prevalence of apnea was large. Seventy-eight percent of commercial drivers had an apnea/hypopnea index of >5 episodes/hour, i.e., at least mild apnea. This is to be compared with other studies on males: 24.0% in state employees in Wisconsin [Young et al, 1993]; 25.9% in Busselton, Australia [Bearpark et al, 1995]; 15.9% in Pennsylvania [Bixler et al, 1998]. Crashes of these commercial

drivers were also assessed and it was stated that drivers with sleep apnea had double the crash risk [Stoohs et al, 1994]. This difference was, however, <u>not statistically significant</u> since the increased crashes in drivers with apnea was primarily the result of crashes of two drivers. It is unfortunate that this non-significant result is still quoted and is the basis of recommendations for policy [McNicholas, 1999].

While this study [i.e., Stoohs et al, 1994] on commercial drivers did not find a statistically significant increase in crash rates in individuals with sleep apnea, it nevertheless drew national attention to this important issue. The issue is important not only because sleep apnea may be common in commercial drivers, but also since it is reversible with treatment. There are effective treatments available. The most effective treatment is use of a mask that is put over the nose during sleep and is connected to a positive pressure source. This positive pressure keeps the airway open and prevents apneas and other breathing abnormalities from occurring. This treatment is called nasal continuous positive airway pressure (CPAP) [Sullivan et al, 1981]. The system is highly portable and could be used by commercial drivers while employed. It is a safe treatment with few major side effects [Pack, 1994]. There are a number of minor side effects such as runny nose which can be troublesome to the patient. While the treatment is safe, the major problem is long-term adherence to therapy given the cumbersome nature of it [Kribbs et al, 1993a]. Alternative treatments are available such as use of a device, a dental appliance that is placed in the mouth over the teeth during sleep to move the jaw forward [for reviews see Schmidt-Nowara, 1995; Ayas and Epstein, 1998]. These devices are not as efficacious as CPAP. Surgical treatments are also available but success rates of the commonly used surgery (of the order of 50%) limit their application [Sher et al, 1996].

All of this information led to this study of sleep apnea in commercial drivers to address its prevalence, risk factors and its impact on performance measures relevant to the driving task. The objectives of the study are given in the next section.

1.2 <u>Objectives of the Study</u>

Our study had three objectives:

a. To estimate the prevalence of sleep apnea among a sample of commercial drivers within the United States. Our sample was based on a random sample of holders of commercial drivers licenses who lived in Pennsylvania within a 50 mile radius of Philadelphia.

b. To examine the relationship in commercial drivers between severity of sleep apnea and decrements in function related to driving tasks.

c. To develop a profile of our overall sample of commercial drivers with regard to their sleep apnea-related characteristics and risks.

1.3 Study Design and Subjects Studied

A. Overall Study Design

Our study design was based on that proposed by Gislason et al [1988] and used in the seminal study on sleep apnea prevalence by Young et al [1993]. This is a two-stage design. In the first stage the population of interest is defined. Then the population is surveyed to obtain information relevant to whether they are likely or not to have sleep apnea. Based on responses to this survey, two samples are formed: (a) a sample at higher risk for apnea and (b) a sample at lower risk for apnea. A larger percent of subjects in the higher risk sample are selected for in-depth laboratory testing to see whether they have sleep apnea as is a smaller percentage of the lower risk group. This design enriches the sample in terms of number of subjects with apnea that are available for assessment of functional consequences while allowing a population-based estimate of prevalence.

B. Establishing the Sampling Frame

For our study we chose as our sampling frame the population of holders of commercial drivers licenses (CDLs) in Pennsylvania who lived within 50 miles of our sleep center at the University of Pennsylvania. This approach has the advantage that the sampling frame is precisely defined, permitting a true random sample of drivers in a precisely defined population. This random sample was implemented by the Department of Motor Vehicles in the State of Pennsylvania who gave us, from their data bases, a random sample of 4826 holders of CDLs in this geographical area.

The alternative approach we considered was to base the sample on a company or companies. For an individual company there are concerns about generalizeability since the hiring practices of that company may have certain specific features affecting the nature of their driver population. For a sample based on companies, a fixed number of drivers from each company would be recruited for study from companies selected at random, with probability proportional to their number of employed drivers. Thus, this approach would require, in advance of the randomized selection process, a precise delineation of the sampling frame comprising of all possible companies that employ eligible drivers along with their numbers of such drivers. It would be difficult to establish the proper sampling frame unless one pre-selects certain characteristics of the company, e.g., they belong to a particular national organization from which a list of companies could be obtained. Moreover, if a selected company refused to participate, this would have a much more deleterious effect on the question of generalizeability than if an individual driver refused to participate.

For these reasons, we chose the sampling frame of holders of CDLs. This strategy has two potential problems. First, not all holders of CDLs will be working as commercial drivers. Indeed, this was the case. But they do have the capability to do so and there is a large amount of job turnover in this industry [ATA Foundation, 1997]. Moreover, we were able to give estimates of sleep apnea prevalence both for individuals working full-time as commercial drivers as well as those who were no longer employed in this capacity. The second issue with the strategy is the lower response rate in the initial survey phase of the study. This was evaluated in pilot studies

and we found that even with realistic financial incentives, a response rate of the order of 30% was obtained. In the study itself we obtained a 31.5% response rate. But such response rates are to be anticipated in studies of subjects with demographic characteristics similar to those in our population of CDL holders. Moreover, we found that the demographic characteristics of the respondents and non-respondents and their specific geographic location were virtually identical. 92.2% of respondents were male as were 92.1% of non-respondents. The average age of male respondents was 44.6 years and of non-respondents 44.5 years. The age distribution of the two groups was virtually identical. Thus, there is no evidence of bias in respondents. This issue of response rate is discussed more fully in Chapter Three. Moreover, we provide in our estimate of prevalence of sleep apnea predictions of prevalence of sleep apnea of different severities based on the major risk factors—age and body mass index. This will permit our data on sleep apnea prevalence to be extrapolated to other populations of commercial drivers if distribution of age and body mass index are known. Finally, we note that the issue of response bias does not affect our second objective, i.e., to determine the relationship between respiratory index during sleep and performance decrements during wakefulness.

C. <u>The Nature of our Initial Sample</u>

For our initial survey, we obtained responses from 1391 holders of commercial drivers licenses. This is shown schematically in Figure 1.1. 7.2% of the sample was female. The mean age of the total sample, i.e., including males and females, was 44.5 years. 21.6% of the respondents were under the age of 35 years. The percentage of drivers who were female (7.2%) is close to that reported by the U.S. Department of Labor for commercial drivers (5.7%). Likewise, the percentage of our respondents who were under 35 years of age (21.6%) is similar to that reported by the Gallup organization (24.6%). While these characteristics are, therefore, similar to that in other surveys of commercial drivers, the number of drivers in our sample who were over 50 years (30.8%) is considerably higher than the 1% of drivers in an American Trucking Associations' (ATA) survey. The reason for this difference is unknown. These issues are discussed more fully in Chapter Three.

D. <u>Risk Stratification in Higher and Lower Risk Group</u>

To determine how likely an individual driver was to have sleep apnea, we used the multivariable apnea prediction (MAP) described by Maislin et al [1995] (see copy of full paper in Appendix D). This instrument allows calculation of the relative likelihood of apnea between zero and one, based on age, gender, body mass index (BMI) (a measure of degree of obesity) and responses to questions about symptoms of sleep apnea such as loud snoring. We rank ordered the responses to questions on our survey, which allowed calculation of MAP for each driver from highest to lowest. The top 551 scores were all above 0.4356 and we defined this group of respondents as the higher risk group. For this group, we enrolled 44.8% for our in-laboratory studies (n=249). The remaining 778 respondents, i.e., with MAP scores below 0.4356, constituted the lower risk group and we enrolled 20.4% in a random order for in-laboratory studies (n=159). Thus, the overall design of our study is that shown in Figure 1.1.
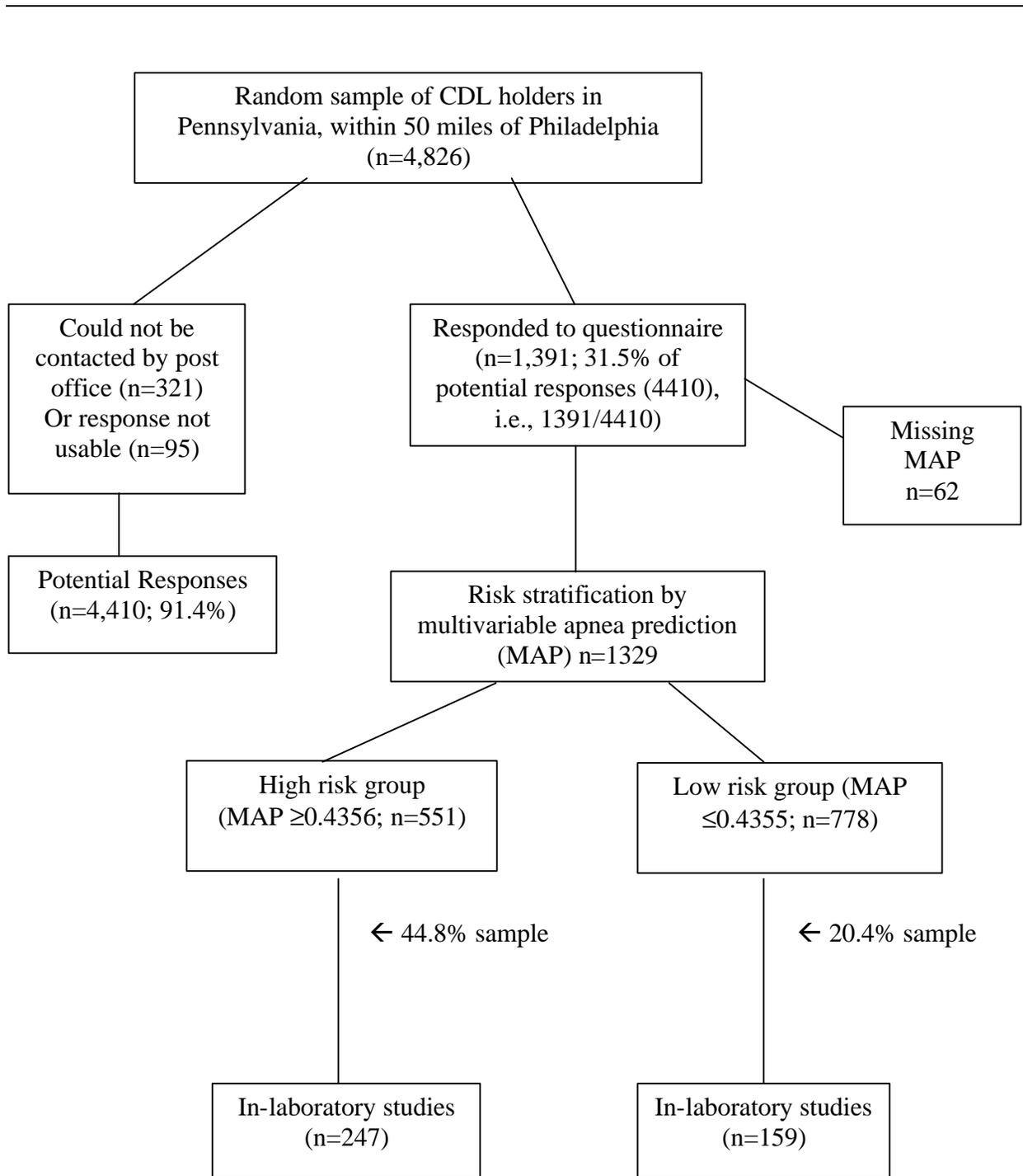
Figure 1.1. Overall sample, respondents and subject disposition status leading to in-lab studies of n=249 higher risk and n=159 lower risk holders of Pennsylvania commercial drivers licenses.

E. Methods for In-Laboratory Testing

All drivers in the in-laboratory phase of our study had the following studies:

1. Actigraphy at Home

The drivers wore a small device like a wrist-watch that measures wrist motion. It measures movements over one minute intervals and hence gives reasonably precise estimates of sleep and wakefulness. The drivers wore this small device for seven consecutive days and nights. The device, which is a wrist activity monitor, stores and records the number of wrist movements in the device's memory for later computer analysis. From examination of the characteristic wrist movement patterns the amounts of activity and/or sleep can be inferred [Redmond and Hegge, 1985]. There are problems applying this to individuals who have sleep apnea. They move as a result of the breathing abnormalities and it can be difficult, therefore, to assess whether they are awake or asleep. This is discussed more fully in Chapter Four.

2. Sleep Study

The drivers underwent an overnight sleep study at the Clinical Research Center for Sleep at the University of Pennsylvania. Brain activity, heart rate, oxygen level and breathing were recorded. This night-time sleep testing allowed us to measure the apnea/hypopnea index. The drivers stayed during the next day in the Sleep Research Center for tests of daytime performance, including evaluating how sleepy they were. These tests are described briefly below.

3. Measurements of Self-Reported Sleepiness

The drivers completed questionnaires to assess the degree of self-reported sleepiness. First, they completed the Epworth Sleepiness Scale [Johns, 1991, 1992, 1993, 1994] (see Appendix J). This self-rating scale asks individuals to rate how likely they are to doze on a four-point scale (0, 1, 2 and 3) in eight situations, e.g., being a passenger in a car, watching TV, etc. The questionnaire thus gives a total score of between 0 and 24. A score above 10 is considered excessive sleepiness. Individuals with scores in this range have a three- to four-fold increased risk of having a car crash due to falling asleep [Stutts et al, 1999].

The second questionnaire that they completed was the Functional Outcomes of Sleep Questionnaire (FOSQ) [Weaver et al, 1997]. This questionnaire was specifically designed to evaluate the effect of sleepiness on relevant aspects of daily living. Both the Epworth Sleepiness Score and the FOSQ thus evaluate self-reported sleepiness that has been present over some period of time before the subject is evaluated.

In addition, during the day of testing in the Sleep Research Center, subjects filled out questionnaires to determine how sleepy they were at that moment. This was done on four occasions during the day of testing in our laboratory. This day of testing followed the overnight sleep study. The specific instruments that we used were the Karolinska Sleepiness Scale

[Akerstedt and Gillberg, 1990] and the Stanford Sleepiness Scale [Hoddes et al, 1973]. These instruments are included in Appendix J.

During the day in our laboratory, subjects also had a number of tests of performance and measurement of degree of sleepiness. Specifically, we used the following:

1. Multiple Sleep Latency Test (MSLT) [Carskadon and Dement, 1982; Carskadon et al, 1986]. This measures how quickly subjects fall asleep on daytime naps and hence the pressure they have for sleep. Subjects are asked to fall asleep in a dark room at 2-hour intervals. The time to fall asleep is measured and the average across all nap opportunities calculated. This average is called the multiple sleep latency and is a measure of the physiological pressure for sleep.

2. Psychomotor vigilance test (PVT), a reaction time task [Dinges and Powell, 1985; Dinges, 1992]. This simple test measures the reaction to stimuli presented at random intervals over a ten-minute period. It is based on a simple handheld device with a small visual display. For this test a number of different measures are obtained: median reaction time; number of performance lapses (i.e., an absent response); domain of lapse duration, which refers to shifts in lapse duration calculated from the 10% lowest reaction times (a metric that reflects vigilance response slowing); optimum response times; the fatigability function that is the slope of the slowing in response over the period of testing. All of these metrics from the PVT are sensitive to the effects of sleep loss [Dinges, 1992].

3. Divided Attention Driving Task (DADT) [Moskowitz and Burns, 1997; George et al, 1996]. This test was designed to simulate the cognitive load of driving tasks. It uses a driving-like paradigm in which subjects must keep a cursor within the middle of a randomly moving target using a steering wheel device (lane tracking), while at the same time identifying and responding to numbers that appear at irregular intervals at the corners of the screen. The version we employed took 30 minutes/test. We computed the average deviation from the desired center point for each 2-minute period, giving 15 measurements/test. We computed the average of these and the slope of increase in deviations over the time of the test.

All of these measurements were repeated four times during the day of testing in our laboratory. For analysis purposes, the average of the four trials was computed and used in statistical models. Each of the four testing battery of measurements required of the order of 60 minutes: 20-25 minutes for multiple sleep latency test; 10-12 minutes for Psychomotor Vigilance Reaction Time Test; and 30-35 minutes for Divided Attention Driving Task.

1.4 Sleep Duration in Our Sample of CDL Holders

Depending on the method used to assess sleep duration, the results of the actigraphy showed that the average duration of the major sleep bout was between 6 and 1/3 hours/night to 7.4 hours/night. Thus, we do not find, on average, the extremely short sleep durations reported by Mitler et al [1997]. There were, however, substantial differences in the amount of sleep between individuals ranging from less than 4 hours to just over 9 hours per night. Those subjects with the shortest sleep durations also had the most variable durations of sleep across nights. We

found that 9.8% of our sample slept for less than 6 hours/night. This was associated with decrements in performance while awake during our laboratory testing. This duration of sleep has been shown to be associated with an increased risk of fall-asleep crashes in drivers of passenger cars [Stutts et al, 1999]. Thus, these short sleep durations are an issue of concern.

There were a number of factors identified that led to short sleep durations. First, a large percentage of our sample terminated sleep early in the morning. 12.3% did so before 5:00 am and 35.6% before 6:00 am. Those who terminated sleep at these early morning hours had shorter sleep durations. Early termination of sleep was not the only factor that was related to shorter sleep durations. Males slept shorter than females by an average of 46 minutes. The presence of sleep apnea also affected sleep durations. Those with moderate or severe sleep apnea had significantly shorter sleep than those without apnea. Whether this is related to the concept that sleep deprivation can make sleep apnea worse [Stoohs and Dement, 1993] or that the sleep apnea with frequent awakenings leads to short sleep durations cannot be discerned from these data. It does mean that individuals with severe sleep apnea can have two factors leading to impaired performance, i.e., sleep fragmentation as a consequence of sleep apnea and repeated short durations of sleep resulting in cumulative partial sleep deprivation [Dinges et al, 1997].

1.5. <u>Prevalence of Sleep Apnea</u>

We found that 17.6% (95% CI: 7.9-27.3%) had mild sleep apnea (AHI≥5 and <15 episodes/hour); 5.8% (95% CI: 0.0-16.1%) had moderate sleep apnea (AHI≥15 and <30 episodes/hour); 4.7% (95% CI: 0.0-9.6%) had severe sleep apnea (AHI≥30 episodes/hour). These prevalence estimates are similar to those in other population studies and considerably less than those reported by Stoohs et al [1995] in a previous study of commercial drivers.

As in previous studies of prevalence of sleep apnea, the prevalence was strongly dependent on age and degree of obesity as measured by the body mass index. Being over 60 years old gives an increased relative risk for the presence of sleep apnea of 3.1, while for a BMI above 35 (class 2 and 3 obesity) the increased relative risk is 30.1. Our results show an interaction between the risk factors of age and obesity, that is the increases in risk are multiplicative with respect to these two variables. The effect of age on prevalence of sleep apnea is more marked in those who are obese, while the effect of obesity on prevalence is more marked in older individuals. These relationships have allowed us to develop prediction equations that could be used for any population of drivers to compute the prevalence of sleep apnea of different severities in that population, if the distribution of age and body mass index of the drivers are known.

Not surprisingly, the prevalence of sleep apnea was related to the magnitude of the multivariable apnea prediction (MAP). The prevalence of any form of sleep apnea and for severe sleep apnea increased as the MAP value increased. This relationship may have some value in practice. In particular, the instrument was found to have high negative predictive value, i.e., be able to rule out the presence of sleep apnea. No driver with a MAP under 0.215 had any sleep apnea; this represents 16.3% of the population we studied. No subject with a MAP value less than 0.323 (34.9% of the population) had severe sleep apnea. Moreover, if we use a cut-point of 0.47 (57.2% of the population) the negative predictive value for severe sleep apnea is

close to perfect, i.e., 99.4%. Severe sleep apnea was found to be the issue of importance since in these holders of CDLs, we found that this was associated with performance impairment as measured by the Psychomotor Vigilance Reaction Time Test (PVT) and Divided Attention Driving Task (DADT).

Our study also revealed that sleep duration was a determinant of apnea prevalence with those sleeping shorter durations having a higher prevalence of apnea.

Certain of the CDL holders in our sample were no longer working as full-time commercial drivers. We wondered whether some of them may have ceased employment because they had experienced difficulty driving as a result of sleep apnea. In support of this postulate, we found that the prevalence of moderate and severe sleep apnea is nearly twice as high in those no longer currently employed as commercial drivers, as compared to those employed full-time as commercial drivers. For those no longer employed as commercial drivers, 17.2% had mild sleep apnea; 9.3% had moderate sleep apnea; and 7.3% had severe sleep apnea. In contrast, for those currently employed as full-time commercial drivers, 16.9% had mild sleep apnea, 4.7% had moderate sleep apnea, and 4.3% had severe sleep apnea.

Although these prevalence estimates are considerably lower than those reported by Stoohs et al [1995], they nevertheless are a source of concern if approximately one in twenty-five commercial drivers have severe sleep apnea.

1.6 Self-Reported Sleepiness

A. Epworth Sleepiness Scale

As discussed above, this scale is a self-report measure based on stated likelihood to doze in eight typical situations. A score above 10 is said to be excessively sleepy. 32.6% of our sample had scores in this range. There was, however, surprisingly no evidence of a relationship between the Epworth Sleepiness Score and the presence or severity of sleep apnea.

While sleep apnea was not a determinant of the Epworth Sleepiness Score, the presence of frequent self-reported snoring (≥3 times/week) was. Since frequent snoring will also be present in a form of sleep-disordered breathing called the Upper Airway Resistance Syndrome [Guilleminault et al, 1993], this might explain the role of snoring in determining sleepiness. Alternatively, since both the Epworth Sleepiness Score and the presence of snoring are both self-report measures, this relationship might simply reflect that some individuals are more likely to report problems and complaints (report bias). In support of this, we did not find that the presence of frequent snoring was related to objective measures of sleepiness.

B. Functional Outcomes of Sleep Questionnaire (FOSQ)

The FOSQ results also showed no relationship to sleep apnea severity and also no relationship to average sleep duration. The presence of frequent snoring was, however, a determinant of the total score for this questionnaire (for discussion see above).

C. Karolinska and Stanford Sleepiness Scales

The scores for neither of these scales, which assess degree of sleepiness at a moment in time, showed a relationship to average sleep duration nor to severity of sleep apnea.

D. Summary

While the various measures did not generally show the anticipated relationships with sleep apnea severity and/or average sleep duration, the results for the different subjective rating scales were correlated providing evidence of reliability of our measures. The correlation between the two measures of sleepiness over a period, i.e., the Epworth and FOSQ, was 0.40. The correlation between the measure of degree of sleepiness at a moment in time—the Karolinska and Stanford Sleepiness Scales—was even higher, 0.89.

The basis for the unexpected lack of relationship between apnea severity or average duration and self-report sleepiness is unclear. It is not because self-report sleepiness does not occur. A high proportion of drivers do have self-reported sleepiness. It does mean, however, that one cannot rely on self-report measures of sleepiness to identify drivers at risk for severe sleep apnea.

1.7 Objective Measurements of Sleepiness and Performance

Unlike self-report measures of sleepiness, our objective measures did reveal relationships with sleep apnea severity and for some of our measures average sleep duration.

A. Multiple Sleep Latency Test (MSLT)

The average time to fall asleep on four daytime naps is measured. Values less than 5 minutes are considered indicative of pathological excessive sleepiness and values above 10 minutes without sleepiness. Values between 5 and 10 minutes are said to be in a gray zone. In our sample, the mean value across all subjects of average MSLT across all four nap opportunities was 8.30 minutes. 25.6% of our sample had values of MSLT less than 5 minutes, i.e., levels of sleepiness that are of concern, and 43.1% had values in the gray zone.

The MSLT was related to apnea severity (p=0.0005). The estimated values for MSLT in the different categories were as follows: no apnea, 8.42 minutes; mild sleep apnea, 9.00 minutes; moderate sleep apnea, 7.24 minutes; and severe sleep apnea, 4.98 minutes. Thus, particularly in the group with severe sleep apnea (AHI≥30 episodes/hour) there was evidence of excessive sleepiness. In this group more than 50% of the subjects had MSLT values less than 5.0 minutes.

MSLT values were also related to average sleep duration in the week prior to testing. Thus, these drivers had cumulative partial sleep deprivation [Dinges et al, 1997]. Individuals with less than 6 hours of sleep/night had significantly shorter MSLT values as compared to any longer duration of sleep.

To assess the relative effect of these different factors, we performed multiple linear regression analyses. We found that controlling for other variables the effect of sleep apnea severity remained significant as did the effect of average sleep duration. These variables had effects of approximately the same magnitude on MSLT. Severe sleep apnea reduced MSLT by an average 3.00 minutes, as compared to no apnea; a sleep duration of less than 6 hours reduced MSLT by 2.95 minutes as compared to that with an average sleep duration of >8 hours/night. Explanatory models revealed that the effects of these different variables on impairment on the MSLT, i.e., a value less than 5 minutes, is complex. The probability of impairment increases with increasing severity of sleep apnea and with reductions in durations of sleep. Moreover, the probability of impairment is determined by both factors such that those who have the highest probability of impairment have both severe sleep apnea and short average sleep durations.

B. Psychomotor Vigilance Task

From this task we extracted a number of parameters—median response time; number of performance lapses; duration of lapse domain; optimum response times; fatigability function, i.e., reaction time slowing while a task.

Analysis of the data led to relatively consistent observations and conclusions.

1. Median Response Time

The median response time was related to severity of sleep apnea. It was significantly larger in those with severe sleep apnea and in those with moderate sleep apnea. The average across subjects of the median responses for those with severe sleep apnea was 289 msec compared to 214 msec in those with no apnea. For comparison purposes, trials of 35 male subjects given alcohol such that their breath alcohol was 0.084 (over the legal limit in California, where the study was done), increased reaction times from an average of 236 msec to 271 msec. Thus, the median reaction time for drivers with severe sleep apnea is similar to those who are over the legal limit for breath alcohol.

A multiple linear regression model showed that controlling for other variables the effects of sleep apnea severity remained significant (p=0.019) but that for average sleep duration did not. Explanatory models again revealed the complex interaction between these factors in determining probability of impairment.

2. Performance Lapses

The number of performance lapses was significantly related to the severity of sleep apnea (p=0.0009) with the number of lapses in those with severe sleep apnea (AHI≥30 episodes/hour) being significantly greater than that in all other categories. (87.4% higher as compared to those with no apnea.) The effect size for sleep apnea was found to be very large. The number of lapses was also related to average sleep duration. The effect size for average sleep duration was again found to be of substantive magnitude, and highly significant differences in lapses was found in those with average sleep durations of less than 6 hours compared to those with all other categories of longer sleep durations. Multiple regression analysis revealed that the

estimated effects of severe sleep apnea as compared to no sleep apnea (an increase in lapses of 2.3 lapses/10 minute interval) was similar in magnitude to the effect of short sleep duration of less than 6 hours compared to that for >8 hours of sleep/night (increase of 1.51 lapses/trial).

The analysis of the other metrics obtained from the PVT confirmed several of these findings. In particular, sleep apnea severity was associated with worse performance and was particularly found in those with severe sleep apnea.

C. Divided Attention Driving Task (DADT)

As described above, this task is designed to evaluate lane tracking ability. The DADT uses a driving-like paradigm in which subjects must keep a cursor within the middle of a randomly moving target using a steering wheel device (lane tracking), while at the same time identifying and responding to numbers that appear at irregular intervals at the corners of the screen. Our version of the test took 30 minutes. Average deviation from the desired center point is computed for each 2-minute period. The major outcome measure is the average (over 15 intervals) two-minute sum of the absolute values of the tracking errors. In addition, we computed a *tracking decrement function* using multiple linear regression that measured the rate at which performance became worse during the 30 minute testing bout. Thus, we extracted two metrics from Divided Attention Driving Tasks: (a) the mean tracking error and (b) the decrement in tracking performance over the time on the task.

1. Mean Tracking Error

We found that mean tracking error is related to the severity of sleep apnea (p=0.002). Subjects with severe sleep apnea had a mean tracking error of 380.0 cm compared to 251.6 cm in those with no sleep apnea (p=0.008). The effect of differences in sleep duration on this particular measure was not as clear.

2. Tracking Decrement Function

This measurement also was significantly associated with increasing severity of sleep apnea and, in this case, shorter sleep durations. For this measurement, the effect of increasing severity of sleep apnea was particularly marked. The decrement in lane tracking performance in those with severe sleep apnea (39.2) was about 10 times as fast as those with mild sleep apnea (4.1) (p=0.002). Multiple linear regression analysis revealed that the magnitude of the effect of sleep apnea on this measure was considerably greater than for shorter sleep durations. This is an important observation since maintenance of lane tracking ability is so critical to the driving task.

Thus, these studies give largely consistent results. For key measures of wake performance, they show that both increasing severity of sleep apnea and shorter durations of sleep are associated with excessive daytime sleepiness and performance impairments. These effects are found for individuals with severe sleep apnea and for those with average sleep durations of less than 6 hours. For several of our performance assessments, the magnitude of the effects of severe sleep apnea and sleep duration less than 6 hours on performance are similar.

For maintenance of lane tracking, however, larger effects are found with severe sleep apnea. The effects of these two variables—severe sleep apnea and shorter sleep duration—are multiplicative such that the probability of performance impairment is determined by both the severity of sleep apnea and the average duration of sleep. Those with the highest probability of impairment have both severe sleep apnea and short sleep duration.

1.8 Conclusions and Recommendations

Based on these results, we draw a number of conclusions and make a number of recommendations. These are described in Chapter Eight of the report.

CHAPTER TWO

Background and Study Objectives

Obstructive sleep apnea is a common disorder that in its symptomatic form is estimated to affect 4% of middle-aged males and 2% of middle-aged females [Young et al, 1993]. This estimate comes from a study of state employees in Wisconsin. It is a condition where during sleep there is reduction of the activity of the throat muscles surrounding the upper airway. These muscles keep the upper airway open. Hence, reduction in their activity leads to airway narrowing and, in some cases, closure (apnea).

A. Definition of Abnormal Breathing Events During Sleep

A number of breathing abnormalities during sleep can result from this process (see Table 2.1). The definitions of these breathing abnormalities are taken from a recent consensus statement by the American Academy of Sleep Medicine [1999]. First, breathing may cease due to airway closure. Such events are called *obstructive apneas*. (This distinguishes them from the much less common *central apneas* where breathing stops due to cessation of activity of the diaphragm [the major muscle for breathing].) As a result of an apnea, the oxygen level in the blood declines and the carbon dioxide level rises. This leads to the subject "waking up" at least to a lighter stage of sleep. With this awakening, called an "arousal", activity of the upper airway muscles is quickly restored and breathing restarts. As the subject returns to sleep, the airway again closes and the process thus occurs repeatedly during sleep.

The same effects can result even if breathing does not stop but merely declines. This is called a *hypopnea* (for definition, see Table 2.1) [Gould et al, 1988]. More recently, it has been suggested that sleep can also be interrupted if subjects are working hard to breathe due to upper airway narrowing, even if obvious decreases in breathing are not observed. This type of event has recently been named a *respiratory effort-related arousal* (RERA) (see Table 2.1). (For summary of all abbreviations used throughout this report, see Appendix A.) It has been proposed that in some patients this is the predominant breathing abnormality or event and the term "upper airway resistance syndrome" has been introduced to describe such patients [Guilleminault et al, 1993]. The presence of this new type of syndrome is not universally accepted [Douglas, 2000]. Moreover, identifying such abnormalities requires invasive monitoring and insertion of a tube through the nose to the gullet to measure breathing effort by assessing pressure changes in the esophagus [Guilleminault et al, 1993].

Sleep-disordered breathing (a general term to describe all types of abnormalities) is a continuum of abnormality (see Figure 2.1). Thus, the condition is not simply either present or absent. The continuum progresses through the following stages: a) normal; b) snoring being present but without effect on sleep quality; c) upper airway resistance syndrome [Guilleminault et al, 1993]; d) sleep hypopnea syndrome (i.e., with consequences being present but no apneas) [Gould et al, 1988]; and e) sleep apnea syndrome. Since apneas and hypopneas have the same effects, the condition is now frequently called the "sleep apnea/hypopnea syndrome". There is some evidence that over time the disorder progresses, albeit slowly, and the subject moves up the

continuum of abnormality [Bliwise et al, 1984; Phoha et al, 1990; Svanborg and Larsson, 1993; Sforza et al, 1994; Pendlebury et al, 1997; Linberg et al, 1999].

Table 2.1. Definition of different abnormalities during sleep from a report of the American Academy of Sleep Medicine [1999].

Type of Abnormality (Events)	**Definition**
Obstructive apnea/hypopnea*	The events must fulfill criterion 1 or 2, plus criterion 3 of the following: 1. A clear decrease (>50%) from baseline in the amplitude of a valid measure of breathing during sleep. Baseline is defined as the mean amplitude of stable breathing and oxygenation in the two minutes preceding onset of the event (in individuals who have a stable breathing pattern during sleep), or the mean amplitude of the three largest breaths in the two minutes preceding onset of the event (in individuals without a stable breathing pattern). 2. A clear amplitude reduction of a validated measure of breathing during sleep that does not reach the above criterion but is associated with either an oxygen desaturation of >3% or an arousal. 3. The event lasts 10 seconds or longer.
Respiratory effort-related arousal (RERA)	These events must fulfill both of the following criteria: 1. Pattern of progressively more negative esophageal pressure, terminated by a sudden change in pressure to a less negative level and an arousal. 2. The event lasts 10 seconds or longer.

*The report indicates that there is no need to separate apneas from hypopneas, since they have the same clinical significance.

Normal	Snoring with no effect on sleep quality	Upper airway resistance syndrome	Sleep hypopnea syndrome	Sleep apnea syndrome

| Not clinical significant | Questionable | Definite clinical significance |

Progression of disorder over time ▶
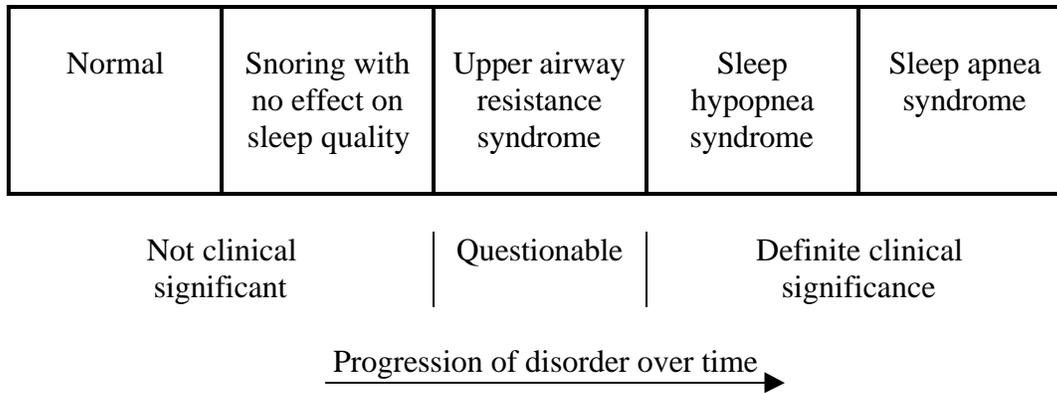
Figure 2.1.   Continuum of abnormality in sleep-disordered breathing.   For more detail on definitions of events and syndrome, the interested reader is referred to reference American Academy of Sleep Medicine [1999].

B. __What is Normal? Values for the Apnea/Hypopnea Index__

Given that there is a continuum of abnormality, it raises the question of what is normal, and what is abnormal. This has been addressed in a number of studies. Currently it is proposed that the major measure of abnormality is the average number of apneas plus hypopneas per hour of sleep. This is called the *apnea/hypopnea index (AHI)*. It is alternatively called the *respiratory disturbance index (RDI)*, which is identical to the apnea/hypopnea index.

Originally it was proposed that the threshold of abnormality was an apnea/hypopnea index above 15 episodes/hour [Gould et al, 1988]. This was based on studies in a small sample of normal individuals of different ages, none of whom had an apnea/hypopnea index above 15 episodes/hour.

More recently, it has been proposed that an apnea/hypopnea index of 5 episodes/hour should be the threshold of abnormality [American Academy of Sleep Medicine, 1999]. This proposal is based on three lines of evidence:

a. In some studies there is an increased risk of hypertension in subjects with an AHI>5 episodes/hour [Young et al, 1997a].

b. In some studies [Young et al, 1997b; Terán-Santos et al, 1999], but not all [George and Smiley, 1999], there is an increased risk of motor vehicle crashes even in subjects with an AHI>5 episodes/hour. (The issue of relationship between crashes and sleep apnea is discussed more fully below.)

c. Some benefit is found with treatment in patients presenting to sleep centers who have an apnea/hypopnea index between 5 and 15 episodes/hour [Engelman et al, 1997; Engelman et al, 1999].

These data, and others, led a committee of the American Academy of Sleep Medicine to propose that obstructive sleep apnea should be considered to have three levels of severity (see Table 2.2).

Table 2.2. Proposed definitions of severity of sleep apnea based on number of apneas and hypopneas/hour of sleep [from American Academy of Sleep Medicine, 1999].

Severity Rating	Apnea/Hypopnea Index (Episodes/hour)
Mild	5<AHI≤15
Moderate	15<AHI≤30
Severe	AHI >30

C. Concept of Sleep Apnea/Hypopnea Syndrome

We discuss here the concept of a syndrome, i.e., not just where subjects have breathing abnormalities during sleep, but where there are clinical consequences of such abnormalities. The syndrome is said to be present when both breathing abnormalities during sleep are found and there are clinical consequences. Apneas or hypopneas during sleep lead to transient reductions in oxygen and increases in carbon dioxide in the blood. As a result, the subject wakes up, usually not to complete wakefulness, but to a lighter stage of sleep. With this awakening, typically called an "arousal", the upper airway is suddenly opened as a result of the action of the muscles surrounding the airway and normal breathing is resumed. As breathing resumes, the subject returns to sleep and the process repeats. Thus, apneas and/or hypopneas are repetitive throughout sleep.

It is believed that the repeated falls in oxygen are the major determinant of cardiovascular consequences. There is growing evidence of the role of sleep apnea as an independent risk factor for high blood pressure (systemic hypertension) [Nieto et al, 2000; Peppard et al, 2000]. The Sleep Heart Health Study, in which over 6000 subjects were studied, shows that sleep apnea is an independent risk factor for high blood pressure when the apnea/hypopnea index is greater than 30 episodes/hour [Nieto et al, 2000]. There is also some evidence that sleep apnea is likely to be a risk factor for heart attack (myocardial infarction) [Hung et al, 1990] and stroke [Dyken et al, 1996; Basseti and Aldrich, 1999; Wessendorf et al, 2000], but the data for this are not as yet conclusive.

The repeated interruption of sleep in sleep apnea, also called "sleep fragmentation," leads to poor quality sleep and hence subjects with this disorder are excessively sleepy during the day. Excessive daytime sleepiness leads to performance impairment and, as discussed below, an increased risk of crashes. Degraded performance in patients with sleep apnea has been shown on tests well known to the traffic safety community such as the Divided Attention Task [Moskowitz and Burns, 1997]. On this test patients with sleep apnea perform as poorly as drivers who are over the legal limit for blood alcohol [George et al, 1996] (see Figure 2.2). Their performance on the Divided Attention Task improves with treatment for sleep apnea [George et al, 1997].
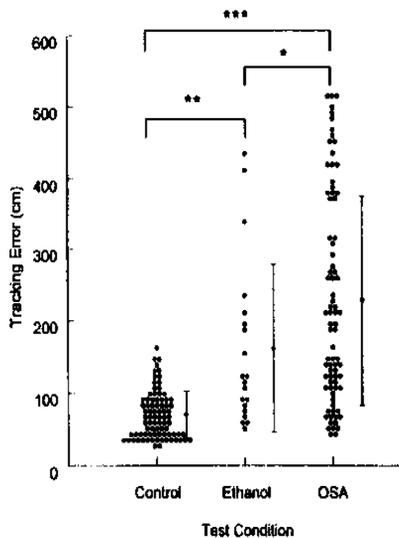
Figure 2.2. Tracking errors measured in Divided Attention Task in controls (left panel), drivers over the legal limit for ethanol (middle panel) and sleep apnea patients (OSA) (right panel). *p<0.05, **p<0.001, ***p<1x10^{-9}. From George et al [1996].

There are, however, important differences between individuals in how much effect disturbed breathing during sleep has on them. Thus, some subjects with relatively low levels of apnea/hypopnea index may be quite sleepy during the day, whereas some subjects with high levels of respiratory disturbance index during sleep may not be excessively sleepy.

This has led to the introduction of the concept of the *sleep apnea/hypopnea syndrome*. In brief, the disorder is only said to be present when there are both apneas and/or hypopneas during sleep and complaints of excessive daytime sleepiness. This acknowledges that the breathing disturbance during sleep may be present without consequences such as excessive daytime sleepiness, and that the presence of the laboratory-identified breathing abnormality during sleep is not sufficient to define the presence of a disorder. While this is an important concept, which will be discussed throughout this report, currently experts in sleep apnea have not made clear how to operationalize the definition of excessive sleepiness, i.e., how will this be measured? This is a void in our current knowledge base and one that is particularly challenging when considering the issue of sleep apnea in populations such as commercial drivers. We took an approach based on the concept of convergent validity, i.e., assessing sleepiness by multiple tools and instruments. Both self-reports of sleepiness (subjective sleepiness) and performance changes related to sleepiness (objective sleepiness) were assessed by multiple methodologies.

D. How Common is Sleep Apnea? Prevalence of Sleep Apnea and Sleep Apnea/Hypopnea Syndrome

This question has been addressed in a large number of studies (see Appendix B for summary of published literature in this area). The prevalence depends on whether one is simply describing the number of people who have different levels of respiratory disturbance during sleep, as outlined above in Table 2.2, or who have the sleep apnea/hypopnea syndrome (i.e., who have complaints of excessive daytime sleepiness). The importance of this distinction is seen from the results of the large study done in a random sample of Wisconsin state employees by Young et al [1993]. While 24% of males and 9% of females had an apnea/hypopnea index above

5 episodes/hour, only 4% of males and 2% of females had the symptoms that allowed the investigators to classify subjects as having sleep apnea/hypopnea syndrome [Young et al, 1993].

Although a large number of investigations have been done (see Appendix B), many of these studies were small and/or did not involve subjects getting full sleep tests to identify the breathing abnormality during sleep. Thus, out of the studies performed to date, there are essentially three that have been done in large, randomly selected populations and in whom some form of sleep study was done in the subjects investigated. These three studies were done in the following: (1) a random sample of middle-aged state employees in Wisconsin, i.e., between 30 and 60 years of age, by Young et al [1993]; (2) a random sample of males aged 40 to 65 years of age in the town of Busselton in Western Australia by Bearpark et al [1995]; and (3) a random sample of males aged 20 years and older in two counties of Pennsylvania (Dauphne and Lebanon) by Bixler et al [1998]. The main results of these three studies are shown in Table 2.3. There is substantial concordance between the results of these different prevalence studies.

Table 2.3. Prevalence of respiratory disturbance during sleep and sleep apnea/hypopnea syndrome (right column) in different large population studies. It is these prevalence estimates that the results of the present study need to be compared to in assessing whether commercial drivers have a higher risk for sleep apnea. (We do not discuss here the important prevalence studies of Ancoli-Israel et al [1991] in the elderly, since these are not germane to our investigation.)

Name/Citation of Study	Prevalence of AHI ≥5/Hour (Males)	Prevalence of AHI ≥15/Hour or 20/Hour (Males)	Prevalence of AHI >5/Hour and Symptoms of Sleepiness (Males)
Young et al [1993]* (n=602; 353M)	24.0%	9.1%[a]	4.0%
Bearpark et al [1995] (n=486; all male); 294 studied in lab	25.9%	3.4%[b]	3.1%
Bixler et al [1998] 4364 (all males); 741 studied in lab	15.9%	4.7%[b]	3.3%

*Young et al [1993] also studied females, but these data are not reported in this table. The sample of commercial drivers we studied was predominantly male. No females were included in the studies of Bixler et al [1998] or Bearpark et al [1995].
[a]AHI≥15 episodes/hour (reported by Young et al [1993]).
[b]AHI≥20 episodes/hour (reported by Bearpark et al [1995] and Bixler et al [1998]).

E. Risk Factors for Sleep Apnea

The presence of sleep apnea is not the same in all segments of society. While there are a number of risk factors such as craniofacial structure [Lowe et al, 1986; Partinen et al, 1988; Lowe et al, 1995; Nelson and Hans, 1997] and altered size of soft tissue structure in the upper

airway [Lowe et al, 1986; Schwab et al, 1995], these risk factors are difficult to quantify in population studies. Thus, population studies have focused on the following risk factors: (1) gender; (2) age; (3) body mass index (degree of obesity); (4) neck (collar) size; (5) ethnicity; (6) smoking habit; and (7) alcohol intake. We describe the data showing the role of each of these factors since each was assessed in our study of prevalence of sleep apnea in commercial drivers. As we will describe, the degree of obesity is the most important risk factor for sleep apnea in middle-aged adults.

1. Gender

Sleep apnea and sleep apnea syndrome are more common in men than women [Redline et al, 1994]. This is also described in the study of Young et al [1993]. Typically for various different definitions of abnormality, sleep apnea is twice as common in men than women. This is so whether we consider breathing abnormalities during sleep (see Table 2.4) or the presence of the sleep apnea/hypopnea syndrome (4% in males, 2% in females).

Table 2.4. Prevalence of different levels of respiratory disturbance during sleep as assessed by apnea/hypopnea index. Data shown are prevalence (%) and the 95% Confidence Intervals of the estimate. From Young et al [1993].

AHI Category	Prevalence in Males	Prevalence in Females
≥5 episodes/hour	24.0% (19-28%)	9.0% (5.6-12.0%)
≥10 episodes/hour	15.0% (12-19%)	5.0% (2.4-7.8%)
≥15 episodes/hour	9.1% (6.4-11%)	4.0% (1.5-6.6%)

2. Age

There is little doubt that sleep apnea is particularly common in the elderly [Ancoli-Israel et al, 1991] but, as mentioned above, this is not germane to our study of commercial drivers. The question of relevance is whether over the ages typical for commercial drivers, sleep apnea prevalence is affected by age. Over the middle-aged years, two of the three major studies show a clear effect of age (Bixler et al [1998] and Young et al [1993]). Given the different definitions that can be employed, we illustrate this effect with one measure that is reported in both of these studies (percentage of subjects with AHI≥5 episodes/hour). Since commercial drivers are predominantly male, we show, for illustration purposes, results for only males (see Table 2.5). Different studies report results in different age ranges. Data like this are not available from the study of Bearpark et al [1995] but they report no significant effect of age on prevalence of sleep apnea.

Table 2.5.   Percentage of males in two of the major studies who had an apnea/hypopnea index above 5 episodes/hour. The ages reported for the two studies are different given how their results are presented. Common to both studies is the result that prevalence increases significantly as a function of age across the middle-aged years.

Age	Young et al [1993]	Age	Bixler et al [1998]
20-29 years	n/a		
30-39 years	17%	20-44 years	7.9%
40-49 years	25%	45-64 years	18.8%
50-59 years	31%	65-100 years	24.8%
60-69 years	n/a		

The age-specific prevalence of the sleep apnea/hypopnea syndrome is reported in the study of Bixler et al [1998] and is shown in Figure 2.3.
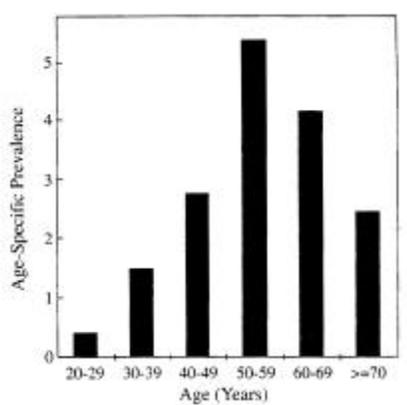
Figure 2.3.   Age-specific prevalence of obstructive sleep apnea diagnosed according to sleep disorders clinic criteria. These results are for males and are from the study of Bixler et al [1998].

3. Body Mass Index (Degree of Obesity)

In middle-aged adults the major risk factor for sleep apnea is obesity. In the study of Young et al [1993], a one standard deviation increase in body mass index (BMI)[1] (a measure of obesity) increased the risk of having an AHI≥5 episodes/hour by 4.17 times. In the study of Bearpark et al [1995], BMI was a significant risk factor for an AHI≥30 episodes/hour increasing the risk by 17.59 times. These studies do not report prevalence of apneas in different categories of BMI. It is to be emphasized that while obesity is a risk factor, not all subjects with sleep apnea are obese.

4. Neck Circumference

It is not simply obesity that is important, but rather the distribution of body fat. Studies have shown a better correlation between the degree of sleep apnea and neck circumference than with a simple measure of overall obesity [Stradling and Crosby, 1991].

---

[1] BMI is a convenient measure of obesity since it represents weight corrected for height. It is calculated as weight in kilograms divided by the square of the height in meters.

Other studies have shown, however, that waist circumference is a better predictor of sleep apnea than neck circumference or BMI [Grunstein et al, 1993].

5. Ethnicity

There are data that indicate that prevalence of sleep apnea is higher among non-Caucasians than Caucasians. In a study by Kripke et al [1997] in San Diego, a higher prevalence of sleep-disordered breathing was found among Hispanics and non-whites; combined prevalence: 16.3% (95% Confidence Interval (CI): 8.1%-24.5%) compared to whites--4.9% (95% CI 2.5-72.). This difference was independent of BMI, age and gender. Likewise, Redline et al [1997b] have reported a higher prevalence of sleep apnea in African Americans. This difference is, however, only significant in males between 20 to 30 years of age and not in older adults.

6. Smokers

A number of cross-sectional epidemiological studies have reported a higher prevalence of snoring (a major symptom of sleep apnea) among smokers compared to non-smokers [Bloom et al, 1988; Schmidt-Nowara et al, 1990a,b; Jennum and Sjol, 1993; Lindberg et al, 1997]. Smoking was not addressed in two of the major prevalence studies of sleep apnea itself, i.e., Young et al [1993] and Bixler et al [1998]. In the study of Bearpark et al [1995], smoking consumption was not related to any of the measures of sleep-disordered breathing. Current smokers did, however, snore for a greater percentage of the night than non-smokers. Thus, smoking does not seem to be a major risk factor for objectively measured sleep-disordered breathing.

7. Alcohol

Alcohol administration in the laboratory acutely increases the frequency of apneic events during sleep and makes them longer, i.e., intensifies the severity of the disorder [Issa and Sullivan, 1982]. Epidemiological data are, however, less clear. Some studies show a relationship between alcohol consumption and snoring on sleep-disordered breathing [Jennum and Sjol, 1993; Koskenvuo et al, 1994] while others do not [Bloom et al, 1988; Schmidt-Nowara et al, 1990a,b; Bearpark et al, 1995; Davies and Stradling, 1995; Worsnop et al, 1998].

Thus, while a number of different risk factors have been assessed, the major positive data are for age (increased prevalence with age), gender (more common in males than females), and obesity (increased prevalence with increased BMI). All of these major factors are relevant to commercial drivers and were considered in the study that we describe in this report.

F. Motor Vehicle Crashes and Sleep Apnea

Since, as discussed above, excessive daytime sleepiness is a consequence of sleep apnea, it is not surprising that it is likely that patients with this disorder have an increased rate of car crashes. This has been addressed in a number of studies (for summary, see Appendix C). These studies have recently been critiqued by Conner et al [2000]. Many of the studies are based on small case series of patients with sleep apnea, who were evaluated in sleep disorders centers.

They compare the crash rates in patients with those in a control group. Given the small size of the samples in many of these studies, there are concerns about selection bias. Moreover, the small sample size limits the precision by which one can determine an estimate of the increase in crash risk. Potential confounding effects of age, gender and driving mileage are often not accounted for. Nevertheless, with these caveats certain features emerge from these studies. First, all studies, apart from one, report an increased risk of crashes in patients with sleep apnea (see Appendix C). Second, in those studies judged by a critical review [Conner et al, 2000] to have at least a modestly robust design, the following estimates of relative increased risk of crashes were obtained: 7.2 (95% confidence intervals 2.4-21.8%) [Terán-Santos et al, 1999]; 3.4 (95% confidence interval 1.4-8.0) [Young et al, 1997b]; 2.58 (95% confidence intervals 1.06-6.31) [Wu and Yan-Go, 1996].

While there are a larger number of smaller studies, three recent major studies, employing different strategies to estimate crash risk associated with sleep apnea in different populations, are worthy of further comment. First, the study of Young et al [1997b] evaluated crash risk in a population-based sample—the Wisconsin Sleep Cohort. The cohort was established from a random sample of state employees in whom a prevalence study of sleep apnea was conducted [Young et al, 1993] (see results of prevalence study in Table 2.3). The vast majority of subjects identified with sleep apnea were never treated. Thus, they became a natural cohort that have been followed over time. Crash records for this group were obtained from the state for a five year period, and the increased relative risk of crashes was determined for different levels of severity of apnea. The results are presented in Table 2.6. An increased risk of crashes, corrected for self-reported miles driven/year, was found in males, but not females. (The number of females in the study is less than males due to the prevalence differences previously discussed.) The reason for the lack of a risk association in females is unknown. The increased risk of crashes was not related to severity of sleep apnea, and was present even at low levels of respiratory disturbance during sleep (see Table 2.6). Given the design of the study, there was not a large number of total crashes over the period of observation, i.e., 165 crashes.

Table 2.6. Increased relative risk of crashes (odds ratio) at different levels of severity of respiratory disturbance during sleep in males (left column) and females (right column). There was a significant increased risk of crashes in males with even low degrees of sleep apnea but not in females.

Any MVA in 5 years (n=165; data are from 913 subjects)

	Increased Relative Risk	
	Men	Women
No SDB*	Reference category = 1.0	
Snorer, AHI<=5	3.4**	0.9
AHI 5-15	4.2**	0.8
AHI >15	3.4**	0.6

*Sleep-disordered breathing
**Significant increase

An alternative approach, i.e., a case-control study, was used by Terán-Santos et al [1999] again based on a population sample. Cases were defined as drivers who had crashes on rural highways in Spain where the driver sustained injury and was identified at an Emergency Room. Drivers charged with violation of alcohol laws were excluded. Controls were matched for age and gender but not for miles driven/year. These controls were randomly selected from patients in primary care centers in the same area of Spain. Controls were excluded if they had "chronic illness" or had a traffic accident in the two previous months. All subjects had a simplified sleep study at home. Those in whom the home study revealed evidence of sleep apnea, or in whom sleep apnea was strongly suspected had full in-laboratory sleep studies. The major results of the study are shown in Table 2.7.

Table 2.7. Relation between sleep apnea and crashes in rural highways.

Apnea/Hypopnea Index*	Case Patients (N=102)	Controls (N=152)	Unadjusted OR (95% CI)	Adjusted OR (95% CI)**
	number of patients (%)		number of controls (%)	
≥5	29 (28.4)	7 (4.6)	8.2 (3.4-19.6)	11.1 (4.0-30.5)
≥10	21 (20.6)	6 (3.9)	6.3 (2.4-16.2)	7.2 (2.4-21.8)
≥15	17 (16.7)	5 (3.3)	5.8 (2.1-16.5)	8.1 (2.4-26.5)

NOTE: OR denotes odds ratio, CI denotes confidence interval

*The three categories of the apnea/hypopnea index are not mutually exclusive, because the number of cases in each category was not high enough to make a proper analysis.

**The logistic-regression model for the adjusted odds ratio considered the presence or absence of an accident as the dependent variable and the apnea/hypopnea index as the independent variable; potential confounders were entered, such as alcohol consumption, visual refraction disorders, body mass index, years of driving, age, involvement in previous accidents, use of medication causing drowsiness, smoking, work and sleep schedule, kilometers driven per year, and coexisting conditions (including psychiatric disorders and arterial hypertension). The apnea/hypopnea scores ≥5, ≥10 and ≥15 are all based on a single continuous index divided at three different levels.

The study, thus, shows a large increased relative risk (odds ratio) of crashes in subjects identified as having sleep apnea. This risk does not appear to change with severity of illness. This conclusion is, however, not definitive since the three categories of apnea/hypopnea index shown are not mutually exclusive. Analyses are not presented for different independent levels of severity of the breathing abnormality during sleep. While the odds ratios are impressive, there must be concern about the control group. The prevalence of AHI≥5 episodes/hour in this group—4.6%—is considerably lower than those reported in previous prevalence studies, i.e., 15.9-25.9% (see Table 2.3). It may be that excluding controls with poorly defined chronic illness, which might have included hypertension, a known consequence of apnea, removed a number of controls with sleep apnea. The prevalence of AHI≥5 episodes/hour in the subjects

who crashed (28.4%) is not much higher than that reported in general population studies (see Table 2.3). Thus, the high odds ratios in this study may be an artefact of the selection of the control group.

The final study to be discussed is that by George and Smiley [1999]. This recent study is the largest case series of sleep apnea patients from a sleep disorders center. Records about crashes were obtained on 600 patients diagnosed as having sleep apnea; the crash records were obtained from state authorities (Ontario, Canada). An increased risk of crashes was found but only in patients with an apnea/hypopnea index above 40 episodes/hour (see Figure 2.4).
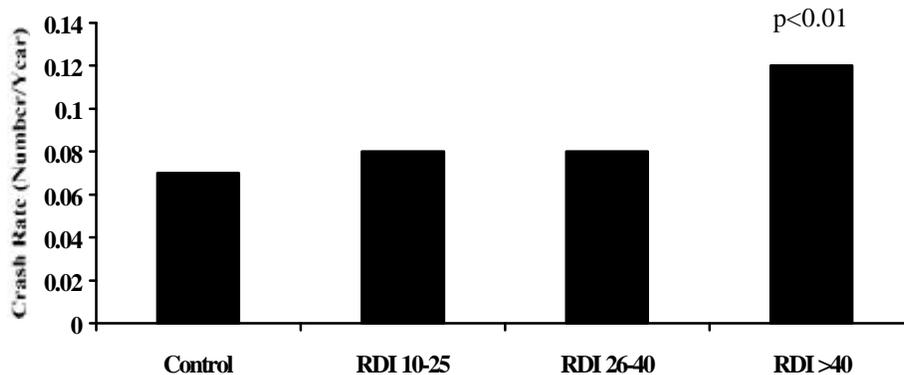
Figure 2.4. Crash rates in controls and patients with different levels of breathing abnormalities during sleep. A significant elevation (p<0.01) of crash rates is only found in patients with an apnea/hypopnea index above 40 events/hour. [George and Smiley, 1999].

These recent results, together with the review by Conner et al [2000], highlight the methodological difficulties in this area. That crashes are relatively rare events, even in persons with marked sleepiness, challenges the power of any study to estimate elevated risk. Thus, the probability of a crash occurring in drivers impaired by sleepiness is relatively low. The probabilistic nature of a crash event might, in part, explain why some studies show elevated crash rates at low levels of respiratory disturbance during sleep while the study of George and Smiley [1999], just discussed, only shows effects at high levels of breathing abnormalities during sleep. While the totality of these studies support the view that is congruent with our knowledge of the physiology of the condition, that sleep apnea is a risk factor for increased automobile crashes, we do not know the precise relationship between severity of the disorder and this risk. This represents a critical gap in our information for policymakers in the traffic safety area.

G. Treatment of Sleep Apnea—A Reversible Cause of Excessive Sleepiness

An important reason to consider obstructive sleep apnea in traffic safety is that it represents a common, treatable condition in which sleepiness can be reversed. Effective treatments for this condition exist. The mainstay of treatment is nasal continuous positive airway pressure (nasal CPAP) [Sullivan et al, 1981]. The concept is simple: air is supplied under a low level of pressure to the back of the throat through the nose via a tight fitting mask. The mask is worn during sleep. The pressure supplied through the mask pushes the airway open and prevents

narrowing and closure of the airway that would occur during sleep, as well as abolishing snoring. It effectively improves sleepiness both objectively measured and reported by the patient, as has been shown in a double-blind trial using as a placebo (control) "sham CPAP", i.e., wearing the mask but not with effective pressure [Jenkinson et al, 1999]. (Double-blind means that neither the patient nor their physician knew whether the patient was on active or "sham" CPAP.) There are few major side effects of CPAP treatment [Jarjoun and Wilson, 1989; Bamford and Quan, 1993; Nino-Murcia et al, 1989; Pack, 1994], although many patients have "nuisance" side effects—runny nose (due to the pressurized air), claustrophobia (due to the mask), and skin abrasions (due to the tight-fitting mask). These side effects are experienced by many people on this form of therapy [Nino-Murcia, 1989; Pack, 1994].

While there are few major side effects with nasal CPAP, the major problem with this therapy is adherence, i.e., continued use of this cumbersome therapy. A large percentage of patients initiating therapy with nasal CPAP, in some cases as high as 45% [Kribbs et al, 1993a], do not use it enough to be effective [Waldhorn et al, 1990; Krieger and Kurtz, 1988; Rolfe et al, 1991; Kribbs et al, 1993a; Reeves-Hoche et al, 1995]. Those who do use it wear the mask on average for less hours during sleep than the desired amount. Attempts to improve adherence to therapy have been recently introduced [Hoy et al, 1999], and produce some improvements. Even with problems with adherence, the double-blind clinical trial alluded to above show significant improvement in sleepiness on an "intent to treat basis," i.e., without considering who is, or who is not, adherent to therapy. Treatment studies also report that nasal CPAP improves driving performance in a controlled double-blind study [Hack et al, 2000]. There is some evidence of reduction in crash rates. This has been shown in two studies where crashes and/or reports of near-misses [Cassel et al, 1996; Yamamoto et al, 2000] were obtained from the patients by self-report. Both studies showed lower crash and/or near-miss rates following one year of beginning to use nasal CPAP (sleep apnea/crashes). But the occurrence of crashes in these patients was not independently verified and hence both studies are subject to recall bias. Small case series are starting to appear that also report reduction in crashes, as obtained from state driving records, following institution of nasal CPAP therapy [Findley et al, 2000].

While nasal CPAP is effective in reducing sleepiness, there remain issues as to whether patients with mild-to-moderate disease benefit from therapy (for definitions of severity, see Table 2.2) [see Davies and Stradling, 2000]. Some studies are starting to address this but more data are needed. Engleman et al [1999] have shown in a placebo-controlled double-blind trial that in patients with an apnea/hypopnea index between 5 and 15 episodes/hour nasal CPAP improves patients self-report of sleepiness (subjective sleepiness) but not objectively measured sleepiness. This is in contrast to the results in patients with more severe disease where improvements in both subjectively and objectively measured sleepiness are found [Jenkinson et al, 1999]. This is an important issue for policymakers in traffic safety since the number in the population with an AHI between 5 and 15 is large and benefit in terms of improvement of performance has not been conclusively demonstrated. This suggests that, at least initially, our focus should be on those with more severe disease.

Since adherence to CPAP is a problem, the question is raised whether there are other therapies. There are, but they have been demonstrated in randomized trials to be not as effective as nasal CPAP. The first is use of an intra-oral device put in by the dentist (or maxillo-facial

specialist) that is worn during sleep. This device moves the jaw forward, thereby making the upper airway larger and reduces the narrowing of the airway during sleep [Ayas and Epstein, 1998]. This approach to therapy is not effective in all patients. There are some data suggesting that the success rate of this form of therapy is best in those with mild disease [Marklund et al, 1998]. No studies have addressed the effects of this form of therapy on crash risk. [For review of use of these devices, see Schmidt-Nowara, 1995; Ayas and Epstein, 1998.]

Other alternative therapies include a number of surgical approaches. Both surgery to the soft palate of the upper airway and bony surgery have been proposed. Soft tissue surgery involves removing part of the uvula and soft palate in an operation called a uvulopalatopharyngoplasty (UPP). It has a variable success and substantial reduction in breathing abnormalities during sleep occur in only about 50% of patients having this surgery [for review of literature on results of surgery, see Sher et al, 1996]. One study reports reduced crash rates following this form of surgery [Haraldsson et al, 1995a] while another indicates improvements in driving performance [Haraldsson et al, 1995b]. Surgery to reconstruct the facial bones has a higher success rate, of the order of 75% to 90% [Riley et al, 1993; Waite et al, 1989, Hochban et al, 1997; Conradt, et al, 1997; Bettega et al, 2000], but many patients do not choose to undergo this form of therapy [Riley et al, 1993]. There are no data on change in crash rates following this type of surgery.

Thus, while there are alternative therapies, the mainstay of treatment is nasal CPAP. In evaluating the issue of sleep apnea in the commercial drivers, an important consideration is whether this form of therapy is compatible with the driver's lifestyle and whether they will be adherent to therapy. The machines to deliver the pressure are small and extremely portable. Thus, there seems no reason why they could not be used by commercial drivers should they be diagnosed as having this condition.

H. Sleep Apnea and the Commercial Driver

Given that excessive daytime sleepiness is a major consequence of obstructive sleep apnea, and there are data supporting that patients with sleep apnea have an increased risk of car crashes (see Appendix C), it is not surprising that attention turned to the role of obstructive sleep apnea in commercial drivers. Since commercial drivers are more obese as a group [Rather et al, 1981], a major risk factor for obstructive sleep apnea (see above), it was suspected that they might have an increased prevalence of the disorder. Stoohs et al [1995] studied sleep apnea in commercial drivers. This study was done at a single company hub and drivers going to the hub during the period of the study were recruited. Thus, this is in the nature of a convenience sample and there was no attempt to sample all drivers in the company. All drivers attending the hub during the period of study (n=388) were approached and all of these subjects filled out questionnaires about sleep apnea symptoms etc. Out of the 388 total drivers, 213 were scheduled to spend the night at the facility and from this group 159 had overnight studies of breathing abnormalities during sleep. Thus, 41% of the drivers studied had sleep studies. The basis for inclusion in this group was that the driver was overnight at the hub and agreed to be studied. Thus, there is the potential for selection bias. The sleep studies were done at the hub using portable equipment, i.e., is not a full sleep study. The equipment (MESAM 4) measured drops in oxygen as a surrogate measure of hypopneas and apneas but it did not record polysomnography,

which is the standard sleep disorder evaluation. There was no difference between body mass index and sleep apnea symptoms between those getting sleep studies and those who did not. The group getting sleep studies were, moreover, slightly younger (35.1 years vs. 37.3 years, p<0.02).

The study found a large prevalence of sleep apnea. <u>Seventy-eight percent</u> of drivers studied had an apnea/hypopnea index ≥5 respiratory events/hour and 10% had ≥30 events/hour. These numbers are the <u>highest</u> reported in any study of prevalence. 78% should be compared to the results presented in Table 2.3 where the prevalence for an AHI≥5 episodes/hour were 24% [Young et al, 1993], 25.9% [Bearpark et al, 1995] and 15.9% [Bixler et al, 1998] in the different studies reported. The high prevalence in the study of commercial drivers might, at least in part, be related to the higher degree of obesity in this group. However, based on what we know about the effect of increased BMI on apnea prevalence [Young et al, 1993; Bearpark et al, 1995], the large difference between prevalence in this study and others cannot be explained on this basis. It might be argued that the high prevalence is an artefact of the portable testing procedures. This, however, seems unlikely. The same equipment (MESAM 4) was used by Bearpark et al [1995] in their study of apnea prevalence in Busselton, Australia, and they obtained similar prevalence numbers to those using in-laboratory studies in other investigations. Thus, the very high prevalence of sleep apnea in the study of Stoohs et al [1995] in commercial drivers remains an outlier that requires explanation.

As part of their study, Stoohs et al obtained crash records for the drivers they studied. They found a doubling of the commercial vehicle crash rates in drivers with apnea [Stoohs et al, 1994]. This difference was, however, <u>not statistically significant since the crashes were primarily contributed by two drivers</u>. It is unfortunate that this statistically unreliable difference was widely reported by the national media in the United States, and is still quoted in articles advocating policy in this area [McNicholas, 1999]. Thus, currently there is no statistically sound evidence of an increased crash rate among commercial drivers with sleep apnea although the matter has received very limited study. The study of Stoohs et al [1994] had limited power to address this question. This issue, therefore, represents another critical gap in knowledge needed to inform public policy debate in this area.

I. <u>Study Goals</u>

The study of Stoohs et al [1995] raised obvious concerns about how common sleep apnea is in commercial drivers. Moreover, since sleep apnea is a continuum of abnormality, it raises important questions as to what level of abnormality one sees in commercial drivers' and whether there are resulting decrements in performance that would degrade the driving task. In commercial drivers there are multiple reasons for excessive sleepiness: (1) shift-work schedules that make sleep difficult [Mitler et al, 1997]; (2) habitual short sleep durations [Mitler et al, 1997]; (3) as well as sleep disorders such as sleep apnea. Investigating the relative role of each of these in degrading performance is an important issue. Thus, the objectives of our study were the following:

1. To estimate the prevalence of sleep apnea among a sample of commercial drivers. Our sample was based on a random sample of holders of a commercial drivers license who lived in Pennsylvania within a 50-mile radius of Philadelphia.

2. To examine the relationship in commercial drivers between severity of sleep apnea and decrements in function related to driving tasks.

3. To develop a profile of our overall sample of commercial drivers with regard to their sleep apnea-related characteristics and risks.

CHAPTER THREE

Study Design and Characteristics of Population Studied

3.1 Overall Study Design

The overall study design we adopted was that initially proposed by Gislason et al [1988]. It was used in the landmark study by Young et al examining the prevalence of sleep apnea in a random sample of Wisconsin state employees [Young et al, 1993]. The study design is based on a *two-stage sampling strategy*. In the first stage the population of interest is determined. Information is obtained from the population to allow classification of the population into two groups: (a) those at higher risk for the presence of sleep apnea; and (b) those at lower risk for the presence of sleep apnea. In the study of Young et al [1993], this risk stratification was based on the presence of symptoms of sleep apnea (see Chapter Two). In the second stage of the study, weighted samples are taken from the higher risk and lower risk groups for in-laboratory testing so that the presence and degree of sleep-disordered breathing can be established. In this weighted sampling scheme the ratio of the number of higher risk subjects to lower risk subjects is larger in the selected sample compared to the same ratio in the population. This design enriches the sample in terms of the number of subjects with apnea that are available for assessment of functional consequences while allowing a population-based estimate of prevalence. Furthermore, since apnea prevalence is likely to vary greatly between the higher and lower risk strata, statistical precision of estimated proportions is likely to be improved [Cochran, 1977; see further in Appendix E] as a consequence of the blocked sampling.

In implementing this design, one needs to decide on the following: (a) the target population to be studied or *sampling frame*; (b) the information to be obtained to allow risk stratification into groups at higher risk and lower risk for sleep apnea; (c) the nature of the second stage sample for in-laboratory testing; and (d) the methods to be employed for in-laboratory testing. We discuss each of these steps for our study in subsequent sections of this chapter. The overall study design is shown schematically in Figure 3.1.

**TWO-STAGE SAMPLING STRATEGY FOR ESTIMATION OF
PREVALENCE OF SLEEP APNEA**
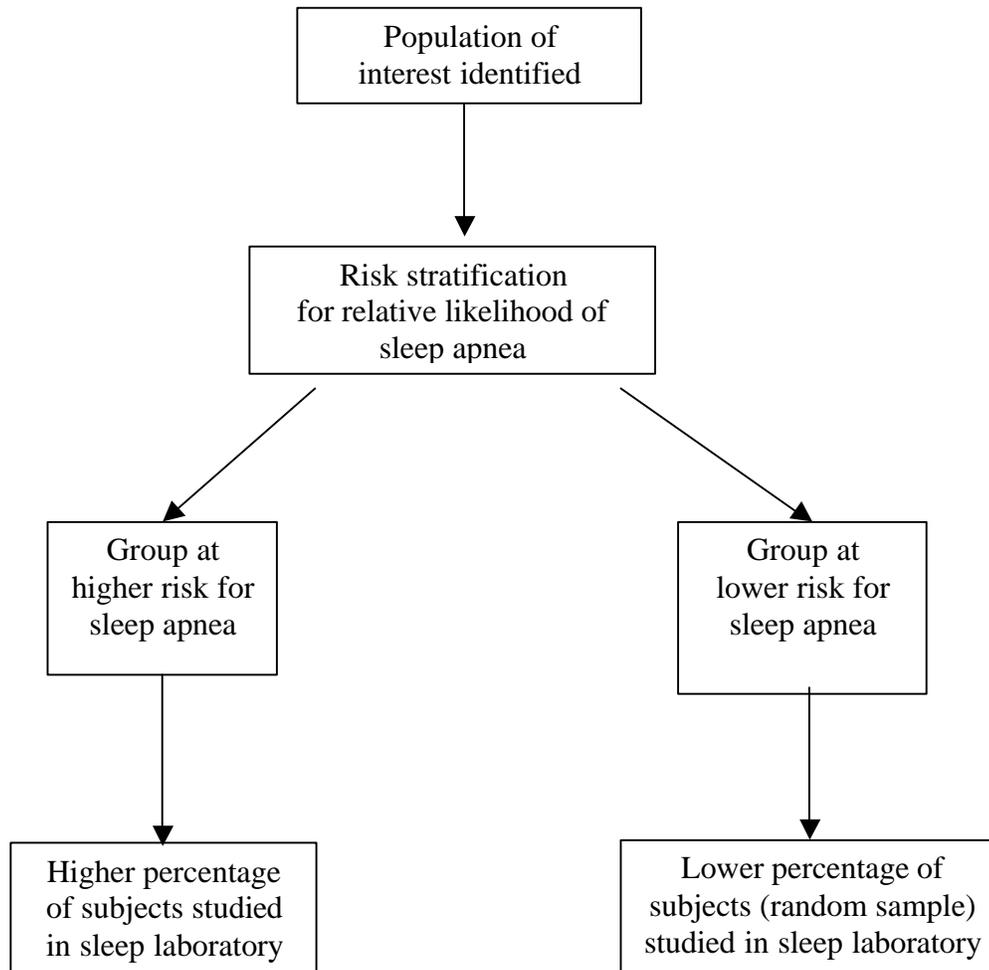
Figure 3.1. Two-stage sampling scheme proposed by Gislason et al [1988] and used by Young et al [1993] to estimate prevalence of sleep apnea in a population of interest.

3.2 Population Sampled

The first step in this design is to decide the population of interest or the sampling frame. For commercial drivers, at least two approaches could be considered. The first is to base the sample on specific companies. This is the approach used by Stoohs et al [1995] in the study described in Chapter Two, and is that being used in a recently initiated study of prevalence of commercial drivers in Australia [see Howard et al, 2000]. The concern about this approach is whether the company, or companies, chosen have selective recruitment policies that will influence the type of drivers employed. Moreover, in order for sampling by company to be both feasible yet provide unbiased prevalence estimates, sampling with probability proportional to size would need to be done [Cochran, 1977]. In this approach, a fixed number of drivers from

each company would be recruited for study from companies selected at random, with probability proportional to their number of employed drivers. Thus, this approach would require, in advance of the randomized selection process, a precise delineation of the sampling frame comprising of all possible companies that employ eligible drivers along with their numbers of such drivers. It would be difficult to establish the proper sampling frame unless one pre-selects certain characteristics of the company, e.g., they belong to a particular national organization from which a list of companies could be obtained. Moreover, if a selected company refused to participate, this would have a much more deleterious effect on the question of generalizability than if an individual driver refused to participate.

The second approach, and the one adopted here, is to define the sampling frame in terms of drivers in some well-defined population, with selection independent of which company they drive for. To identify drivers, we obtained a random sample of 4826 holders of commercial drivers licenses from the State of Pennsylvania. The only criterion for selection was that they needed to live within 50 miles of the Sleep Center at the University of Pennsylvania in Philadelphia, where laboratory studies were to be conducted. This approach has the advantage that the sampling frame is precisely defined, permitting a true random sample of drivers in a precisely defined population. It has the disadvantage that not all drivers so identified were currently working as commercial drivers. They do, however, have the capability to do so and there is a large amount of job turnover in this industry [ATA Foundation, 1997]. The second disadvantage of this approach is the response rate. This is considered more fully below.

Thus, for the population of interest, we chose to study a random sample of commercial drivers license (CDL) holders living in Pennsylvania within 50 miles of Philadelphia. This sample was provided by state licensing authorities.

3.3 Risk Stratification

In the study of Young et al [1993], risk stratification was based on the presence of symptoms of sleep apnea. In our study described here, we used a more precise tool that has arisen from our investigations of patients with sleep apnea in our sleep clinic population—the Multivariable Apnea Predication (MAP) [Maislin et al, 1995]. This instrument arose from factor analysis of questionnaire responses from individuals having sleep studies for clinical evaluation at the University of Pennsylvania; Johns Hopkins Center for Sleep Disorders, Baltimore, Maryland; and West Penn Center for Sleep Disorders, Pittsburgh, Pennsylvania. (For full description of development of this tool, see Appendix D, in which the manuscript describing the instrument is presented.) It combines responses to questions about the frequency of symptoms of apnea, measurement of body mass index (the major risk for sleep apnea [see Chapter Two]), age and gender to compute the relative likelihood of sleep apnea on a scale between 0 and 1.

For each of the apnea symptom frequency questions, on which this prediction is based, the subject is asked the following: During the last month, have you had, or have been told about the following symptom (Show the frequency): (0) Never; (1) Rarely, Less Than Once a Week; (2) 1-2 Times Per Week; (3) 3-4 Times Per Week; (4) 5-7 Times Per Week; (.) Don't Know. The questions asked about: (1) Loud snoring; (2) Snorting or gasping; and whether (3) Your breathing stops or you choke or struggle for breath.

Answers to these questions provide an apnea symptom frequency score with values between 0 and 4. This score is called *Index 1* and is computed as the mean of the responses. Zero means that the individual never has any of these three symptoms. A score of 4 means that all three symptoms are present 5-7 nights per week. This symptom frequency score is combined with body mass index (for definition, see footnote in Chapter Two), age and gender to calculate the relative likelihood of apnea, i.e., multivariable apnea prediction (MAP), using the following formula:

$$MAP = e^x/(1 + e^x)$$
where $x = -8.160 + (1.299 * \text{Index 1}) + (0.163 * BMI)$
$\quad - (0.028 * \text{Index 1} * BMI) + (0.032 * Age)$
$\quad + (1.278 * MaleGender)$
where MaleGender=1 if subject is male and 0 if female.

The behavior of this instrument is shown in Figure 3.2. This figure shows the relative likelihood of apnea (Y axis) on a scale of zero to one as a function of body mass index (BMI)[2] (X axis) for a 49-year-old male. The lines shown are for different values of the symptom frequency score (Index 1). For females and for younger males, the lines would be shifted downward. For older males, the lines would be shifted upward. As can be seen in the figure at very high levels of BMI, i.e., around 45 kg/m^2, the lines converge since at this level of BMI the likelihood of apnea is very high with or without symptoms. At lower levels of BMI, e.g., around 30 kg/m^2, the likelihood of sleep apnea is much more dependent on whether the individual has symptoms or not.
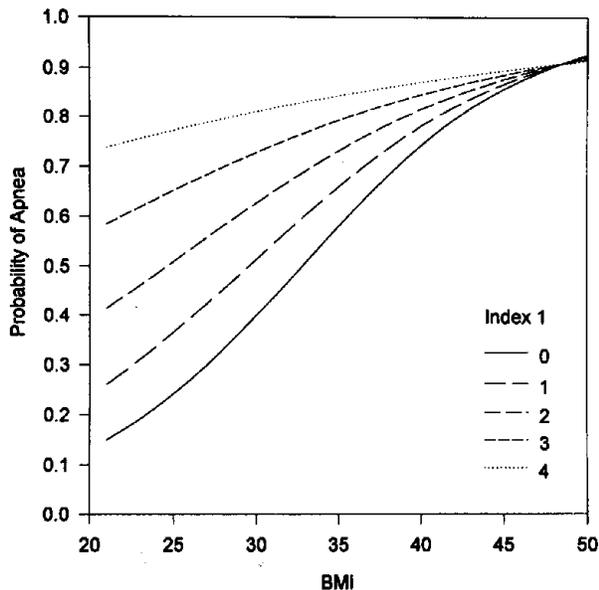
Figure 3.2. Relative likelihood of sleep apnea as a function of body mass index (BMI) for different values of the symptom frequency scores for sleep apnea (Index 1) between zero and four. Likelihood depends on age and gender. The example shown is for a 49-year-old man. (Reproduced from Maislin et al [1995].) For patients presenting at sleep disorders centers, the MAP represents the probability of having sleep apnea. For all other populations, the MAP reflects a relative likelihood of sleep apnea facilitating relative risk comparisons between subjects and between groups.

---

[2] BMI is a convenient measure of obesity since it represents weight corrected for height. It is calculated as weight in kilograms divided by the square of the height in meters.

We used this instrument in our study to rank order all survey respondents from highest to lowest values of MAP, i.e., from highest to lowest relative likelihood of sleep apnea. The instrument is not precise enough to tell whether an individual subject has apnea or not, but can be used in population studies, such as that reported here, to stratify subjects into classes with different relative likelihoods of sleep apnea. This instrument is also being used for a similar purpose in a study of sleep apnea in commercial drivers that has recently been started in Australia [Howard et al, 2000].

Table 3.1 summarizes the responses (see below). From the survey of 4826 CDL holders, we received 1486 (33.0%) responses, of which 1391 (31.5%) were potentially usable. The 95 unusable responses included those from spouses indicating that the CDL holder had died or from others indicating that the CDL holder had moved, etc. Among the 1391 useable responses, 62 did not have sufficient data to calculate a value of their apnea risk (MAP) necessary for risk stratification. Among these respondents with MAP values, i.e., 1329, our goal was to recruit 250 CDL holders at the highest risk for sleep apnea and then select a random sample from the remaining respondents to form a lower risk group. We recruited 247 from the top 551 (44.8%) highest risk respondents. We then randomized the order of the remaining 778 respondents. Taking into account the two-stage sampling design, we determined that if 160 lower risk CDL holders were added to the sample, the true prevalence of at least moderate sleep apnea (e.g., AHI$\geq$15 events/hr) in our population would be estimated with a margin-of-error equal to approximately between $\pm$3% and $\pm$4%. At the end of the recruitment period, 159 lower risk CDL holders had been studied. For further discussion of this approach to estimation of prevalence and justification for numbers of subjects studied, see Appendix E. Details regarding comparisons between respondents and non-respondents are provided in Section 3.5.

3.4 <u>In-Laboratory Studies</u>

All of the higher and lower risk subjects studied in the sleep laboratory were studied using identical protocols. This protocol is outlined in Figure 3.3.

Actigraphy and sleep diary	For one week at home
Admission to Clinical Research Center for Sleep	Evening
Medical history/physical examination/questionnaires	Evening
Training in performance tests	Evening
Full sleep study (polysomnography; PSG)	Overnight
Repeated tests of daytime performance (every 2 hours)	Daytime
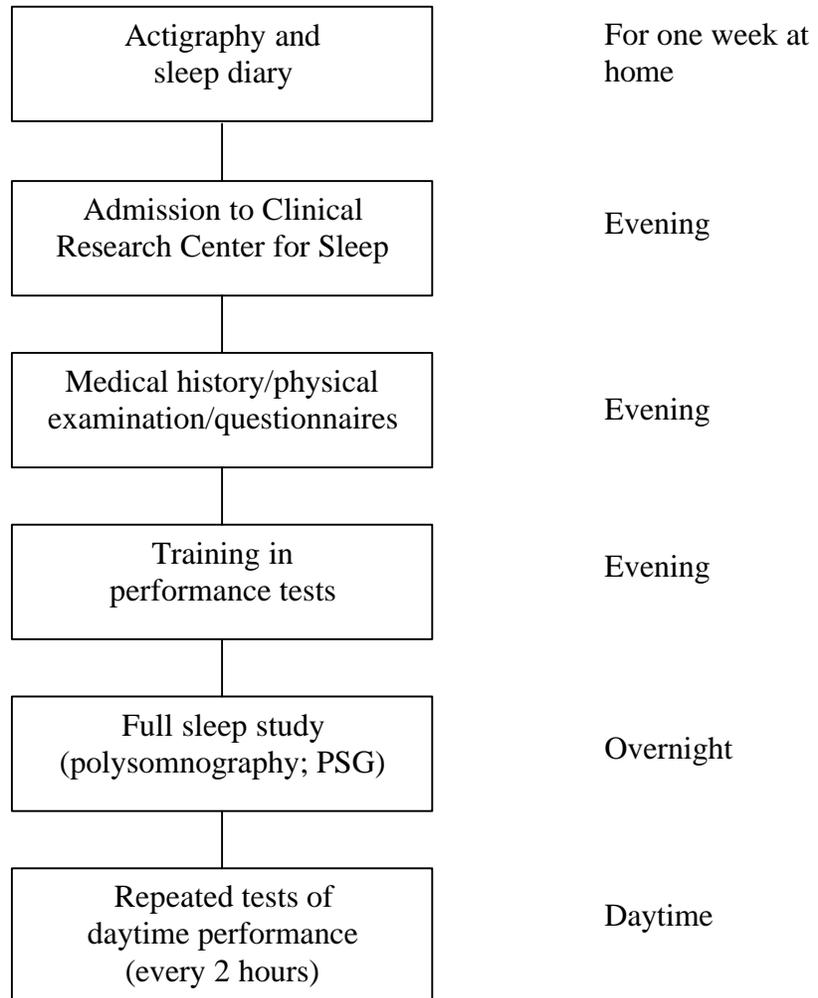
<u>Figure 3.3</u>. Steps in conduct of in-laboratory phase of the investigation.

In brief, all subjects recruited for this part of the study first wore a wrist activity monitor for one week and filled out diaries about their sleep before coming to the sleep laboratory. The activity monitor is worn on the wrist and detects rotational motion. When awake, the subject is moving and the instrument detects motion; when asleep, the subject is relatively motionless. Thus, the instrument provides reasonably precise estimates of sleep duration at least in normal individuals [Blood et al, 1997; Jean-Louis et al, 1997]. The results of this part of our study are described in more detail in Chapter Four. We also discuss in Chapter Four the problems that occur estimating sleep duration in individuals with sleep apnea, since movements occur during sleep in such individuals as they arouse at the end of their apneic events.

After wearing the actigraph for one week, subjects reported to the Clinical Research Center for Sleep at the University of Pennsylvania. They had a full medical history and physical examination. In the evening they also had training on the various performance tests that were to be conducted the next day. They also filled out questionnaires about their sleep, symptoms of sleep apnea, medical history, etc.

On the first night in our Center, the subjects had a full sleep study performed. We measured their brain waves (electroencephalogram, EEG) and muscle activity (electromyogram, EMG) to permit measurement of sleep staging, breathing, breathing effort and oxygen level. This allowed us to determine breathing abnormalities during sleep so that the presence and degree of sleep apnea could be detected. The techniques employed and the results are described in more detail in Chapter Five.

Finally, on the day after the overnight sleep study, the subjects had experimental tests of performance done. We assessed the degree of physiological sleepiness by the multiple sleep latency test (MSLT) [Carskadon and Dement, 1982; Carskadon et al, 1986]; performance on the Divided Attention Task (DADT) [Moskowitz and Burns, 1977; George et al, 1996]; and we determined responses on the Psychomotor Vigilance Test (PVT), Reaction Time Task [Dinges and Powell, 1985; Dinges and Kribbs, 1991; Wyatt et al, 1997] and other objective and subjective functional measures contained in our computerized Neurobehavioral Assessment Battery (NAB) [Dinges and Kribbs, 1991; Dinges, 1992; Kribbs and Dinges, 1994]. The natures of these various performance tests are described in more detail in Chapters Six and Seven, where the results are also presented.

3.5 Nature of Sample of Commercial Drivers

As outlined above, the first phase of our study was to obtain a random sample of holders of a commercial drivers license (CDL) of Pennsylvania residents within 50 miles of the Sleep Center at the Hospital of the University of Pennsylvania. We asked the State of Pennsylvania for a random sample of 5,000 drivers meeting these criteria. The State, for reasons not explained, supplied a list of 4,826 drivers. For each driver we were provided name, address, gender and date of birth from the database on licenses. We sent out an initial letter to each such driver informing them of the study and asking them for the following: basic information about themselves, including current driving occupation; number of miles driven per year; height and weight to allow us to calculate body mass index (BMI); and the frequencies of symptoms related to sleep apnea to allow us to compute the multivariable apnea prediction (MAP). The questionnaire we used for this stage of our investigation is shown in Appendix F). We provided $5.00 to each driver who returned a completed questionnaire. CDL holders who did not respond initially were sent a second mailing (for numbers see Phase 2 mailing in Table 3.1). This approach was based on a set of earlier pilot studies that we had conducted.

Overall Response Rate

Table 3.1 summarizes the response rates to the two mailings.

Table 3.1. Questionnaire response rates.

	Phase 1 Mailing	Phase 2 Mailing	Total
Number Mailed	4826 (100%)	3559 (100%)	4826 (100%)
Returned to sender	222 (4.6%)	99 (2.0%)	321 (6.7%)
Possible (i.e., not returned)	4604 (100%)	3459 (100%)	4505 (100%)
No response	3559 (77.3%)	3018 (87.3%)	
Total Responses	1045 (22.7%)	441 (12.7%)	1486 (33.0%)
Total Usable Responses	980 (21.5%)	409 (11.9%)	1391 (30.8%)

Of the 4,826 names provided by the State of Pennsylvania, 321 could not be contacted with letters to them returned by the U.S. Postal Service. Another 95 were deemed unusable and included those from spouses indicating that the CDL holder had died or from others indicating that the CDL holder had moved, etc. Thus, the final number of potentially recruitable CDL holders was 4410. From these, we received 1391 (31.5%) responses.

Comparisons between Respondents and Non-Respondents

As mentioned above, the records provided by the Pennsylvania Department of Motor Vehicles contained age, gender, and geographic information, allowing for comparisons between respondents and non-respondents. Table 3.2 compares the gender distribution, mean ages, and mean ages within gender between respondents and non-respondents. These were nearly identical between these two groups.

Table 3.2. Age and Gender Distributions in Respondents and Non-Respondents.

	Respondents* (n=1391)	Non-Respondents (n=3437)	Difference
Male (%)	92.2%	92.1%	$\chi^2 = 0.02$, p=0.90 Not Significant
Age	Mean (SD)	Mean (SD)	
Males	44.6 (11.3)	44.5 (11.9)	$t = 0.02$, p=0.94 Not Significant
Females	42.1 (10.4)	43.5 (10.4)	$t = -1.21$, p=0.23 Not Significant

To further illustrate the similarity between respondents and non-respondents, we show in Figure 3.4 the detailed age distribution of the initial total sample and of respondents. As can be seen, there is a remarkable similarity in the age distribution histograms between CDL holders in the sampling frame and CDL holders who responded to the survey.
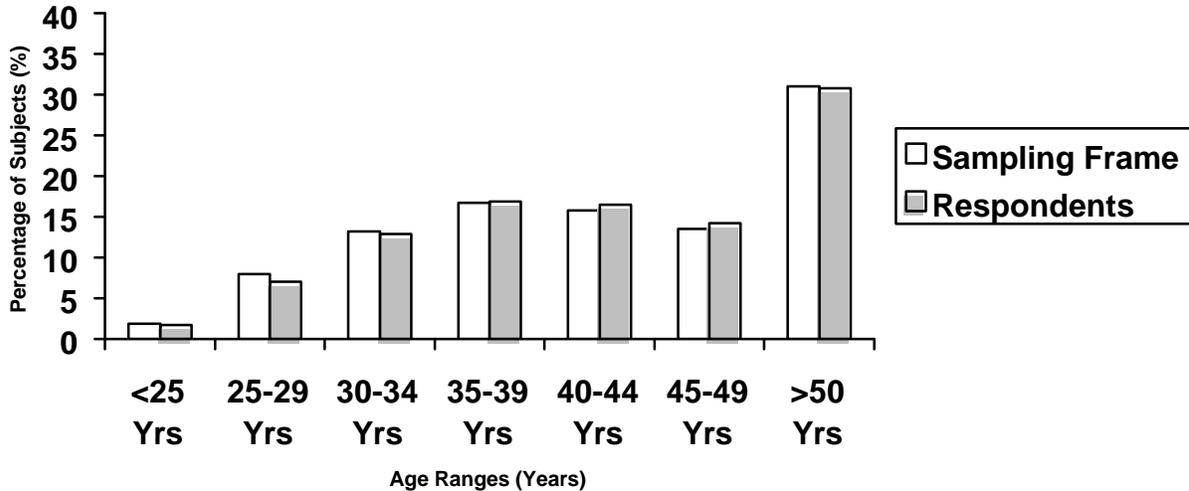
Figure 3.4. Percentage of subjects in the sampling frame and of respondents in different age categories. There is, as can be seen, a remarkable similarity between these two distributions.

There are also minimal differences between the location of drivers in respondents and non-respondents. We assessed this by calculating the percentages of both groups samples residing in the seven Pennsylvania counties falling within 50 miles of the Center for Sleep. These data are shown in Table 3.3.

Table 3.3. County codes in respondents and non-respondents.

County	Responder n (%)	Non-Responder n (%)	Total
A	1 (0.07)	3 (0.09)	4
B	345 (24.84)	668 (19.44)	1013
C	174 (12.53)	479 (13.94)	653
D	0 (0)	1 (0.03)	1
E	229 (16.49)	564 (16.41)	793
F	270 (19.44)	603 (17.54)	873
G	370 (26.64)	1119 (32.56)	1489
Total	1391	3437	4826

We cannot, however, exclude the possibility that there were differences between respondents and non-respondents in other aspects that we do not have data for. Particularly important is degree of obesity as measured by body mass index (BMI) since this is a major determinant of apnea prevalence. This issue is discussed in more detail at the end of this chapter.

3.6 Risk Stratification—Distribution of Multivariable Apnea Prediction (MAP)

As outlined above, we used the multivariable apnea prediction (MAP) [Maislin et al, 1995] to separate the group of respondents into higher and lower risk groups, i.e., for sleep apnea. To do so, we obtained height, weight and information about the frequency of symptoms of sleep apnea from respondents. Complete information was obtained from 1329 of 1391 respondents. The multivariable apnea prediction is dependent, at least in part, on body mass index (BMI). The distributions of BMI in our sample of respondents and the group selected to be studied in-laboratory are shown in Figure 3.5, with respect to different degrees of obesity [Clinical Guidelines NIH, 1998; Flegal et al, 1998]. The shift toward larger BMI values in the in-lab sample was expected and is a consequence of the over-sampling on the basis of the MAP.

Relative Frequency Distributions of BMI in Responders and In-lab Samples
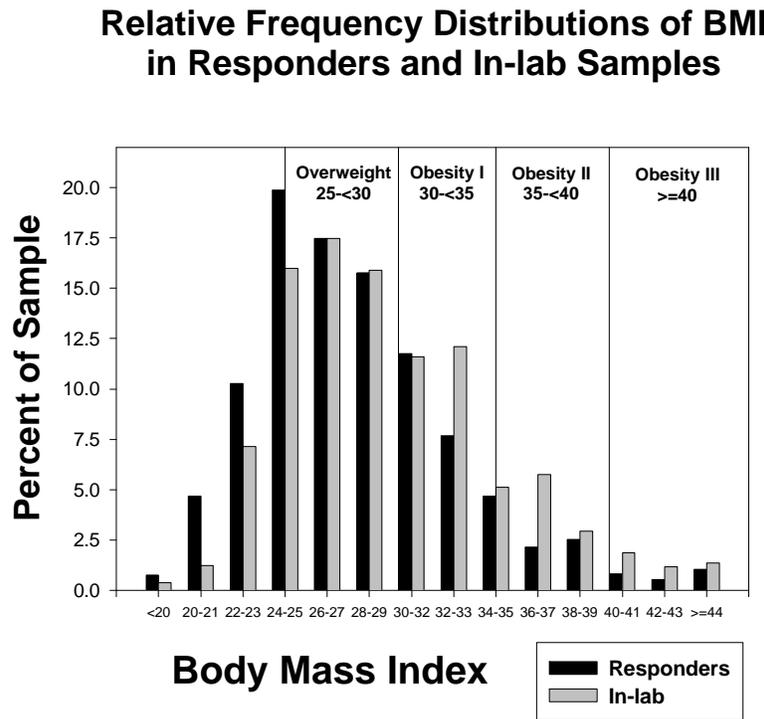
Figure 3.5. Percentage of overall respondents and those studied in the laboratory in different categories of obesity, i.e., overweight (BMI ≥25 & ≤30), obesity stage 1 (BMI ≥30 & ≤35), obesity stage 2 (BMI ≥35 & <40), and obesity stage 3 (BMI ≥40).

A substantial percentage of drivers who responded were obese. Among male respondents, 77.5% were overweight, i.e., with a BMI above ≥25 m/kg^2. This compares to 65% for males in Pennsylvania [Pennsylvania Department of Health, 1998]. To more accurately describe obesity rates, we show in Table 3.4 the percentage of male drivers in our sample at

different levels of obesity compared to the most recent published data about obesity rates in the United States [Flegal et al, 1998]. The different levels of obesity are those defined by the recent guidelines of the National Heart, Lung and Blood Institute [Clinical Guidelines, 1998]. We describe this for only males since the vast majority of our sample (92.8%) was male.

Table 3.4. Percentage of population in different age groups who were overweight (BMI: 25-29.9), class 1 obesity (BMI: 30-34.9), class 2 obesity (BMI: 35-39.9) and class 3 obesity (BMI ≥40). Data are shown for all male respondents in whom we had these data and for all males studied in-laboratory (n=384). These data are compared to recently published national data [Flegal et al, 1998].

	Ages (Years)					
	20-29	30-39	40-49	50-59	60-69	70-79
BMI Category						
25-29.9 (overweight)						
National data (male)	30.6	40.9	42.4	44.1	45.4	43.1
Male respondents	43.9	46.7	43.9	48.8	49.2	47.4
Male drivers studied in lab*	32.0	44.8	38.0	45.7	50.0	50.0
30-34.9 (Class 1 Obesity)						
National data (male)	8.4	11.6	16.5	22.5	20.4	15.2
Male respondents	20.6	20.1	20.1	26.2	25.0	21.1
Male drivers studied in lab*	26.8	26.3	28.1	29.0	35.2	35.8
35-39.9 (Class 2 Obesity)						
National data (male)	2.9	2.8	4.4	4.5	3.5	4.6
Male respondents	2.8	7.1	6.8	6.9	4.2	5.3
Male drivers studied in lab*	9.8	7.8	14.1	8.7	9.9	8.9
≥40 (Class 3 Obesity)						
National data (male)	1.2	2.9	2.2	1.9	1.0	0.2
Male respondents	1.9	1.9	3.1	1.6	0.0	0.0
Male drivers studied in lab*	2.6	5.7	4.3	3.8	0.0	0.0
Sample Sizes						
Male respondents	107	368	383	248	120	19
Male drivers studied in lab*	19	92	130	81	50	12
*All prevalence values for drivers studied in laboratory were estimates as weighted average of higher and lower risk group prevalence within age strata. Weights were determined within age category from population-based screen (n=1282 males).						

As can be seen in Table 3.4, the driver population from which we recruited had higher prevalence of obesity in all categories in all age groups as compared to national data for males. Chi-square goodness-of-fit tests demonstrated significant differences between respondent obesity category distributions and national norms for males with ages 20-29 (χ^2=38.2, df=4, p<0.001), ages 30-39 (χ^2=78.5, df=4, p<0.001), ages 40-49 (χ^2=36.6, df=4, p<0.001), and ages 50-59 (χ^2=16.02, df=4, p=0.003). In the general population, the prevalence of obesity at a level of at least class 1, i.e., BMI greater than or equal to 30 m/kg^2, increases across the middle aged years between 20 and 60 years from 12.5% to 24.9%. In contrast, among the holders of CDLs there is

little increase with age since CDL holders in their 20s have already high rates of obesity of at least class 1 (25.3%). These high obesity rates should be a source of concern. Obesity leads to an increased risk for a number of medical conditions including: hypertension [Hermansen, 2000; National Task Force on the Prevention and Treatment of Obesity, 2000; Kannel, 2000; Coulhoun et al, 1998], diabetes [Sakurai, 2000; Seidel, 2000; Fujimoto, 2000; Sheehan and Jensen, 2000; National Task Force on the Prevention and Treatment of Obesity, 2000; Kopelman, 2000], ischemic heart disease [Kopelman, 2000; Anonymous, 1998; Brochu et al, 2000] and stroke [Gillum, 1999; Anonymous, 1998]—in addition to the increased risk of sleep apnea that is the focus of this report (see discussion in Chapter Two).

Using BMI data and symptom scores for sleep apnea, we computed for each respondent the multivariable apnea prediction index. The distribution of this in our sample is shown in Figure 3.6. High scores for MAP, that is close to 1.0, indicate high relative likelihood of sleep apnea, while lower scores imply low relative likelihood.
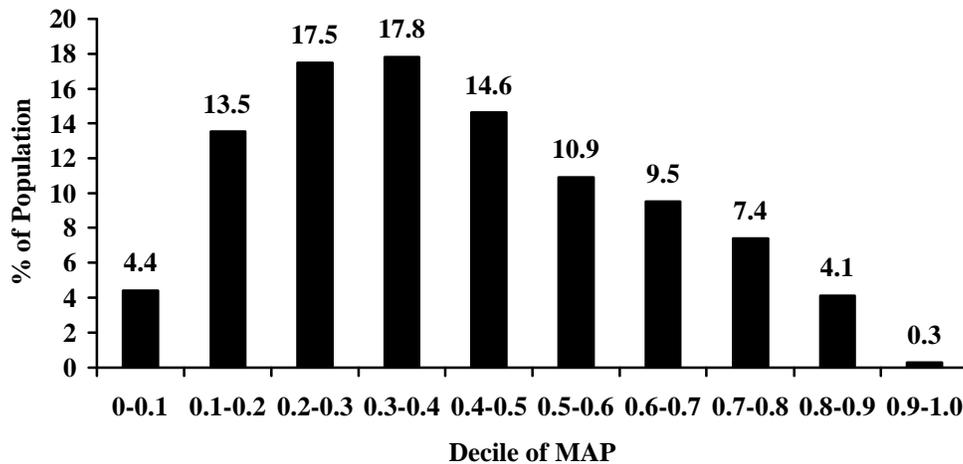
Figure 3.6.   Distribution of multivariable apnea prediction (MAP) between 0 and 1 in respondents to survey.

As can be seen from Figure 3.6, the median of this distribution is at a MAP value of 0.382. The distribution is asymmetrical, and there is a long tail to the right of the median with some drivers having a very high relative likelihood of apnea, e.g., above 0.8. This distribution is different from that one sees in a population of patients seeking evaluation at a sleep disorders center. This is shown in Figure 3.7 to provide a basis of comparison. The majority of patients attending a sleep disorders center have a MAP value above 0.6.
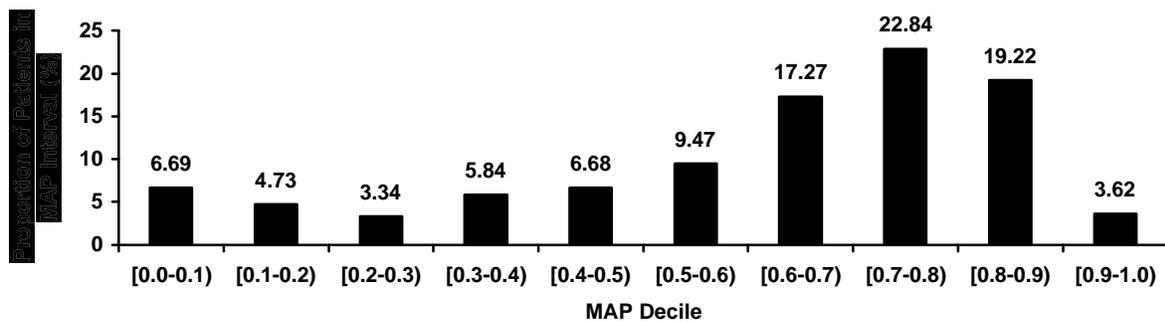
Figure 3.7. Distribution of scores of multivariable apnea prediction in a sample of 359 patients being evaluated at the Penn Center for Sleep Disorders.

3.7 **The Higher and Lower Risk Groups**

We rank ordered the MAP scores of all respondents from the lowest to highest. The plan was to recruit in descending MAP order from the higher risk until 250 in-lab assessments were performed. Recruitment of higher risk subjects ended after 247 subjects from among the top 551 scores had in-lab assessments. The minimum MAP value was 0.4356 and we defined this as the higher risk group. Thus, from the population of subjects with MAP values larger than 0.4356, we enrolled 44.8% for in-laboratory studies. The MAP scores of the higher risk group enrolled for in-laboratory studies ranged from 0.44 to 0.94 with a mean of 0.64 and a median of 0.62. The mean (SD) age of this group was 49.3 (11.6) and the mean (SD) BMI was 33.0 (5.5) kg/m^2.

The remaining 778 subjects, i.e., below MAP values of 0.4356 was defined as the lower risk group. After randomizing the order of the lower risk group, we enrolled 20.4% for in-laboratory testing (n=159). The range of MAP values in this group was 0.03 to 0.43 with a mean of 0.26 and a median of 0.30. The mean (SD) age of this group was 42.6 (9.8) and the mean BMI was 27.6 (3.6) kg/m^2. Thus, as expected, based on the risk stratification, this group was younger and less obese than the higher risk group. Tables 3.5 and 3.6 summarize demographic and physical exam characteristics of the higher and lower risk groups for continuous and categorical variables, respectively. These tables also provide population estimates taking into account the sampling scheme we employed. Table 3.5 shows, that as a consequence of risk stratification on the basis of the multivariable apnea prediction (MAP), the higher risk group had a significantly larger mean age (p<0.0001), and a significantly larger mean BMI (p<0.0001) than the lower risk group. Moreover, systolic blood pressure (p<0.0001), diastolic blood pressure (p<0.0001), and heart rate (p=0.019) all exhibited significantly larger mean values in the higher risk group compared to the lower risk group. Similarly, Table 3.6 shows that the higher risk group was more likely to be male (p=0.004). However, there was no significant difference in the racial distributions between risk groups (p=0.850). The lower risk group tended to be more highly educated (Wilcoxon rank sum p<0.001) with 13.9% completing 4 years of college compared to 6.8% in the higher risk group. Marital status also significantly differed between risk group (p=0.006); fewer of the lower risk group were married presumably as a consequence, at least in part, of their younger age. In both groups, age was significantly associated with marital status (p<0.002 for both risk strata) with single CDL holders having the lowest mean age.

Table 3.5: Demographic and physical exam characteristics of higher and lower risk subjects - continuous variables. The higher risk group, as compared to the lower risk group, was significantly older (first row), more overweight (rows 3 and 4), had higher MAP values (rows 5 and 6), had higher systolic and diastolic blood pressure (rows 7 and 8) and heart rate (row 9). Two values of MAP are shown, i.e., from initial survey (screen) and that repeated when subjects visited our sleep laboratory (in lab).

	Higher Risk					Lower Risk					Weighted		
	n	Mean	Std	Min	Max	n	Mean	Std	Min	Max	Mean[†]	SE[‡]	p value
Age (yrs)	247	49.3	11.6	26	77	159	42.6	9.8	24	75	45.4	0.55	<0.0001
Height[@] (in.)	247	70.1	2.9	60	78	159	69.6	4.1	54	79	69.8	0.20	0.094
Weight[@] (lbs.)	247	230.8	42.2	148	426	159	187.6	29.0	118	280	205.5	1.75	<0.0001
BMI (kg/m^2)	247	33.0	5.5	20.4	53.1	159	27.6	3.6	19.6	39.38	29.9	0.22	<0.0001
MAP - Screen	247	0.64	0.13	0.44	0.94	159	0.26	0.10	0.03	0.43	0.41	0.01	<0.0001
MAP – In Lab	247	0.61	0.16	0.10	0.94	159	0.32	0.14	0.04	0.73	0.44	0.01	<0.0001
Epworth sleepiness	229	9.12	4.67	0	23	157	8.54	4.14	0	20	8.8	0.23	0.2084
BP - systolic	241	139.2	17.1	100	200	157	126.9	17.2	90	180	132.0	0.92	<0.000
BP - diastolic	241	80.8	9.3	55	105	158	75.8	11.3	58	140	77.9	0.58	<0.000
Heart rate	243	77.2	11.4	48	109	157	74.9	7.5	52	104	75.8	0.46	0.019

Notes:

Std = standard deviation; min = minimum value; max = maximum value

[†] Weighted mean computed as (0.415*Higher risk mean) + (0.585*Lower risk mean)

[‡] Weighted standard error (SE) computed as the square root of $(0.415)^2$ * (Higher risk SE)^2 + $(0.585)^2$ * (Lower risk SE)^2

[@] Height and weight from physical exam

p value is for t test comparing differences between groups

Table 3.6. Demographic characteristics of higher and lower risk subjects and weighted estimate of population percentage. The higher risk group, as compared to the lower risk group, had less advanced education and more of them were married.

Characteristic	Higher Risk			Lower Risk			Weighted	
	n	Percent %	SE	n	Percent %	SE	Percent %[†]	SE[‡]
BMI Levels: Normal - <25	8	3.2%	6.2%	38	23.9%	6.9%	15.3%	4.8%
Overweight - 25-<30	73	29.6%	5.3%	83	52.2%	5.5%	42.8%	3.9%
Obesity I - 30-<35	89	36.0%	5.1%	33	20.8%	7.1%	27.1%	4.6%
Obesity II - 35-<40	51	20.6%	5.7%	5	3.1%	7.8%	10.4%	5.1%
Obesity III - >=40	26	10.5%	6.0%	0			4.4%	2.5%
Epworth > 10	77	33.6%	5.4%	50	31.8%	6.6%	32.5%	4.5%
Male Gender	240	97.2%	1.1%	144	90.6%	2.4%	93.3%	1.5%
Race White	206	83.4%	2.6%	137	86.2%	2.9%	85.0%	2.0%
African-American	35	14.2%	5.9%	18	11.3%	7.5%	12.5%	5.0%
Hispanic	5	2.0%	6.3%	3	1.9%	7.9%	1.9%	5.3%
American Indian	1	0.4%	6.3%	1	0.6%	7.7%	0.5%	5.2%
Marital Status Married	194	79.5%	2.9%	103	64.8%	4.7%	70.9%	3.0%
Single	22	9.0%	6.1%	31	19.5%	7.1%	15.1%	4.9%
Separated/Divorced	22	9.0%	6.1%	21	13.2%	7.4%	11.5%	5.0%
Widow(er)	6	2.5%	6.4%	4	2.5%	7.8%	2.5%	5.3%
Education <=Junior High	20	8.5%	6.2%	1	0.6%	7.7%	3.9%	5.2%
High school	164	69.8%	3.6%	107	67.7%	4.5%	68.6%	3.0%
2 year college	35	14.9%	6.0%	28	17.7%	7.2%	16.5%	4.9%
4 year college	11	4.7%	6.4%	13	8.2%	7.6%	6.7%	5.2%
Graduate school	5	2.1%	6.4%	9	5.7%	7.7%	4.2%	5.2%

Notes:

[†] Weighted percent computed as (0.415*Higher risk percent) + (0.585*Lower risk percent)

[‡] Weighted standard error (SE) computed as the square root of $(0.415)^2$ * (Higher risk SE)^2 + $(0.585)^2$ * (Lower risk SE)^2

Finally, in this section we summarize the overall design of our study and number of subjects at each step of our experimental protocol in Figure 3.8.

```
┌─────────────────────────────────────┐
│     Random sample of CDL holders in  │
│ Pennsylvania, within 50 miles of Philadelphia │
│              (n=4,826)               │
└─────────────────────────────────────┘

┌──────────────────────┐        ┌──────────────────────────┐
│   Could not be       │        │ Responded to questionnaire│
│ contacted by post    │        │      (n=1,391; 31.5% of   │
│ office (n=321)       │        │   potential responses (4410),│
│ Or response not      │        │      i.e., 1391/4410)     │
│ usable (n=95)        │        └──────────────────────────┘
└──────────────────────┘                              ┌──────────┐
                                                      │ Missing  │
                                                      │  MAP     │
                                                      │  n=62    │
                                                      └──────────┘
┌──────────────────────┐        ┌──────────────────────────┐
│ Potential Responses  │        │   Risk stratification by  │
│  (n=4,410; 91.4%)    │        │ multivariable apnea prediction│
└──────────────────────┘        │       (MAP) n=1329        │
                                └──────────────────────────┘

      ┌──────────────────────┐        ┌──────────────────────────┐
      │  High risk group     │        │  Low risk group (MAP      │
      │ (MAP ≥0.4356; n=551) │        │    <0.4356; n=778)        │
      └──────────────────────┘        └──────────────────────────┘

         ← 44.8% sample                    ← 20.4% sample

      ┌──────────────────────┐        ┌──────────────────────────┐
      │ In-laboratory studies│        │  In-laboratory studies    │
      │      (n=247)         │        │        (n=159)            │
      └──────────────────────┘        └──────────────────────────┘
```

Figure 3.8.   Overall sample, respondents and subject disposition status leading to in-lab studies of n=247 higher risk and n=159 lower risk holders of Pennsylvania commercial vehicle drivers licenses.

3.8 Relationship of our Sample to Other Data About Commercial Drivers

We show in Table 3.7 the gender, age distribution and, when available, the ethnic background of the following: our total population (i.e., our sampling frame); all respondents; the subjects in our higher risk group; the subjects in our lower risk; and the weighted average. The weighted average is computed from knowledge of the percentage of the population who were in the higher risk group and in the lower risk group. In order for the weighted average to provide population estimates accounting for the over-sampling of the higher risk group, higher risk percentages and mean values had to be multiplied by 0.415 and lower risk percentages had to be multiplied by 0.585. We also show for comparison the demographic information from other population studies of commercial drivers. The other information on gender and minorities was obtained from the 1997 report of the U.S. Department of Labor; the percentage of drivers under 35 years of age from a 1996 Gallup survey of 1,000 drivers; while the age distribution of drivers was obtained from a 1997 survey by the American Trucking Associations [ATA Foundation, 1997].

Table 3.7. Demographic comparisons among sample groups.

	Population (n=4826)		Respondents (n=1391)		Case Studies Higher Risk (n=247)		Case Studies Lower Risk (n=159)		Weighted*	Reference[&]
	n	%	n	%	n	%	n	%	%	%
Female	382	7.9	109	7.2	7	2.8	15	9.4	6.7	5.7
Mean Age	44.5		44.5		49.3		42.6		45.4	
Age <35 years old	1135	23.1	300	21.6	28	11.3	36	22.6	17.9	24.6
<25 years	90	1.9	24	1.7	0	0	2	1.3	0.7	
25-29 years	387	8.0	97	7.0	8	3.2	10	6.3	5.0	3.0
30-34 years	638	13.2	179	12.9	20	8.1	24	15.1	12.2	13.0
35-39 years	806	16.7	235	16.9	26	10.5	28	17.6	14.7	37.0
40-44 years	761	15.8	229	16.5	41	16.6	35	22.0	19.8	38.0
45-49 years	650	13.5	198	14.2	37	15.0	25	15.7	15.4	9.0
≥50 years	1494	<u>31.0</u>	429	<u>30.8</u>	115	<u>46.6</u>	35	<u>22.0</u>	<u>32.2</u>	<u>1.0</u>
		100		100		100		100	100	101
Caucasian	n/a		n/a		204	83.3	137	86.2	85.0	75.4
African-American	n/a		n/a		35	14.3	18	11.3	12.5	
Hispanic	n/a		n/a		5	2.0	3	1.9	1.9	
Native American	n/a		n/a		1	0.4	1	0.6	0.5	

*Weighted percent computed as (0.415 * higher risk percent) + (0.585 * lower risk percent). Weights determined at the population level.

[&]Gender and minorities, 1997 U.S. Department of Labor; Age<35 1996 Gallup; Age categories [ATA, 1997].

The percentages of drivers who were female in our total population (7.9%), all respondents (7.2%), and weighted average of those undergoing in-laboratory testing (6.7%) were close to that reported by the U.S. Department of Labor (5.7%). Likewise, the percentage of

subjects who are Caucasian in our final sample (85.0%) is not dissimilar to that reported by the U.S. Department of Labor (75.4%).

With respect to age, the percentage of our initial population who are under 35 years (23.1%) and respondents (21.6%) are also similar to that reported by the Gallup Organization (24.6%).

While these data suggest that there are major similarities between our population and the national population of commercial drivers, there are differences when we look at age distribution in more detail as reported by the survey of the American Trucking Associations (ATA) in 1997. The results of this survey are, however, different from those from the Gallup survey since only 16.0% of drivers were under 35 years of age in the ATA survey compared to 24.6% in the Gallup survey. Our results are closer to those of the Gallup survey.

The major difference between our population and that described by the ATA is in older drivers, i.e., over 50 years of age. In our global population of those surveyed (n=4826), 31% of holders of CDLs were over age 50 years of age; 30.8% of respondents were in this age group as were 32.2% of the weighted average of those studied in the laboratory of this group. In contrast, only 1% of drivers in the ATA survey were over 50 years of age. The basis for this large difference is unclear, but will be discussed further below.

3.9 Types of Occupation of the Drivers Surveyed

Since the population we surveyed is those who hold commercial drivers licenses, there is no reason to expect that all those approached were actually working at the time of the survey as commercial drivers. We show in Table 3.8 for each of the relevant populations in our study design, the percentage of CDL holders who were working full-time as commercial truck drivers, working part-time as commercial truck drivers, not currently employed as a commercial truck driver, currently employed as drivers of any commercial vehicle (trucks, busses, etc.) and not currently employed in any capacity driving any commercial vehicle.

As can be seen in Table 3.8, the vast majority of those we studied in our laboratory sample, i.e., 82.2%, were currently employed driving a commercial vehicle. 62.5% of the total laboratory sample was currently employed as full-time drivers of commercial trucks.

For those employed as commercial truck drivers, we present data on the types of driving they did, i.e., over-the-road, local, or both; their driving schedule with number of miles reported driven per year. These data are shown in Table 3.9.

There are no exactly comparable data on which to assess the type of driving done by commercial drivers in our study with what occurs nationally. The Truck Use Inventory Study does, however, give relevant data. The Truck Use Inventory Survey found that 47.3% of commercial vehicles operated within a 50-mile radius from their home base, 16.9% between 50 and 100 miles, and 10% between 100 and 200 miles. Thus, a total of 74.2% of vehicles operate within 200 miles of their base and 64.2% within 100 miles. Because the Truck Use Inventory Survey collects information on vehicles, not drivers, its data are not in the same form as ours.

Nevertheless, the observation that about two-thirds of drivers (62-68% in different parts of our sample) in our study said that they were primarily local is congruent with the observation that 64.2% of commercial vehicles operate within 100 miles of their base.

The gender, age distribution and ethnic background of these different groups of drivers are shown in Table 3.10. There were minimal ethnic differences between drivers in these different groups. The age distributions of drivers in the various groups of employment were also similar. There were small differences in the percentage who were female and, in particular, there were more females in the group with CDL but not currently employed as a commercial driver (12%).

Table 3.8. Employment status, with respect to type of commercial driving, in all respondents and those studied in the laboratory.

	Respondents (n=1391)		Case Studies Higher Risk (n=247)		Case Studies Lower Risk (n=159)		Population Estimate (Weighed Prevalence[&])	SE^
	n	%	n	%	n	%	%	%
Employment status								
Currently employed full-time as a commercial truck driver	791	56.9	151	61.1	101	63.5	62.5	3.2
Currently employed part-time as a commercial truck driver	244	14.8	48	19.4	30	18.9	19.1	4.8
Not currently employed as a commercial truck driver	356	20.4	48	19.4	28	17.6	18.4	4.8
Currently employed as commercial vehicle driver[$]	1051	75.6	200	81.0	132	83.0	82.2	2.2
Currently not employed as commercial vehicle driver	340	24.4	47	19.0	27	17.0	17.8	4.8
Type of Driving[%]								
Over the road	63	6.1	14	7.1	14	10.8	9.3	5.6
Local	707	68.2	122	62.2	88	67.7	65.4	3.4
Both	267	25.7	60	30.6	28	21.5	25.3	5.2
Driving schedule[%]								
Only days	505	48.8	88	43.6	59	44.7	44.2	4.4
Only nights	32	3.1	10	5.0	6	4.5	4.7	5.7
Both	497	48.1	104	51.5	67	50.8	51.1	4.1
Number of miles driving a truck[%]								
70k-130k	300	23.8	63	25.5	39	24.5	24.9	4.6
30k-<70k	336	26.7	77	31.2	35	22.0	25.8	4.6
15-<30k	258	20.5	41	16.6	33	20.8	19.0	4.8
<15k	366	29.0	66	26.7	52	32.7	30.2	4.4

[&] Weighted percent computed as (0.415*Higher risk percent) + (0.585*Lower risk percent). Weights determined at the population level.

[$] Includes respondents indicating they drove a truck or a bus either full-time or part-time.

^ Weighted standard error (SE) computed as square root of $(0.415)^2$ * (Higher risk SE)^2 + $(0.585)^2$ * (Lower risk SE)^2

Table 3.9. Types of driving for individuals employed as commercial truck drivers in all respondents and those studied in the laboratory.

	Respondents Employed as Commercial Driver[$] (n=1035)		Higher Risk Employed as Commercial Driver[$] (n=199)		Lower Risk Employed as Commercial Driver[$] (n=131)		Weighted Average[†]	SE[‡]
	n	%	n	%	n	%	%	%
Type of Driving								
Over the road	62	6.0	14	7.1	14	10.8	9.3	5.6
Local	700	68.0	122	62.2	88	67.7	65.4	3.4
Both	267	25.9	60	30.6	28	21.5	25.3	5.2
Driving schedule								
Only days	500	48.7	86	43.9	59	45.4	44.8	4.4
Only nights	31	3.0	10	5.1	6	4.6	4.8	5.8
Both	496	48.3	100	51.0	65	50.0	50.4	4.2
Number of miles driving a truck								
70k-130k	253	26.1	56	30.0	35	27.3	28.4	5.1
30k-<70k	291	30.0	68	36.4	34	26.6	30.6	5.0
15-<30k	205	21.2	31	16.6	30	23.4	20.6	5.3
<15k	220	22.7	32	17.1	29	22.7	20.4	5.3

[$] Includes respondents indicating they drove a truck either full-time or part-time.
[†] Weighted percent computed as (0.415*Higher risk percent) + (0.585*Lower risk percent).
[‡] Weighted standard error (SE) computed as the square root of $(0.415)^2$ * (Higher risk SE)^2 + $(0.585)^2$ * (Lower risk SE)^2.

Table 3.10. Age, BMI, gender and age distribution and ethnic background of subjects who indicated that they were or were not full-time drivers doing at least 30,000 miles/year and who indicated that they were or were not full- or part-time drivers doing at least 15,000 miles/year.

	All Case Studies (n=406)	Full-time drivers with >=30k miles/year (wt=167.9)	Less than full-time drivers or <30k miles/year (wt=238.0)	Full- or part-time drivers with >=15k miles/year (wt=253.6)	Not currently employed or <15k miles/year No (wt=152.4)
	Mean[&]	Mean[&]	Mean[&]	Mean[&]	Mean[&]
Age	45.4	43.7	46.6	44.1	47.5
BMI	29.9	30.5	29.4	30.1	29.4
	Percent[&]	Percent[&]	Percent[&]	Percent[&]	Percent[&]
Female gender	6.7%	2.6%	9.6%	3.5%	12.0%
<25 years	0.7%	0%	1.3%	0%	2.0%
25-29 years	4.7%	3.1%	5.8%	4.9%	4.2%
30-34 years	11.8%	16.9%	8.2%	15.0%	6.6%
35-39 years	15.2%	16.4%	14.4%	15.1%	15.4
40-44 years	19.4%	23.2%	16.7%	22.2%	14.7%
44-49 years	15.2%	13.1%	16.3%	12.5%	19.9%
>=50 years	32.9%	27.3%	36.9%	30.4%	37.2%
Ethnicity					
White	85.0%	87.1%	83.5%	88.1%	79.8%
African-American	12.5%	10.7%	13.8%	10.1%	16.4%
Hispanic	1.9%	0.9%	2.7%	0.9%	3.8%
Native American	0.5%	1.3%	0%	0.9%	0%

[&] Mean and percentages computed from weighted sample

3.10 Discussion

In conclusion, we employed a now standard two-stage sampling strategy to investigate the prevalence of sleep apnea [Gislason et al, 1988; Young et al, 1993]. The population we identified as our sampling frame was based on a random sample of holders of commercial drivers licenses living in Pennsylvania and residing within 50 miles of the Sleep Center at the University of Pennsylvania. While this has the advantage of being a random sample with a precisely defined sampling frame, this approach has a potential disadvantage. Since we recruited these subjects by mail, we only had a 31.5% response rate.

This response rate needs to be seen in context [Young, 1999]. To quote: "If respondents differ from nonrespondents on characteristics that correlate with the prevalence of sleep apnea, the crude prevalence will not be correct. However, a low response rate does not necessarily mean there is a selection bias with respect to the probability of having a particular condition (e.g., sleep apnea). The following points temper the degree to which the survey response rate at

issue is an incapacitating limitation. Numerous methodological studies have shown that higher response is associated with higher education, female gender, health consciousness, non-smoking, and better physical and psychological health. It follows that very high response may be expected from a sampling frame of registered nurses (i.e., the Harvard Nurses Health Study), while a sampling frame for a profession that is, as a group, lower on the characteristics related to good response, would result in lower response. The sample at issue [i.e., commercial truck drivers] could not be a better example of a group for which a low response would be expected. Obtaining a response of 30% from a sampling frame of nurses would be alarming, and selection bias other than the "healthy volunteer characteristics" would be suspect. The rate in the CDL sample is not surprising. Consideration of the numeric "%" in the response is not the crucial measure of potential bias: whether the respondents differ from the entire sample frame on factors relevant to the study aims is the crucial factor." Importantly, there were no differences in gender distribution between non-respondents and respondents and the age distribution of these two groups was essentially identical (see Figure 3.4).

While there is, therefore, no evidence of response bias on the basis of available data, it is conceivable that there are other unknown factors that determined response, e.g., respondents could have had higher BMIs and hence more likely to have apnea, i.e., hidden bias [Rosenbaum, 1991, 1995]. Our study design does not allow us to exclude this possibility but the nearly identical age and gender distributions make this, in our view, very unlikely. We further report in our prevalence estimates of sleep apnea (Chapter Five), the relationship between prevalence and the major determinants of apnea—age, gender, BMI, as well as the multivariable apnea index (MAP) [Maislin et al, 1995]. This will allow others to extrapolate our findings on prevalence to other populations, if these characteristics of the population are known. It is important to note that this issue of potential response bias does not affect the other components of our study, i.e., the major goal of determining the relationship between respiratory disturbance index during sleep and performance decrements that is the subject of Chapter Seven.

CHAPTER FOUR

Sleep Duration in Commercial Drivers

4.1 Introduction and Techniques Employed

It has previously been reported that commercial drivers have short sleep durations. This result came from the large study of drowsy driving in commercial drivers with different schedules [Mitler et al, 1997]. Drivers were monitored for one week during which they had their daily sleep in a sleep laboratory. The average nightly sleep was short—an average 4-5 hours—but the explanation for this short sleep duration was not elucidated. It may, at least in part, have resulted from the drivers having to sleep in sleep laboratories and not in their natural environment.

In the study we performed, we monitored sleep duration at home for seven consecutive days using an actigraph. This small device detects movements and gives reasonable estimates of sleep and wakefulness. Usable actigraphy data were collected for 208 of 247 (84.2%) higher risk subjects and 132 of 159 (83.0%) lower risk subjects. There was no significant difference in the percentages of subjects in the higher risk and lower risk groups in whom usable data were obtained (p=0.751). Reasons for missing data included the following: data were lost due to equipment malfunction (n=50); subjects did not wear actigraph (n=11); subjects removed batteries, thereby disabling instrument (n=1); actigraph delivered to wrong address (n=1); and actigraph not given to subject (n=3). Mantel-Haenszel stratified analyses [1959] were used to compare the prevalence of sleep apnea at various severity levels between subjects with and without actigraphy data controlling for risk group. There were no significant differences between these groups in the percentages of subjects with AHI$\geq$5 (p=0.948), AHI$\geq$15 (p=0.557), or AHI$\geq$30 (p=0.160) events per hour. Similarly, these groups did not significantly differ with regard to gender (p=0.698). Two-way analysis of variance was used to compare subjects with and without actigraphy data controlling for risk group with regard to age and BMI. Adjusted mean ages among subjects with missing and not missing actigraphy data were 43.2 and 46.5 years, respectively with p=0.023. There was also a slight difference with respect to mean BMI. After controlling for risk group, the adjusted mean BMI values were 31.9 and 30.7 kg/m^2, respectively, among subjects with missing and not missing actigraphy data (with p=0.057). In conclusion, we found that subjects missing actigraphy data were similar to those not missing actigraphy data in terms of sleep apnea prevalence and gender, but tended to be slightly younger and slightly more obese.

The actigraphy device is a small instrument, roughly the size of a wristwatch, and is worn on the wrist. It detects rotated movements of the wrist using a sensitive motion sensor. We set the length of time for the epochs it monitored movements to 1 minute. The instrument stores the number of movements in each such epoch in its memory. Following the week of recording, the instrument was returned to the laboratory and the data on the computer memory in the actigraph were transferred into the laboratory's computer systems. Specialized software (Action-W) [Ambulatory Monitoring, Inc., 1996] was used to analyze the movement data. In parallel with actigraphy, the subjects also filled out a sleep diary twice per day. This provides information

about when the subjects reported they went to bed and woke up etc. as well as about the quality of their sleep. A page from the diary is shown in Appendix G.

We utilize these self-reports in the diary, together with data from actigraphy, to estimate sleep duration. Actigraphy provides reasonable estimates of sleep duration in normal individuals as compared to sleep measured by the electroencephalogram, etc. [Blood et al, 1997; Jean-Louis, et al, 1997]. While this is true in normals, there are problems in individuals with sleep apnea since such individuals have repeated interruption of sleep and may show movements, i.e., activity during sleep [Middelkoop et al, 1995].

We show in Figure 4.1 the actigraph recordings for a 24-hour period for a normal commercial driver (bottom panel) and for one with severe sleep apnea (apnea/hypopnea index = 71.2 episodes/hour) (top panel). The major sleep period is clearly visible as marked. In the subject with sleep apnea, there were more movements during the sleep period that were recorded on the actigraph. This has been described previously [Middelkoop et al, 1995], and is a result of movements occurring when the subject arouses at the end of their apneic episodes. This makes it difficult to estimate actual sleep durations in such individuals.
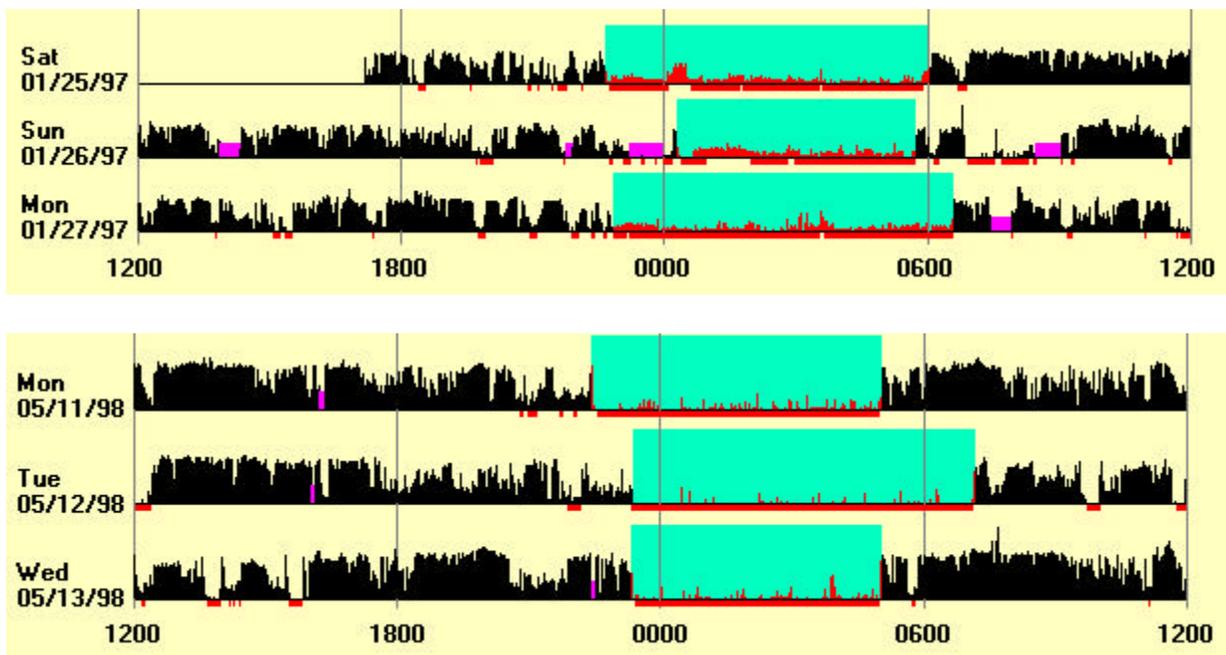
Figure 4.1.   Example of recorded movements (y axis) across three consecutive 24-hour periods from noon to noon in a 57 year old male subject with severe sleep apnea (ID=33, AHI=71.2 episodes/hour, top panel) and a 42 year old male normal subject (ID=344, AHI=1.2 episodes/hours, bottom panel). Both subjects were employed full-time as commercial truck drivers. The sleep periods are marked in gray. Sleep durations in these examples were 336, 227, 429 minutes in the subject with severe sleep apnea and 386, 465, 331 minutes in the normal subject. There was evidence in the apneic (top panel) of continuous movements throughout the sleep period, i.e., sleep was extremely disturbed.

To address this problem we have taken the following approaches. First, we have defined as a variable *main bout length of relative inactivity*. This is from the beginning of the period of relative inactivity to its end. There may be activity (movements) between these start and end-points. In some individuals such activity may represent periods of wakefulness with movement. In other individuals it will represent movement arousals occurring during sleep, i.e., not during wakefulness. Since it is difficult, if not impossible, to determine whether periods of movements during sleep represent wakefulness or movement arousals during sleep, we have decided not to try to distinguish them. We hence report as a variable "bout length of relative inactivity" and examine its determinants. We appreciate that this will be potentially an overestimate of actual sleep. Hence, in addition, we define another variable as the cumulative duration of periods of inactivity during the major bout. We call this variable *cumulative duration of inactivity in main sleep bout*. In some individuals with movements during sleep such as those with sleep apnea, we know that this variable will likely underestimate actual sleep duration. We also report the value of this variable in our population and its determinants. We argue that if analysis of both variables leads to similar conclusions, then discussion about whether our approach underestimates or overestimates sleep is mute. This is particularly important in the analysis represented in Chapter Five on determinants of sleep apnea prevalence and in Chapter Seven on determinants of daytime performance. Cumulative duration of inactivity will be a much closer approximation to sleep duration in individuals without sleep apnea who lack movement arousals. We take advantage of this by doing, throughout this report, secondary analyses of determinants of the cumulative period of inactivity in those subjects who did not have sleep apnea, i.e., had an apnea/hypopnea index of less than 5 episodes/hour.

We formally analyzed the potential extent to which the presence of sleep apnea events results in a downward bias in the cumulative duration of inactivity in the main sleep bout. To this end, we computed the mean difference between the main bout length of relative inactivity and the cumulative duration of inactivity during the main sleep bout. The mean (SD) differences were 148.1 (90.4), 95.7 (68.2), 69.6 (50.0), and 55.8 (47.2) minutes, respectively, for subjects with severe (AHI>30/hr), moderate (15-<30), mild (5-<15), and 'insignificant' (<5/hr) sleep apnea. A two-way analysis of variance was used to determine if these differences remained significant controlling for risk group. Least squares risk group adjusted mean differences were 142.0, 90.0, 66.2, and 56.1 minutes, for the above AHI categories, respectively. Overall differences (F=16.8, df=3,333, p<0.0001) as well as the linear trend (F=49.8, df=1,33, p<0.0001) were highly statistically significant. Furthermore, the partial Pearson correlation (controlling for risk group) between AHI and cumulative durations of inactivity in main sleep bout was r=-0.377 (p<0.0001). In contrast, the partial correlation between AHI and the main bout lengths of relative inactivity was only r_s=-0.143 (p=0.001). This is because the partial correlation between AHI and the difference was found to r=+0.373 (p<0.0001). In other words, the expected 'discrepancy' between the main bout lengths of relative inactivity and the cumulative durations of inactivity during the main sleep bout was found to be linearly related to severity of sleep apnea. Such a relationship is consistent with the hypothesis that actigraphically determined cumulative duration of inactivity contains wakefulness artifacts caused by movements during sleep associated respiratory events. An alternative explanation is that increasing apnea severity is truly associated with increased wakefulness after sleep onset. As noted above, definitive discrimination is not possible within the context of this study. Therefore, our analyses of sleep duration utilize three definitions, the main bout length of relative inactivity, the cumulative

duration of inactivity during the main sleep, and the cumulative duration of inactivity during the main sleep bout restricted to the subsample with less than mild apnea (AHI<5/hr).

This discussion highlights the difficulty in estimating the duration of the major sleep bout in individuals with sleep apnea. We propose that this problem is even more pronounced when trying to estimate the duration of shorter periods of sleep, i.e., naps. Hence, we limit our analysis to reporting estimates of the major sleep bout.

4.2 Main Bout Length of Relative Inactivity

The techniques described above allow us to examine the duration of the main bout length of relative inactivity in all of the subjects. There were no significant differences in mean main bout length of relative activity between subjects in the higher risk group and those in the lower risk group (t=-1.46, df=338, p=0.145). The mean main bout length of relative inactivity was 436.9±62.6 minutes (mean±SD) in the higher risk group (n=208) and 436.8±58.6 (mean±SD) in the lower risk group (n=132). Based on weights determined using respondents from the population survey, the average weighted duration of the main bout length of relative inactivity is 442.7 minutes, i.e., 7.4 hours/night. Since, as discussed above, this is likely to be an overestimate of actual sleep duration, we do not find, however, on average, the extremely short durations reported by Mitler et al [1997].

While the average is more than 7 hours, there were substantial differences in the duration of the main bout length of relative inactivity ranging from less than 201 minutes (3.4 hours) to 606 minutes (just over 10 hours). We show in Figure 4.2 the percentages of subjects in both groups (higher risk and lower risk) with different durations of the main bout length of relative inactivity. Based on the empirical distribution we defined the following categories for mean duration of the main bout length of relative inactivity: <6 hours (total n=33), 6 to less than 7 (total n=77) hours, 7 to 8 hours (total n=151), and more than 8 hours (total n=79). In subsequent regression analyses, the category >8 hours was chosen as reference.
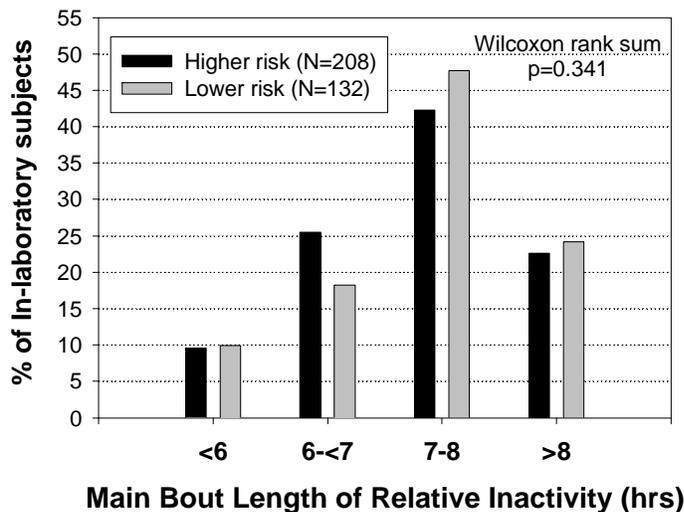
Figure 4.2. Percentages of subjects in the higher risk and lower risk groups with different categories of the average durations of the main bout lengths of relative inactivity over the 7-day monitoring period. There was no significant difference (p=0.341) in these distributions of categories between the higher and lower risk groups.

4.3 Cumulative Duration of Inactivity During the Major Sleep Bout

Because the duration of the *main bout length of relative inactivity* is likely to overestimate mean sleep duration, we now turn our attention to the mean *cumulative duration of inactivity during the major sleep bout* in all of the subjects. In contrast to the results from main bout length or relative inactivity, subjects in the higher risk group, on average, had significantly smaller mean cumulative durations compared to those in the lower risk group (p<0.0001). The average cumulative duration of inactivity was 358.5±84.4 minutes (mean±SD) in the higher risk group (n=208) and 395.2±74.1 384.0±90.6 (mean±SD) in the lower risk group (n=132). Based on weights determined using respondents from the population survey, the average weighted cumulative duration of inactivity is 380.0 minutes, i.e., 6 and 1/3 hours/night. Even though this is likely to be an underestimate of actual sleep duration, we still do not find, on average, the extremely short durations reported by Mitler et al [1997].

While the average is more than 6 hours, there were substantial differences in the amount of cumulative durations of inactivity between individuals ranging from less than 4 hours to 553 minutes (just over 9 hours). We show in Figure 4.3 the percentages of subjects in both groups (higher risk and lower risk) with different cumulative durations of inactivity in the main sleep bouts.
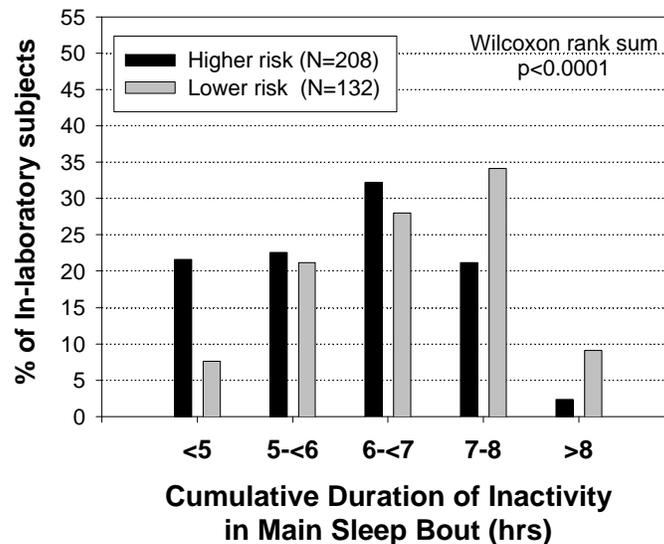
Figure 4.3.   Percentages of subjects in the higher risk and lower risk groups with different categories of average cumulative duration of inactivity over the 7-day monitoring period.   There is a significant difference (p<0.0001) in the category distributions between the higher and lower risk groups.

Again, in contrast to the mean *main bout length of relative inactivity category*, the distribution of the cumulative duration of inactivity categories was significantly shifted toward lower values in the higher risk group compared to the lower risk group (Wilcoxon rank sum p<0.0001). Because of the relatively small numbers of subjects in the >8 hours category, regression models utilizing cumulative duration of inactivity in main sleep bout category defined 7-8 hours as the reference category.

4.4 <u>**Cumulative Duration of Inactivity Among Subjects with RDI<5 Events Per Hour**</u>

The third method used to assess the duration of sleep using home actigraphy involved the cumulative duration of inactivity in main sleep bout but restricted attention to the non-apneic subjects (RDI<5 events/hr). This approach was intended to avoid potential bias from movements associated with apnea that are defined as wakefulness by actigraphy. There were 102 subjects (49.3%) in the higher risk group and 114 subjects (87.0%) in the lower risk group with an apnea hypopnea index less than 5 events per hour.

In the subsamples of subjects with AHI<5/hr, mean values of cumulative duration of inactivity were no longer significantly different between the risk groups (t=-1.58, df=214, p=0.115). The average cumulative duration of inactivity was 381.1±67.8 minutes (mean±SD) in the higher risk group (n=102) and 396.7±75.8 (mean±SD) in the lower risk group (n=114). Based on weights determined using respondents from the population survey, the average weighted cumulative duration of inactivity among those with AHI<5/hr is 390.2 minutes, i.e., 6 and 1/2 hours/night.

The frequency distributions of cumulative duration of inactivity categories restricted to subjects with no significant sleep apnea (i.e., AHI<5/hr) are illustrated in Figure 4.4. The difference between risk groups in their cumulative duration of inactivity categories retained statistical significance (Wilcoxon rank sum=0.025). Even after excluding subjects with AHI>=5/hr, the risk groups still significantly differed with respect to age, BMI, frequency of snoring, and gender. Of these variables, gender appeared to be the most important in explaining this risk group difference. Eleven of the 13 female subjects were in the lower risk group.
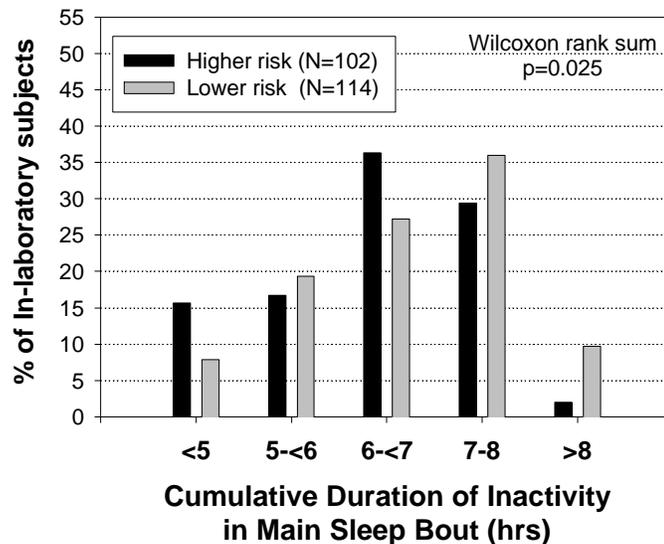
Figure 4.4. Percentages of subjects in the higher risk and lower risk groups with different categories of average cumulative duration of inactivity over the 7-day monitoring period in the subsamples of subjects with AHI less than 5 events per hour. There remained a significant difference (p=0.025) in the category distributions between the higher and lower risk groups largely due to the greater number of females in the lower risk group (11 compared to 2).

4.5 Between Subject Variation

It is anticipated that performance will also be affected by variations in nightly sleep duration. We therefore examined the variation of cumulative periods of inactivity across the 7 days of measurement. Some subjects showed highly variable durations across the 7-day period (e.g., subjects ID240, ID143, Figure 4.5) while others showed more stable durations for this variable (e.g., subject ID215, Figure 4.5). AHI for subjects ID240, ID215, ID143, were 11.0, 25.1, and 1.8 events per hour, respectively. Mean (SD) main bout length of relative inactivity for ID240, ID215, ID143 were 328 (156.7), 402.1 (73.2), and 470.7 (148.1). Thus, for these three subjects, the differences between mean main bout lengths of relative inactivity and mean cumulative durations of relative inactivity were similar.
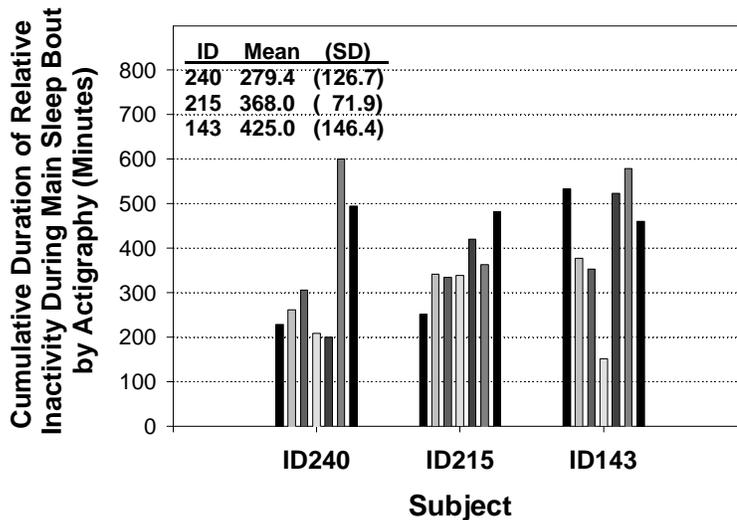
Figure 4.5.   Day-to-day variation in measured duration of cumulative periods of inactivity. Two subjects (ID240, ID143) show highly variable daily patterns, while subject ID215 has a more stable duration of this variable.

To characterize this across our population, we estimated for each subject, two range statistics. The first range statistic is expected to cover the middle 50% of their average daily *main bout length of relative inactivity.* The second range statistic is expected to cover the middle 50% of their daily *cumulative duration of inactivity during the major sleep bout*. Both intervals were computed as a subject-specific mean ± 0.6745 times a subject-specific standard deviation over the seven days. The value of 0.6745 is from the standard normal distribution and corresponds to the 75th percentile value. We show in Figures 4.6a and 4.6b, plots of these "middle 50% ranges" versus mean daily values by risk group.

There were significant negative correlations between variability and mean values for both risk groups using both definitions. Specifically, shorter durations were associated with more variable durations (see Figures 4.6a and 4.6b). For main bout length of relative inactivity, Spearman rank correlations were r_s=-0.24 (p=0.0004) and r_s=-0.31 (p=0.0003), respectively, for subjects in the higher and lower risk groups (Figure 4.6a). Similarly for cumulative duration of inactivity during the major sleep bout, Spearman rank correlations were r_s=-0.34 and r_s=-0.43, respectively (p<0.0001), for subjects in the higher and lower risk groups (Figure 4.6b). Thus,

using both definitions, longer mean daily durations were associated with greater stability. Conversely, subjects with lower average durations had more variable durations across the 7 nights of study. This might reflect more variable durations of sleep and/or variable amounts of movement arousal within sleep. We then defined a stable pattern as one where the middle 50% range was less than or equal to 60 minutes. For main bout length of relative inactivity, in the two groups, 16.4% and 15.9%, respectively, had 'stable' sleep patterns (χ^2=0.01, df=1, p=0.915). Similarly, for cumulative duration of inactivity during the major sleep, 17.8% and 14.4%, respectively, appeared to have 'stable' sleep patterns. These percentages did not significantly differ (χ^2=0.67, df=1, p=0.412) between groups.
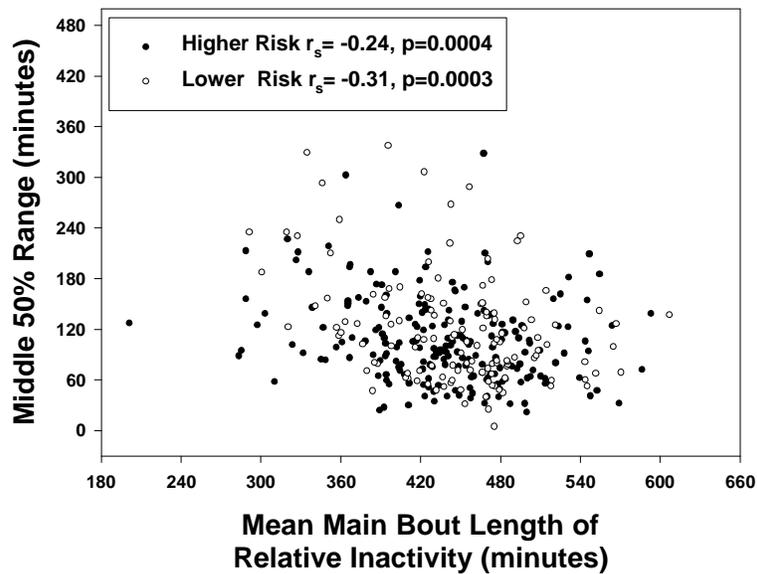
Figure 4.6a. Scatterplot of interval lengths necessary to cover 50% of subject-specific mean main bout length of relative inactivity versus mean main bout length of relative inactivity for subjects in both the higher risk and lower risk groups measured across 7 days. Shorter mean durations are associated with more variable durations.
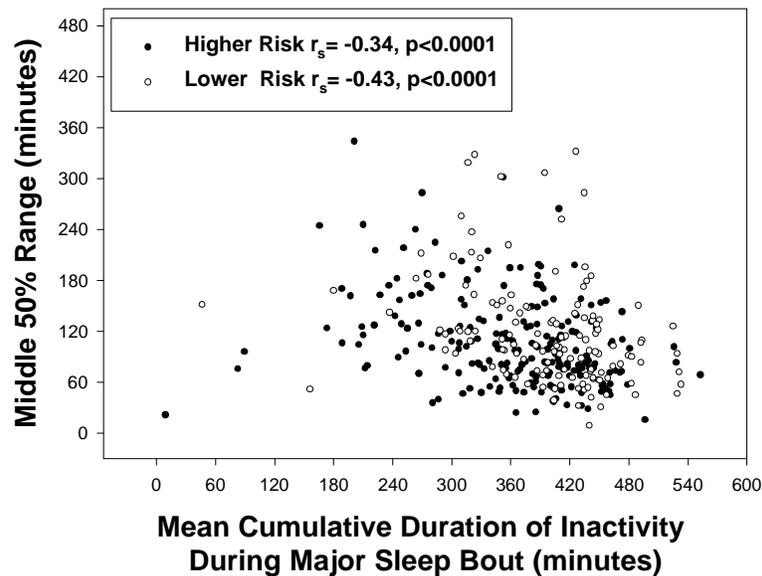
Figure 4.6b. Scatterplot of interval lengths necessary to cover 50% of subject-specific mean cumulative durations of inactivity versus mean cumulative durations for subjects in both the higher risk and lower risk groups measured across 7 days. Again, shorter mean durations are associated with more variable durations.

4.6 <u>Onset Time and Offset Time for Major Sleep Bout</u>

There were not only differences between individuals in durations of the variables we estimated, but also in the time individuals initiated sleep, and when they woke up. We show in Table 4.1 the percentages of individuals in both groups who initiated sleep on average at different times and when they terminated sleep.

As can be seen from Table 4.1, the majority of subjects in this study (54.3%) initiated their major sleep bout between 10:00 PM and midnight. However, approximately 10% (9.2%) initiated sleep between 1:00 AM and 4:00 AM. The data on time to terminate the major sleep bout is more concerning. 35.6% of all subjects studied terminated their sleep before 6:00 AM while 63.3% did so before 7:00 AM. It seems likely that such early morning awakening will lead to reduction in sleep duration.

Table 4.1. Frequency distributions for higher and lower risk subjects and weighted population estimates of mean times of sleep initiation and termination. For each subject, the mean value was computed from 1 week of actigraphy.

	Higher Risk			Lower Risk			Weighted Ave.	
	n	%	SE	n	%	SE	%[†]	SE[‡]
Initiate Sleep								
7:00 PM to 7:59 PM	6	2.9%	1.2%	1	0.7%	0.6%	1.6%	0.6%
8:00 PM to 8:59 PM	11	5.3%	1.6%	7	5.2%	1.5%	5.3%	1.1%
9:00 PM to 9:59 PM	33	15.9%	2.5%	20	14.9%	2.5%	15.3%	1.8%
10:00 PM to 10:59 PM	54	26.0%	3.0%	33	24.6%	3.0%	25.2%	2.2%
11:00 PM to 11:59 PM	52	25.0%	3.0%	43	32.1%	3.2%	29.1%	2.3%
12:00 AM to 12:59 AM	28	13.5%	2.4%	13	9.7%	2.1%	11.3%	1.6%
1:00 AM to 1:59 AM	10	4.8%	1.5%	8	6.0%	1.6%	5.5%	1.1%
2:00 AM to 2:59 AM	7	3.4%	1.3%	4	3.0%	1.2%	3.1%	0.9%
3:00 AM to 3:59 AM	1	0.5%	0.5%	1	0.7%	0.6%	0.6%	0.4%
Other	6	2.9%	1.2%	2	1.5%	0.8%	2.1%	0.7%
Terminate Sleep								
3:00 AM to 3:59 AM	11	5.3%	1.6%	4	3.0%	1.2%	3.9%	0.9%
4:00 AM to 4:59 AM	20	9.6%	2.0%	10	7.5%	1.8%	8.4%	1.4%
5:00 AM to 5:59 AM	60	28.8%	3.1%	26	19.4%	2.7%	23.3%	2.1%
6:00 AM to 6:59 AM	47	22.6%	2.9%	42	31.3%	3.2%	27.7%	2.2%
7:00 AM to 7:59 AM	32	15.4%	2.5%	29	21.6%	2.9%	19.0%	2.0%
8:00 AM to 8:59 AM	17	8.2%	1.9%	10	7.5%	1.8%	7.8%	1.3%
9:00 AM to 9:59 AM	8	3.8%	1.3%	5	3.7%	1.3%	3.8%	0.9%
10:00 AM to 10:59 AM	2	1.0%	0.7%	2	1.5%	0.8%	1.3%	0.6%
11:00 AM to 11:59 AM	1	0.5%	0.5%	1	0.7%	0.6%	0.6%	0.4%
Other	10	4.8%	1.5%	3	2.2%	1.0%	3.3%	0.9%

Notes:
[†] Weighted percent computed as (0.415*Higher risk percent) + (0.585*Lower risk percent)
[‡] Weighted standard error (SE) computed as the square root of $(0.415)^2 *$ (Higher risk SE)^2 + $(0.585)^2 *$ (Lower risk SE)^2

Mean main bout length of relative inactivity±SD (hours) for subjects terminating sleep at various times were 3:00 AM (6.6±0.7), 4:00 AM (7.0±1.1), 5:00 AM (7.1±0.9), 6:00 AM (7.5±0.9), 7:00 AM (7.7±0.9), 8:00 AM (7.8±0.9), and 9:00 AM (7.9±1.0). Two-way ANOVA controlling for risk group demonstrated that the linear trend reflecting reduced mean bout lengths of relative activity for earlier awakenings was highly statistically significant (F=26.6, df=1,313, p<0.001). Thus, we see a strong linear association between sleep termination time and mean main bout length of relative inactivity. Similarly, there was a statistically significant linear trend in mean main bout length of relative inactivity as the times of sleep initiation became earlier from 2:00 AM to 8:00 PM (F=6.6, df=1,315, p=0.011). Mean main bout length of relative inactivity±SD (hours) for subjects initiating sleep at these times were 8:00 PM (7.4±1.7), 9:00 PM (7.5±0.9), 10:00 PM (7.7±0.8), 11:00 PM (7.2±1.0), midnight (7.1±0.8), 1:00 AM (6.9±1.4),

and 2:00 AM (6.8±0.8). Thus, both earlier times of sleep termination and later time of sleep initiations contributed to shorter mean main bout lengths of relative inactivity.

The above analysis was repeated for mean cumulative durations of inactivity during the main sleep bout. Mean sleep±SD (hours) cumulative durations of inactivity for subjects terminating sleep at various times were 3:00 AM (5.4±0.8), 4:00 AM (5.6±1.5), 5:00 AM (6.1±1.4), 6:00 AM (6.3±1.3), 7:00 AM (6.8±1.1), 8:00 AM (6.4±1.4), and 9:00 AM (6.9±1.0). Two-way ANOVA controlling for risk group demonstrated that the linear trend reflecting reduced mean sleep durations for earlier awakenings was highly statistically significant (F=16.0, df=1,313, p<0.001). Thus, the strong linear association with mean time of sleep termination observed for mean main bout length of relative inactivity was maintained for mean cumulative duration of inactivity during the main sleep bout. In contrast, there did not appear to be a statistically significant linear trend in mean cumulative durations of inactivity as the times of sleep initiation became earlier from 2:00 AM to 8:00 PM (F=1.3, df=1,315, p=0.248) although overall, significant differences were observed (p=0.007). Mean sleep±SD (hours) cumulative durations of inactivity for subjects initiating sleep at these times were 8:00 PM (6.1±1.6), 9:00 PM (6.3±1.5), 10:00 PM (6.6±1.1), 11:00 PM (6.2±1.4), midnight (5.7±1.3), 1:00 AM (5.9±1.8), and 2:00 AM (6.1±0.8). Thus, we see a consistent linear trend for time of sleep termination but not for time of sleep initiation for mean cumulative duration of inactivity in the main sleep bout.

4.7 Determinants of "Sleep" Durations

To evaluate the role of different factors in determining sleep duration, we performed three multiple linear regression analyses. The first defined the dependent variable as the mean main bout length of relative inactivity. The second defined the dependent variable as the cumulative duration of inactivity during the main sleep bout. The third also defined the dependent variable as the cumulative bout length of relative inactivity but restricted attention to those without significant sleep apnea (AHI<5/hr). The results are shown in Tables 4.2, 4.3, and 4.4 for these models, respectively. These tables contains the parameter (i.e., slope) estimates of multiple linear regression models for either mean bout length or mean cumulative duration of inactivity as a function of age, BMI, gender, employment status, miles driven per year, apnea severity category (the latter is only shown in Tables 4.2 and 4.3), whether or not sleep initiation time usually took place between 8:00 PM and midnight, and whether or not sleep termination usually was, or was not, prior to 6:00 AM. Age and BMI were entered into the model after subtracting their mean values in order to make the intercept interpretable. Thus, e.g., in Table 4.3, the intercept value of 445 minutes (7.4 hours; top row) is the predicted mean cumulative duration of periods of inactivity for a 45.4 year old male with a BMI equal to 29.9 kg/m^2 who is currently not employed as a commercial driver, has an AHI<5 episodes per hour, goes to sleep on average between 8:00 PM and midnight and wakes up on average between 7:00 AM and 10:00 AM.

In Table 4.2, we find that the effect of full-time employment is to reduce the predicted main bout length of relative inactivity by 11 minutes. The predicted value for females is approximately 38 minutes longer than for males. There appeared to be no association with AHI category controlling for the other variables in the model (F=0.67, df=3,326, p=0.574). The effect of not initiating sleep between 8:00 PM and midnight and not terminating sleep between 7:00

AM and 10:00 AM were both to decrease expected durations by approximately 0.75 hours, i.e., 45 minutes. The multiple linear regression was estimated using sampling weights to account for the stratified sample. The proportion of total variance explained by the model was 23.2%. The standard deviation around the predicted values was 53.7 minutes.

In Table 4.3, we find that the effect of full-time employment is to reduce the predicted average cumulative duration of inactivity during the main sleep bout by 19 minutes; the predicted value for females is approximately 46 minutes longer than for males; drivers reporting more than 100,000 miles per year have predicted cumulative durations of inactivity that are 18 minutes longer; subjects with AHI values >30 episodes per hour (severe sleep apnea) are predicted to have average cumulative durations of inactivity that are much shorter, i.e., 74 minutes. This likely reflects the movement arousals occurring during sleep in such subjects. The effect of not initiating sleep between 8:00 PM and midnight and not terminating sleep between 7:00 AM and 10:00 AM were both to decrease expected durations by almost one hour. The multiple linear regression was estimated using sampling weights to account for the stratified sample. The proportion of total variance explained by the model was 33.1%. The standard deviation around the predicted values was 66.9 minutes. Thus, all of these factors play an important role in determining mean cumulative durations of inactivity during main sleep bouts.

Table 4.2. Variables Determining Main Bout Length of Relative Inactivity

Factor	Parameter Estimate	SE	p-value
Intercept	493.67	9.10	
Age - 45.4	-0.155	0.272	0.569
BMI - 29.9	-0.365	0.642	0.570
Female vs. male	37.74	12.51	0.003
Full vs. unemp	-11.38	8.36	0.175
Part vs. unemp	-8.61	9.55	0.168
Miles/yr>=100k	6.11	8.41	0.468
5≤AHI<15 vs. AHI<5	0.08	8.1	0.993
15≤AHI<30 vs. AHI<5	-17.24	13.06	0.188
AHI≥30 vs. AHI<5	-8.82	15.66	0.574
Begin sleep not 8pm-12	-45.94	7.53	<0.0001
Terminated sleep not 7am-10am	-48.91	7.03	<0.0001

Results of multiple linear regression to evaluate variables that play a role in determining mean main bout length of relative inactivity. Intercept (row one) is the predicted main bout length of relative inactivity for a male 45.4 years old, with a BMI of 29.9 kg/m^2, who is not currently employed as a commercial vehicle driver, has an AHI<5/hr, initiates sleep between 8 PM and midnight on average, and who terminates sleep between 7 AM and 10 AM, on average. The parameter estimates in rows 4 to 12 indicate the differences in main bout length of relative inactivity for the groups shown. The results of this analysis are described more fully in the text.

The results for cumulative duration of inactivity during the main sleep bout were similar when restricting attention to the subjects with AHI<5/hr (Table 4.4). The effect of full-time

employment is to reduce the predicted average cumulative duration of inactivity during the main sleep bout by 17 minutes; the predicted value for females is approximately 41 minutes longer than for males. The effect of not initiating sleep between 8:00 PM and midnight and not terminating sleep between 7:00 AM and 10:00 AM were both to decrease expected durations by slightly more that one hour. The proportion of total variance explained by the model was 26.8%. The standard deviation around the predicted values was 67.8 minutes.

Table 4.3. Variables Determining Cumulative Duration of Inactivity

Factor	Parameter Estimate	SE	p-value
Intercept	445.39	11.34	
Age - 45.4	-0.520	0.339	0.126
BMI - 29.9	-3.670	0.800	<0.0001
Female vs. male	45.70	15.59	0.004
Full vs. unemp	-18.86	10.42	0.071
Part vs. unemp	-16.44	11.89	0.168
Miles/yr>=100k	18.29	10.47	0.082
5≤AHI<15 vs. AHI<5	0.60	10.00	0.953
15≤AHI<30 vs. AHI<5	-37.15	16.27	0.023
AHI≥30 vs. AHI<5	-73.79	19.50	0.0002
Begin sleep not 8pm-12	-54.62	9.37	<0.0001
Terminated sleep not 7am-10am	-56.56	8.76	<0.0001

Results of multiple linear regression to evaluate variables that play a role in determining mean cumulative duration of inactivity during the main sleep bout. Intercept (row one) is the predicted cumulative duration of inactivity during the main sleep bout for a male 45.4 years old, with a BMI of 29.9 kg/m^2, who is not currently employed as a commercial vehicle driver, with an AHI<5/hr, who initiates sleep between 8 PM and midnight on average, and who terminates sleep between 7 AM and 10 AM, on average. The parameter estimates in rows 4 to 12 indicate the expected differences in mean cumulative duration of inactivity during the main sleep bout for the groups shown. The results of this analysis are described more fully in the text.

Table 4.4. Variables Determining Cumulative Duration of Inactivity in Subjects Without Sleep Apnea, i.e., AHI<5 episodes/hr

Factor	Parameter Estimate	SE	p-value
Intercept	452.60	13.42	
Age - 45.4	-0.900	0.414	0.031
BMI - 29.9	-2.54	1.03	0.015
Female vs. male	41.04	17.75	0.022
Full vs. unemp	-16.79	12.93	0.196
Part vs. unemp	-16.99	14.69	0.249
Miles/yr>=100k	11.87	12.33	0.337
Begin sleep not 8pm-12	-61.64	11.27	<0.0001
Terminated sleep not 7am-10am	-62.71	10.17	<0.0001

Results of multiple linear regression to evaluate variables that play a role in determining mean cumulative duration of inactivity during the main sleep bout among subjects with no significant sleep apnea (AHI<5 events/hr). Intercept (row one) is the predicted cumulative duration of inactivity during the main sleep bout for a male 45.4 years old, with a BMI of 29.9 kg/m^2, who is not currently employed as a commercial vehicle driver, who initiates sleep between 8 PM and midnight on average, and who terminates sleep between 7 AM and 10 AM, on average.

4.8 Movements During Sleep (Disturbed Sleep)

As illustrated in Figure 4.1, some subjects with sleep apnea had a large number of movements during their sleep. This has been demonstrated previously [Middelkoop et al, 1995] and these movements are the result of the arousal from sleep, as described in Chapter Two, that occur at the end of the apneic episode. To analyze this in more detail, in Figures 4.7 and 4.8, we show scatterplot measurements of the number of average number of movements/hour of sleep plotted against apnea severity measured by the apnea/hypopnea index (AHI). The data are shown for the mean number of movements during the various sleep bouts (Figure 4.7) and for the median number of events/hour across the various sleep bouts for individual subjects (Figure 4.8).
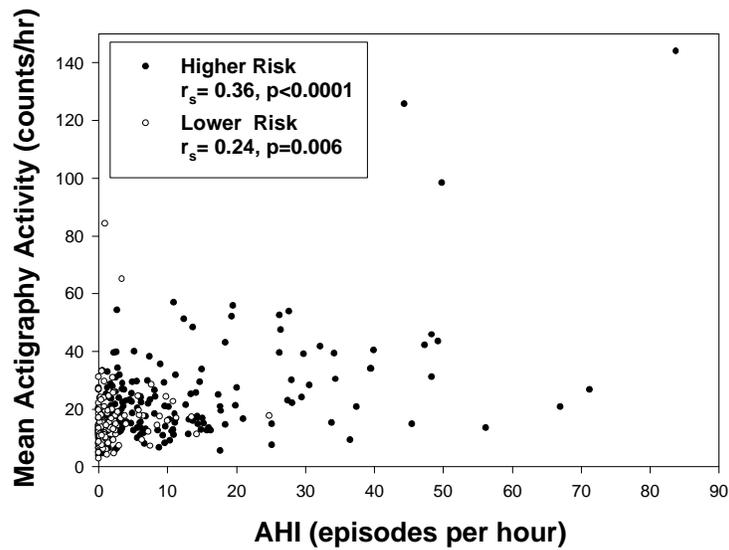
Figure 4.7. Mean actigraphy (counts for movements per hour of sleep) versus apnea/hypopnea index (AHI) (episodes per hour of sleep). There is a correlation between the variables in both the higher risk group (p<0.0001) and in the lower risk group (p=0.006).
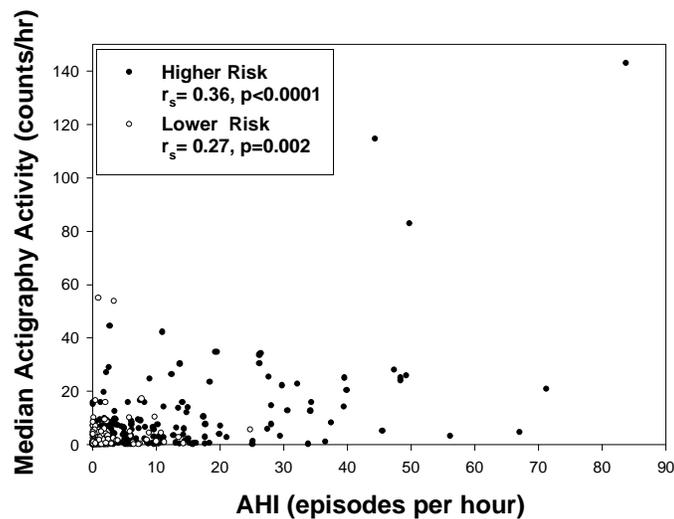
Figure 4.8. Median actigraphy (counts of movements per hour of sleep) versus apnea/hypopnea index (AHI) (episodes per hour of sleep). There is again a correlation between these variables in both the higher risk group (p<0.001) and in the lower risk group (p=0.002).

Thus, subjects with more severe sleep apnea have greater numbers of movements during sleep. That is, at home their sleep is more disturbed. These movements may produce artifactual increases in actigraphically determined cumulative durations of inactivity during the main sleep bouts as described above. However, some evidence of reduced sleep independent of these movements is brought to light by analyses of main bout of relative inactivity. Hence, both factors may interact to lead to decrements in function, as described in more detail in Chapter Seven.

That subjects with more severe sleep apnea have higher levels of movement during sleep might suggest that actigraphy, such as we performed here, could have a role in helping detect drivers with severe sleep apnea. The potential value of this needs to be assessed in future studies evaluating different strategies in combination to identify drivers with severe sleep apnea.

4.9 <u>Summary and Discussion</u>

These actigraphy results reveal several interesting observations. First, we investigated the effect that movements during sleep have on estimates of sleep duration. To this end, we employed two definitions. The definition most nearly equivalent to total sleep duration is the ***cumulative duration of inactivity during the main sleep bout***. However, this definition may be subject to downward bias since movements during sleep secondary to sleep apnea may be interpreted as wakefulness. An alternative definition not subject to this bias is based on the elapsed time from initiation of the main sleep bout to termination of the main sleep bout. Since the main sleep bout is identified by bracketing a period of relative inactivity, we refer to this time interval as the ***main bout length of relative inactivity***. Finally, analyses were preformed for cumulative durations of inactivity during the main sleep bout restricting attention to subjects with no significant sleep apnea (AHI<5 events per hour).

Average sleep duration for these three approaches were 7.38 hours, 6.33 hours, and 6.54 hours, respectively. We do not find the short average durations that were reported by Mitler et al [1997]. Nonetheless, we estimate that in our population of drivers, 9.8% have mean bout lengths of relative inactivity less than 6 hours and 31.0% mean bout lengths less than 7 hours. When assessed by cumulative durations of inactivity during the main sleep bout, 13.5% had less than 5 hours and 35.3% had less than 6 hours. Furthermore, even among those without significant sleep apnea as determined by an AHI<5 events per hour, 10.2% have mean cumulative durations of inactivity during the main sleep bout less than 5 hours and 28.9% have mean cumulative durations less than 6 hours.

While the role of such short sleep durations in increasing the risk of crashes in commercial drivers is unknown, it has been shown in a case-control study of crashes of passenger cars, attributed by the police to the driver falling asleep at the wheel, that such sleep durations are associated with a substantially increased risk of such crashes [Stutts et al, 1999]. Thus, mean main bout lengths of less than 7 hours in 31.0% of the population and mean cumulative durations of inactivity during the main sleep bout of less than 6 hours among 28.9% of drivers with no significant sleep apnea are a source of concern.

Analysis of the factors related to the short sleep durations showed that a number of variables played a role. Not surprisingly, time of sleep initiation and sleep termination were

independently significant important predictors of sleep duration no matter which of the three approaches were employed. Initiation of sleep at times other than between 8:00 PM and midnight and termination of sleep times other between 7:00 AM and 10:00 AM both were simultaneously associated with mean decreases in sleep durations of approximately three quarters of an hour to one hour. Even though there were a small number of females, gender was consistently significant with females sleeping on average about three quarters of an hour more than males. Presence of sleep-disordered breathing with apnea/hypopnea indices of >30 episodes/hour were associated with shorter sleep durations by almost 1.25 hours when estimated by cumulative duration of inactivity during the main sleep bout ($p=0.0002$). However, this finding was not confirmed when examining the main bout length of relative inactivity ($p=0.574$). Interestingly, when restricting attention to subjects with AHI<5/hr, increasing age and increasing obesity were independently associated with decreased mean cumulative durations of inactivity.

Finally, our studies confirm previous reports that there are greater movements during sleep in some subjects with sleep apnea. Whether or not these are associated with true wakefulness, these movements may reflect disturbed sleep potentially associated with impaired daytime performance. This is discussed in more detail in Chapter Seven of this report. Moreover, these findings suggest that actigraphy could play a role in combination with other tools in screening strategies to identify drivers likely to have sleep apnea. One important advantage of this approach is that the technique will provide data not only on the movements that occur during sleep, but also on average sleep duration, variability from night to night in sleep duration, and times of initiating and terminating sleep.

CHAPTER FIVE

Prevalence of Sleep-Disordered Breathing

5.1. Presence of Symptoms of Sleep Apnea in All Respondents

A. Snoring

In our survey of the entire group of holders of CDLs, we asked 12 self-report symptom frequency questions of the form: "During the last month, have you had, or have been told about the following symptom (Show the frequency): (0) Never; (1) Rarely, Less Than Once a Week; (2) 1-2 Times Per Week; (3) 3-4 Times Per Week; (4) 5-7 Times Per Week; (.) Don't Know". One of these questions concerned the frequency of loud snoring. The percentage of respondents with different frequencies of this complaint is shown in Figure 5.1. In our sample, 32.6% had frequent loud snoring (more than 3 times/week) while 48.6% had loud snoring more than 1-2 times/week.
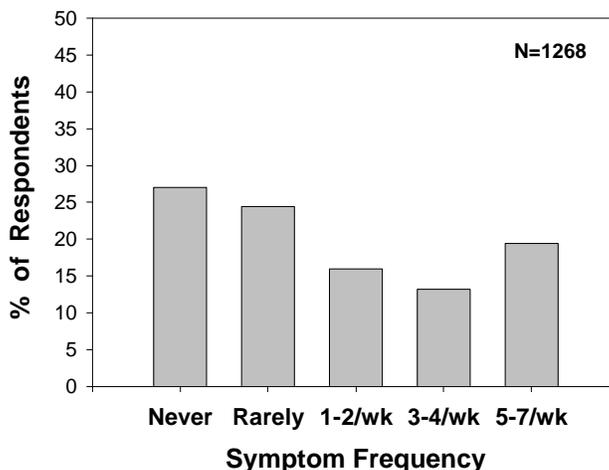
Figure 5.1. Percentage of all respondents who had different frequencies of symptoms of loud snoring. These data were derived from 1268 respondents in whom responses to this question were obtained.

B. Snoring and Gasping/Witnessed Apneas

The complaints of snoring and gasping and witnessed apneas were much less common than snoring (see Figures 5.2 and 5.3). As pointed out in Chapter Two, witnessed apneas is not a particularly sensitive symptom (i.e., many subjects have sleep apnea but do not have this complaint) but is highly specific (i.e., most subjects with this complaint have sleep apnea). In our sample, 7.2% of respondents (N=1296) had witnessed apneas on a frequent basis (more than 3 times/week) while 10.9% of respondents had this at least once/week. This suggests that sleep apnea is relatively common in our sample as we examined in more detail in our in-laboratory studies.
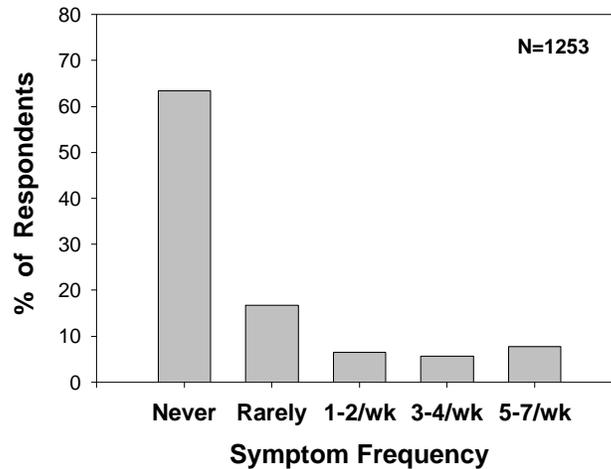
Figure 5.2. Percentage of respondents who reported different frequency of occurrence of snorting and gasping during sleep. The number of respondents (n=1253) is shown in the figure.
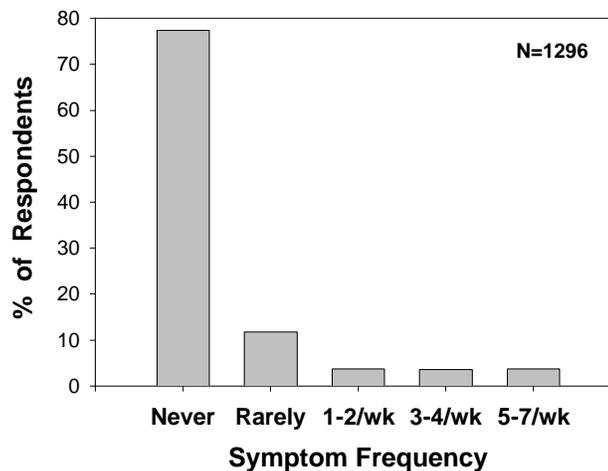
Figure 5.3. Percentage of respondents who reported different frequency of symptoms related to witnessed apneas, i.e., breathing stops, struggle for breath. The number of respondents is given in the figure. The percentage of subjects with this particular symptom is less than for the other symptoms shown.

C. <u>Determinants of Symptoms in All Respondents</u>

As outlined in Chapter Two, there are three known important determinants of sleep apnea risk, i.e., age, gender and body mass index. We examined the relationship between these variables and two of the symptoms of sleep apnea, i.e., frequent loud snoring ($\geq$3 times/week) and the presence of any witnessed apneas ($\geq$1 time/week). The number of females in our sample is small (n=97) but we found that frequent loud snoring occurred in 20 of 81 (24.7%) females who answered this question compared to 394 of 1187 (33.2%) of males. The estimated difference in percentages between males and females is 8.5% but is not statistically significant.

Similarly, we found that reports of breathing stops/struggle for breath at least once per

week occurred in 5 of 91 females giving a frequency of 5.5%, compared to 136 of 1205, i.e., 11.3% of males. The estimated difference (95% confidence interval) in percentages between males and females is 5.8% (0.8% to 10.8). The confidence interval excluded zero reflecting a statistically significant difference. Thus, symptoms of sleep apnea were more common in men than women in our sample but the only statistically significant difference was for reports of breathing stops/struggle for breath at least once per week.

Next, we examined the relationship between BMI and age with frequency of occurrence of these two symptoms. Figure 5.4 shows the relationships with obesity as categorized by BMI according to the guidelines for obesity (from the National Institutes of Health) [Anonymous, 1998]. These guidelines propose that BMI ≤ 25 kg/m^2 is not overweight; BMI >25 and ≤ 29 kg/m^2 is overweight; BMI ≥ 30 and ≤ 34 kg/m^2 is class 1 obesity; BMI ≥ 35 and ≤ 39 kg/m^2 is class 2 obesity; while ≥ 40 kg/m^2 is class 3 obesity. The rates of symptom reports from sleep apnea increase linearly as a function of BMI for both loud snoring $>=3$/wk (Cochran's test for linear trend $\chi^2 = 50.0$, df=1, p<0.0001) and for witnessed apneas ($\chi^2 = 33.9$, p<0.0001).
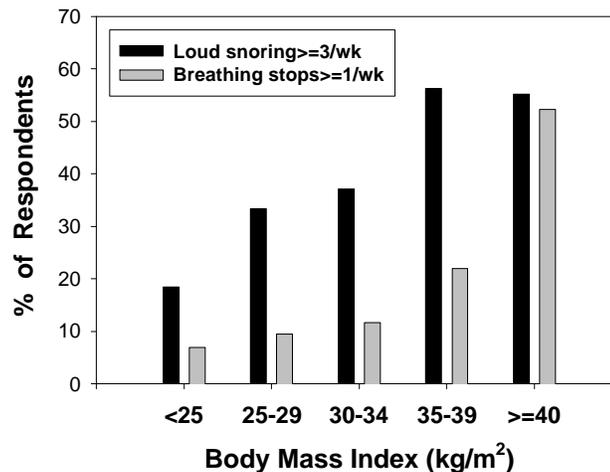
Figure 5.4. Percentage of respondents with symptoms of apnea—frequent loud snoring, witnessed apneas—in different categories of obesity ranging from not overweight (BMI <25, left) to class 3 obesity (BMI ≥ 40, right). The percentage of subjects with either symptom increases with increasing obesity.

Figure 5.5 provides the analogous analyses for age categorized by decade. In contrast to BMI, there was no significant linear trend in the association between age and the rate of reports of frequent (≥ 3/wk) loud snoring (p=0.287), nor between age and the report of breathing stops/struggle for breath at least once per week (p=0.503). Loud snoring is not a particularly specific indication of sleep apnea, i.e., many individuals who snore do not have sleep apnea. Snoring is, however, likely to arise early in the disease process. This likely explains in this group the lack of an association with age. Later in the chapter we provide clear evidence that sleep apnea itself is associated with age.
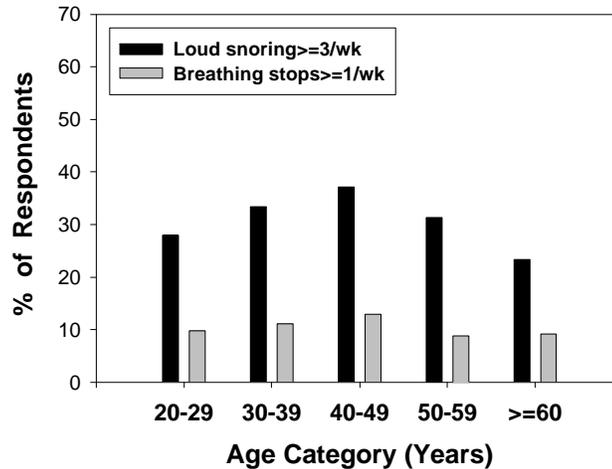
Figure 5.5. Percentage of all respondents with the different symptoms of sleep apnea—
frequent loud snoring and witnessed apneas—as a function of age. There is no clear
effect on age on the frequency of these symptoms.

5.2. Symptoms in Higher and Lower Risk Groups

A. Frequency of Symptoms

The same symptom frequency questions were asked again of subjects enrolled for in-
laboratory studies. The various questionnaires used in the laboratory phase of the study are
given in Appendix J. The revised version of the sleep disorders symptom questionnaire included
a question concerning the frequency of "any snoring" in addition to "loud snoring". The
responses in the higher and lower risk groups are shown in Tables 5.1a and 5.1b. As anticipated,
all of the symptoms are more prevalent in the higher risk group. For example, 1.3% of the lower
risk subjects reported witnessed apneas at least once/week while 11.3% of the higher risk group
reported witnessed apneas of this frequency.

Table 5.1a. Reports of apnea symptom frequencies for in-laboratory higher risk group.

Symptom	Never	Rarely	1-2 times/week	3-4 times/wee	5-7 times/week	Don't know
Loud snoring	63 (25.7%)	40 (16.3%)	47 (19.2%)	37 (15.1%)	46 (18.8%)	12 (4.9%)
Any snoring	57 (23.8%)	40 (16.7%)	49 (20.5%)	37 (15.5%)	39 (16.3%)	15 (6.3%)
Snorting and gasping	109 (44.9%)	52 (21.4%)	20 (8.2%)	25 (10.3%)	12 (4.9%)	25 (10.3%)
Witnessed apneas	154 (63.9%	39 (16.2%)	17 (7.1%)	6 (2.5%)	4 (1.7%)	21 (8.7%)

Table 5.1b. Reports of apnea symptom frequencies for in-laboratory lower risk group.

Symptom	Never	Rarely	1-2 times/week	3-4 times/week	5-7 times/week	Don't know
Loud snoring	54 (34.4%)	40 (25.5%)	26 (16.6%)	10 (6.4%)	13 (8.3%)	14 (8.9%)
Any snoring	77 (49.0%)	39 (24.8%)	19 (12.1%)	7 (4.5%)	7 (4.5%)	8 (5.1%)
Snorting and gasping	120 (76.9%)	18 (11.5%)	9 (5.8%)	1 (0.6%)	--	8 (5.1%)
Witnessed apneas	138 (87.9%)	10 (6.4%)	2 (1.3%)	--	--	7 (4.5%)

B. Association Between Smoking and Alcohol Ingestion with Symptoms of Apnea

For subjects studied in-laboratory we had additional information about risk factors for these symptoms, i.e., smoking history and alcohol intake. Figure 5.6 displays the comparison of percentage of individuals with frequent loud snoring (≥ 3 times/week) between subjects who reported smoking at some point during their lifetime versus those who reported never smoking, and between subjects who were current smokers and subjects who were not current smokers. Never smoking was defined as less than 20 packs in a lifetime, or less than one cigarette a day over a year. Analyses were stratified by risk group in order to account for the sampling design. In both the higher and lower risk groups, the prevalence of "frequent loud snoring" was slightly larger among those who had smoked at some point during their lifetime, but this difference was not significant controlling for risk group. In the comparison between subjects who were current smokers compared to those who were not current smokers, this difference was observed only in the lower risk group. This difference was also not statistically significant. Thus, in this study, we did not find a relationship between smoking history and the report of frequent loud snoring.
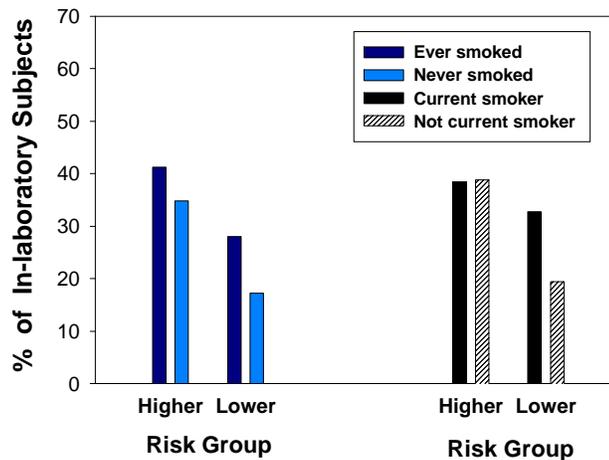
Figure 5.6. Percentage of subjects studied in the two in-laboratory groups with frequent loud snoring (≥ 3 times/week) who ever smoked or never smoked (left panel) and who were current smokers or not current smokers (right panel). For the ever smoker vs. never smoker comparison there was no significant difference in the frequency of this symptom. For the comparison of current smoker vs. not current smoker, a difference was observed only in the lower risk group but overall this difference was not significant.

Current smoking significantly increased the rate of another important apnea-related symptom, i.e., breathing stops or struggle for breath. Comparisons for both higher and lower risk groups are illustrated in Figure 5.7.

In-laboratory subjects were also asked three questions concerning their usual frequency of intake of wine, beer, and spirits over the last month (in terms of drinks per week). These were totaled and categorized as 0-<1 per day, 1-<2 per day, and ≥2 per day. There were no significant associations between frequent loud snoring or witnessed apneas and degree of alcohol intake (data not shown).
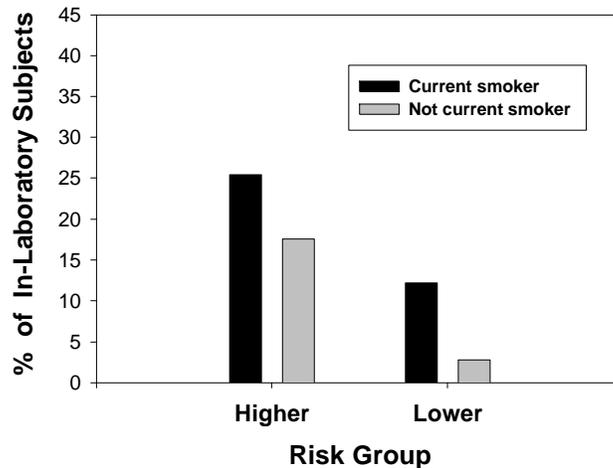
Figure 5.7. Percentage of the two groups studied in-laboratory who had witnessed apneas at least once/week. This symptom was slightly more frequent in current smokers.

5.3 Assessment of Presence of Sleep-Disordered Breathing

A. Methods Employed

Subjects in the higher and lower risk groups had full overnight polysomnographic sleep studies performed. They were admitted to the Clinical Research Center for Sleep at the University of Pennsylvania. They were given an opportunity to sleep throughout the night.

During sleep, the polysomnography was recorded on a paperless polygraph system (NPB Melville Sandman monitoring system). The following measurements were included in each recording: two monopolar electro-encephalographic (EEG) leads, which included scalp electrodes at C3/A2 with C4/A1 as backup and OZ/A2; two monopolar electro-oculograms (EOG: right and left outer canthi reference to A2 and A1, respectively); two bipolar submental (chin) and leg (tibialis) electromyograms (EMG); ECG, which was measured using two electrodes each placed in the subclavical area on the right and left side of the chest. Nasal and oral airflows were recorded using thermocouples (Protech Services, Inc.) which were taped on the face by the nose and mouth to measure respiratory flow. Bands were placed over the chest and abdomen to monitor respiratory effort (EPM, Resp-EZ). Body position was monitored

visually by the technical staff. Oxygen saturation was measured with a finger clip (Nellcor Puritan Bennett). Thus, the variables outlined in Table 5.2 were recorded.

Experienced technologists scored the sleep study data according to standardized criteria of Rechtstaffen and Kales [1968]. The data were scored manually in 30-second epochs. Two types of scoring validation were performed. The first validation focused on scoring the respiratory events. The goal of this validation was to look at both inter-scorer (between scorers) and intra-scorer (repeated scoring by a single scorer) reliability. Four technologists were asked to score the respiratory data for 30 randomly selected subjects among the 300 available at the time of the validation study and then to go through and score each one again (n=240 scored records). A variance components analysis was implemented using random effects analysis of variance. The intraclass correlation coefficient (ICC) was computed in order to determine reliability. The ICC was 0.989 indicating that almost all (i.e., 98.9%) of the variance in these 240 AHI values was explainable by true differences between subjects. Thus, we determined that our primary disease severity measure was obtained with excellent reliability. Throughout the study we assured the overall quality of the scoring. Thus, throughout the study a 10% random sample of the study PSGs was taken and blindly re-scored by the technologists.

Table 5.2. Laboratory assessment of presence of sleep-disordered breathing.

VARIABLE MEASURED	TECHNOLOGY USED	USE OF MEASUREMENT
RESPIRATION		
Airflow	Thermistors	To help detect apneas/hypopneas
Respiratory effort	Respiratory bands	To help detect apneas/hypopneas and reveal obstructive nature of events
Oxygen saturation	Oximeter (Nellcor)	To detect oxygen desaturation
SLEEP STATE		
Electroencephalogram (EEG)	Scalp electrodes	Detection of brain waves associated with different sleep states
Electro-oculogram (EOG)	EOG electrodes	Detection of eye movements to identify rapid-eye-movement sleep
Electromyogram (EMG)	Submental EMG	Detection of characteristic changes in muscle tone in different sleep states
PERIODIC LIMB MOVEMENTS		
Leg jerks	Electrodes over anterior tibialis muscle	To detect rhythmic movements in legs during sleep
CARDIOVASCULAR		
Electrocardiogram (ECG)	Electrodes on chest	To determine heart rate and rhythm

B. Prevalence of Sleep Apnea

As outlined in Chapter Two, there is now a consensus definition of an obstructive apnea/hypopnea. This was given in Table 2.1. One criterion is a clear amplitude reduction in a validated measure of breathing during sleep that is associated with either an oxygen desaturation of >3% or an arousal. We utilized this definition and obtained the prevalence of sleep apnea in the higher and lower risk groups shown in Table 5.3. We provide prevalence estimates for the various severities of sleep apnea [American Academy of Sleep Medicine, 1999]: mild sleep apnea (AHI between 5 and 15 episodes/hour); moderate sleep apnea (AHI between 15 and 30 episodes/hour); severe sleep apnea (AHI greater than 30 episodes/hour). We also show in Table

5.4 percentage of subjects in both groups who are above different cutpoints of abnormality, i.e., from AHI≥5 to AHI≥40 episodes per hour.

Table 5.3. Estimates of prevalence of sleep apnea of different severities. The data shown are for higher and lower risk groups and for the weighted average. Overall, we can see that estimated population prevalence of severe (≥30), moderate (15-<30), and mild (5-<15) obstructive sleep apnea were 4.7%, 5.8%, and 17.6%, respectively.

AHI$^{\&}$	Higher Risk					Lower Risk					Weighted			
	N	Rate	SE	(95%	C.I.)	N	Rate	SE	(95%	C.I.)	Rate	SE	(95%	C.I.)
<5	119	0.482	0.032	0.419	0.544	141	0.887	0.025	0.838	0.936	0.719	0.020	0.680	0.757
5 - <15	70	0.283	0.029	0.227	0.340	16	0.101	0.024	0.000	0.147	0.176	0.018	0.141	0.212
15 - <30	30	0.121	0.021	0.081	0.162	2	0.013	0.009	0.000	0.030	0.058	0.010	0.038	0.077
>= 30	28	0.113	0.020	0.074	0.153		0.000			0.000	0.000	0.047	0.008	0.031	0.063

Notes: $^{\&}$ Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.

Table 5.4. Estimates of prevalence of sleep apnea using various cutoff definitions from 5 to 40 events per hour.

Cutoff	Higher Risk					Lower Risk					Weighted			
	n	Rate	SE	(95%	C.I.)	n	Rate	SE	(95%	C.I.)	Rate	SE	(95%	C.I.)
RDI>= 5	128	0.518	0.032	0.456	0.581	18	0.113	0.025	0.064	0.162	0.281	0.020	0.243	0.320
RDI>=10	86	0.348	0.030	0.289	0.408	6	0.038	0.015	0.008	0.067	0.167	0.015	0.136	0.197
RDI>=15	58	0.235	0.027	0.182	0.288	2	0.013	0.009	0.000	0.030	0.105	0.012	0.081	0.129
RDI>=20	44	0.178	0.024	0.130	0.226	2	0.013	0.009	0.000	0.030	0.081	0.011	0.059	0.104
RDI>=30	28	0.113	0.020	0.074	0.153	0	0.000	0.000	0.000	0.000	0.047	0.008	0.031	0.063
RDI>=40	15	0.061	0.015	0.031	0.091	0	0.000	0.000	0.000	0.000	0.025	0.006	0.013	0.038

5.4 Determinants of Prevalence of Sleep Apnea

As outlined in Chapter Two, there are a number of known determinants of sleep apnea that have been evaluated in previous epidemiological studies of other populations—gender, age, BMI, alcohol intake, and smoking history. In addition, the design of our study allowed us to look at other potentially important determinants of prevalence—sleep duration, employment status and the number of miles driven/year.

A. Gender and Prevalence of Sleep Apnea

The prevalence rates in men and women with different apnea/hypopnea indices are shown in Figure 5.8. The sample of women is small given the small percentage of commercial drivers who are women. The prevalence of sleep apnea appeared slightly smaller in women in the mild category.
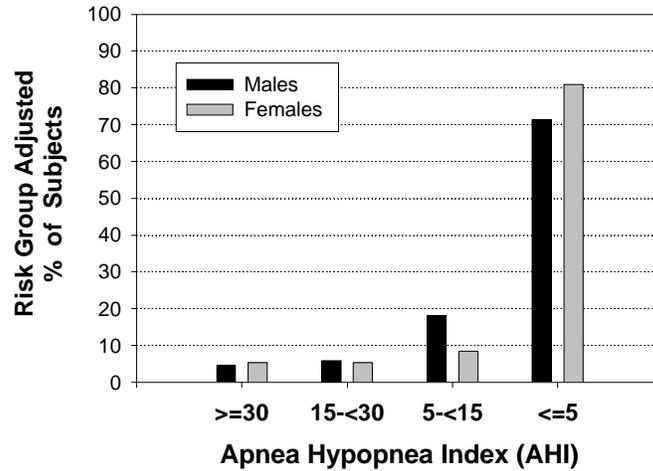
Figure 5.8. Prevalence values were computed based on weighted data to account for sampling by risk group.

B. Age, BMI and Their Effect on Sleep Apnea Prevalence

Given that our sample was predominantly male (n=94.6%, weighted n=93.3%), we examined the interrelationship between prevalence of sleep apnea in our sample and age and body mass index in only males. This also allows comparison of our results with the three previous major epidemiological studies of sleep apnea (see Table 2.3, Chapter Two). We show in Table 5.5 the prevalence of different severities of sleep apnea in different age groups (decades) and in different grades of obesity. Prevalence estimates in this table are weighted to reflect the age and BMI distribution observed among population survey respondents.

Table 5.5. Estimated apnea prevalence in males by age and body mass index (BMI) category (normal weight, <25; overweight, ≥25 & <30; Class 1 obesity, ≥30 & < 35; Class 2 obesity, ≥35 & <40; Class 3 obesity, ≥40. The latter two categories are combined.

Age	BMI	Normal Weight <25	Over-weight 25-<30	Class 1 Obesity 30-<35	Class 2 Obesity >=35	Total
20-29	PSG's[&]	4	7	5	3	19
	AHI>=5	0.234	0.043	0.152	1.000	
	AHI>=10	0.000	0.043	0.000	0.000	
	AHI>=15	0.000	0.000	0.000	0.000	
	AHI>=20	0.000	0.000	0.000	0.000	
	AHI>=30	0.000	0.000	0.000	0.000	
	AHI>=40	0.000	0.000	0.000	0.000	
30-39	PSG's	12	35	25	20	92
	AHI>=5	0.000	0.133	0.296	0.632	
	AHI>=10	0.000	0.034	0.166	0.526	
	AHI>=15	0.000	0.034	0.130	0.368	
	AHI>=20	0.000	0.017	0.086	0.316	
	AHI>=30	0.000	0.000	0.043	0.158	
	AHI>=40	0.000	0.000	0.000	0.105	
40-49	PSG's	16	45	39	29	129
	AHI>=5	0.000	0.233	0.424	0.759	
	AHI>=10	0.000	0.070	0.233	0.586	
	AHI>=15	0.000	0.052	0.169	0.310	
	AHI>=20	0.000	0.017	0.105	0.241	
	AHI>=30	0.000	0.017	0.064	0.172	
	AHI>=40	0.000	0.000	0.032	0.069	
50-59	PSG's	7	35	26	13	81
	AHI>=5	0.147	0.244	0.634	0.615	
	AHI>=10	0.147	0.122	0.492	0.538	
	AHI>=15	0.147	0.018	0.358	0.462	
	AHI>=20	0.147	0.018	0.269	0.385	
	AHI>=30	0.000	0.018	0.090	0.308	
	AHI>=40	0.000	0.000	0.000	0.231	
>=60	PSG's	4	27	24	7	62
	AHI>=5	0.122	0.481	0.304	0.714	
	AHI>=10	0.000	0.297	0.217	0.571	
	AHI>=15	0.000	0.152	0.174	0.286	
	AHI>=20	0.000	0.114	0.174	0.286	
	AHI>=30	0.000	0.114	0.130	0.143	
	AHI>=40	0.000	0.038	0.130	0.143	
	Total	43	149	119	72	383

Notes:
Prevalence estimate is weighted average from higher and lower risk groups. Weights were determined within age and BMI category from population-based screen in males only.
[&] Number of males subject (pooling higher and lower risk groups) with polysomnography within age and BMI strata, on appropriate row.

There are several observations. First, in our sample, the prevalence of sleep apnea generally appeared to increase with age within all strata of obesity but leveled off after the age of 60.   This is illustrated in Figure 5.9 for AHI≥15. We do not believe that much import can be given to the apparent reduction in prevalence of this level of abnormality in subjects with BMI ≥35 who are over sixty given the sample size of this group is extremely small (n=7). Second, the effect of increasing BMI on apnea prevalence depended on age. For example, in the small sample of subjects with Class 1 obesity or above (n=9) in the decade between 20 and 29 years, no subject had moderate or severe sleep apnea despite the presence of the known risk factor for sleep apnea, i.e., obesity. This is compatible with the concept that sleep apnea is a slowly progressive disorder. This is illustrated in Figure 5.10 where we show the effect of increasing BMI on the prevalence of severe apnea (AHI≥30 episodes/hour) in different decades.   An analogous analysis examining the prevalence of at least moderate apnea (AHI≥15) is presented in Figure 5.11.
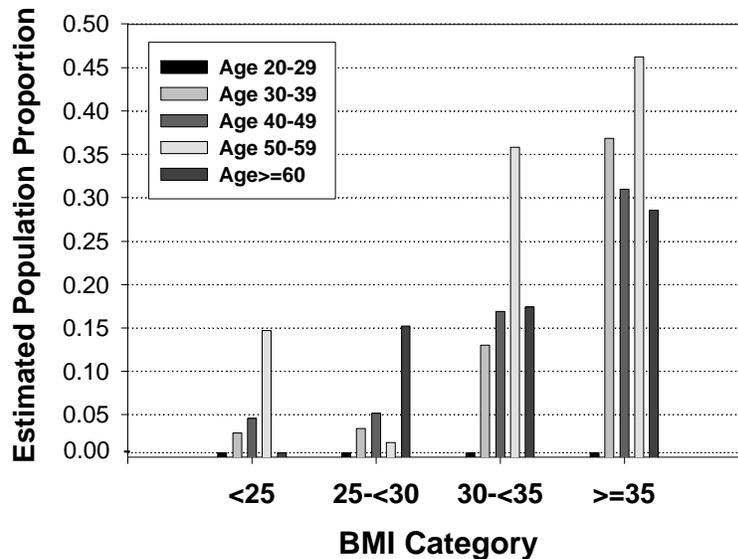
Figure 5.9.   Prevalence values of apnea/hypopnea index ≥15 episodes/hour were computed based on population-derived risk-group weights within age and BMI category strata (post-stratification).

As can be seen in Figures 5.10 and 5.11, and by examination of the data in Table 5.5, there is an interactive effect of BMI and age on prevalence of severe sleep apnea. Increasing degrees of obesity have a marked effect on its prevalence in older subjects but had limited (if any) impact in the youngest individuals. This is particularly the case for individuals with modest degrees of obesity, compatible with the view that the speed of progression of the disorder depends on the degree of obesity. For those who are very obese, the disorder seems to develop over a shorter time window than for those with lesser degrees of obesity.
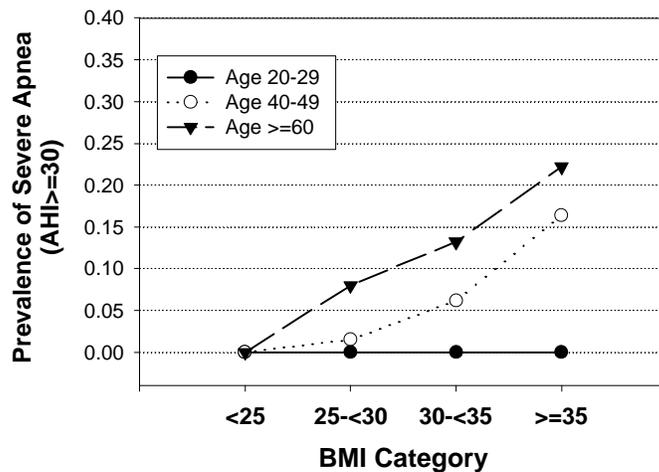
Figure 5.10. Prevalence of severe sleep apnea (AHI≥30 episodes/hour) as function of degree of obesity. The data are shown for only three different age groups—20-29 years (filled circles), 40-49 years (open circles), and ≥60 years (filled triangles) to simplify presentation. Data for other decades are given in Table 5.5.
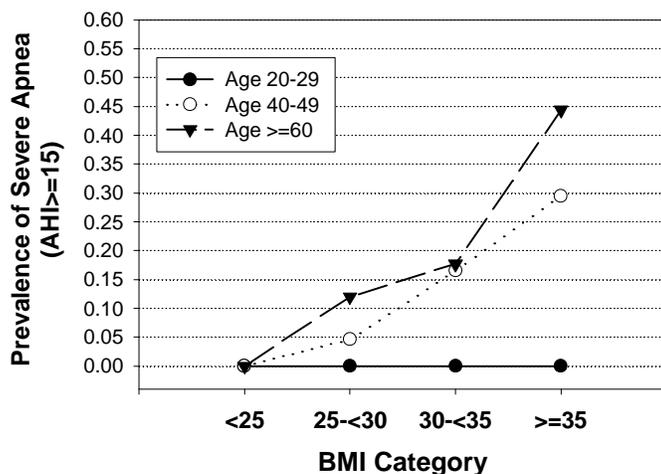
Figure 5.11. Prevalence of at least moderate sleep apnea (AHI≥15 episodes/hour) as function of degree of obesity. The data are shown for only three different age groups— 20-29 years (filled circles), 40-49 years (open circles), and ≥60 years (filled triangles) to simplify presentation. Data for other decades are given in Table 5.5.

When interpreting the finding described above it is critical to consider what is meant by "interactive effect". Here, we are referring to 'public-health interaction' as described by Rothman et al [1980] in contrast to statistical interaction. "The primary concern of public health interaction is the number of cases of disease occurring in a population and the proportional contribution of each risk factor to this case burden ….. [and] that for public-health purposes, interaction is equivalent to a departure from additivity of incidence rate differences". Thus, e.g., if the presence of risk factor A increases risk by a factor of 3 and the presence of risk factor B

increases risk by a factor of 10, and when combined, the risk increases by a factor of 30, we have important public-health interaction between risk factors A and B (i.e., when A is absent, presence of B increases risk by 10 fold; but when A is present, presence of B increases risk by 30 fold). In contrast, statistical interaction is a property of statistical models and depends on the choice of metric. The type of 'public-health interaction' described above reflects a multiplicative relationship that is, in fact, additive on a logarithmic scale. Thus, it is properly modeled using a logistic regression model containing no statistical interaction terms.

These considerations and data from the present study allowed us to develop prediction equations for prevalence of different severities of sleep apnea to inform policy makers. These predictions are based on the (multiplicative) effects of increased BMI and age on risk of sleep apnea. The equations we develop are only for males, the vast majority of this working group. Such equations allow our data to be extrapolated to other populations of truck drivers if the age and BMI distributions in such populations are known. To do this, the age and BMI of each driver would be entered in one of the three equations below. The equation is then used to compute the *logit*, or log odds, for each driver. Then each logit is turned into a predicted probability using the formula, predicted probability = $e^{logit}/(1+e^{logit})$. Finally, the predicted probabilities are added up to produce the expected number of drivers with sleep apnea at the specific severity level. Dividing the expected number of drivers with sleep apnea by the total number of drivers produces an age- and BMI-adjusted estimate of prevalence specific for each company. The measure used to assess the accuracy of predictions on an individual basis was a discrimination index computed as the area under the Receiver Operator Characteristic (ROC) curve based on predicted probabilities (AUC) [Harrell et al, 1996]. The AUC is the estimated probability that a randomly selected driver with sleep apnea has a larger predicted probability that a randomly selected driver without sleep apnea. That is, it is the probability that the model can pick out the driver with sleep apnea from two drivers, one with and one without sleep apnea. It is to be emphasized that these formulas are applicable to estimate the prevalence of drivers with different levels of severity of sleep apnea for a specific population of drivers. They were not designed for use in detecting whether a specific driver has apnea or not. However, AUC values approximately equal to 0.80 or larger typically reflect prediction equations that are accurate enough for risk stratification.

Table 5.6. Equations to allow truck owners to develop estimates of prevalence of sleep apnea of different severity in their population based on ages and BMI's of their drivers. Equations determined using logistic regression for estimating apnea prevalence at 3 severities as a function of age and body mass index (BMI). The equations determine the log odds of apnea (logit). Estimated prevalence is then determined using the equation $e^{logit}/(1+e^{logit})$. *"Goodness of Prediction" is based on Area Under Curve (AUC) of Receiver Operating Characteristic (ROC) curve. The maximum for this measure of Goodness of Fit is 1.0.

Severity of Apnea	Equation	Accuracy of Prediction
Equation 1. Mild and above (AHI≥5 episodes/hour)	logit = 0.1235 * AGE – 0.000800 * AGE^2 +0.3863 * BMI – 0.002285 * BMI^2 – 13.9046	0.748
Equation 2. Moderate and above (AHI≥15 episodes/hour)	logit = 0.2683 * AGE – 0.00240 * AGE^2 +0.4009 * BMI – 0.00278 * BMI^2 – 19.9089	0.768
Equation 3. Severe (AHI≥30 episodes/hour)	logit = 0.2661 * AGE – 0.00189 * AGE^2 +0.3624 * BMI – 0.00175 * BMI^2 – 21.1748	0.813
Predicted Probability = $e^{logit}/ (1+e^{logit})$		

The equations outlined in Table 5.6 allow us to develop graphical representations of the prevalence of sleep apnea of different severities as a function of the major variables—age and body mass index. These graphical representations are shown in Figure 5.12 (for prevalence of at least mild sleep apnea, AHI≥5 episodes/hour), Figure 5.13 (for prevalence of at least moderate sleep apnea, AHI≥15 episodes/hour) and Figure 5.14 (for prevalence of severe sleep apnea, AHI≥30 episodes/hour).

Results for Mild Sleep Apnea and Above
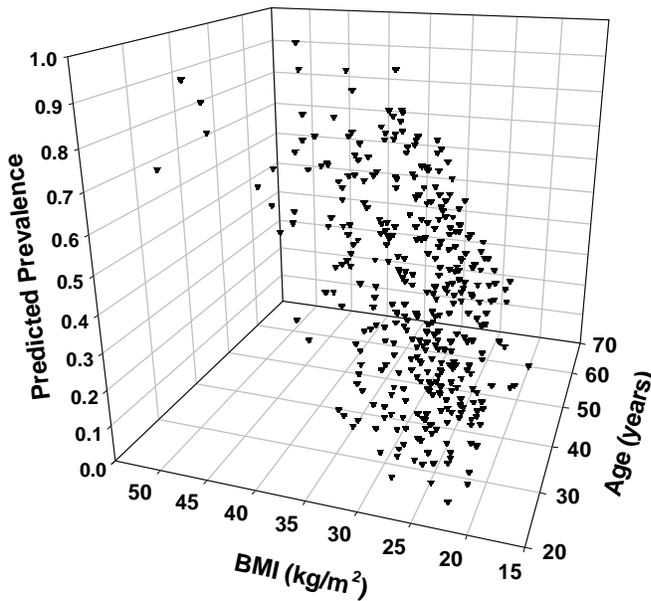
Figure 5.12. Prevalence density of AHI≥5 episodes/hour as function of degree of obesity and age. The points in the plot indicate the predicted probability for AHI≥5 episodes/hour from the non-linear logistic regression model summarized in Table 5.6. The graph also serves to illustrate the relative density of participants with BMI and age that places them in the highest risk portion of the distribution. The clear effect of increasing age, increasing BMI and their interaction are easily seen. The individual points are model-based predictions for the probability that each male CDL holder in our sample has an AHI≥5. The points in the graph were derived from Equation 1 in Table 5.6.

Results for Moderate Sleep Apnea and Above
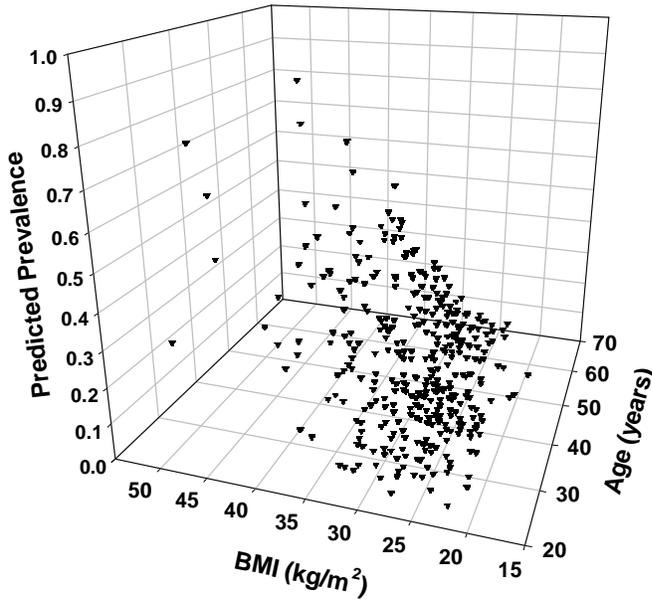
Figure 5.13. Prevalence density of AHI≥15 episodes/hour as function of degree of obesity and age. The points in the plot indicate the predicted probability for AHI≥15 episodes/hour from the non-linear logistic regression model summarized in Table 5.6. The graph also serves to illustrate the relative density of participants with BMI and age that places them in the highest risk portion of the distribution. The effects of age, BMI and their interaction on predicted prevalence are seen, but the predicted prevalence is, as expected, lower than in Figure 5.12. The individual points are model-based predictions for the probability that each male CDL holder in our sample has an AHI≥15. The points in the graph were derived from Equation 2 in Table 5.6.

Results for Severe Sleep Apnea
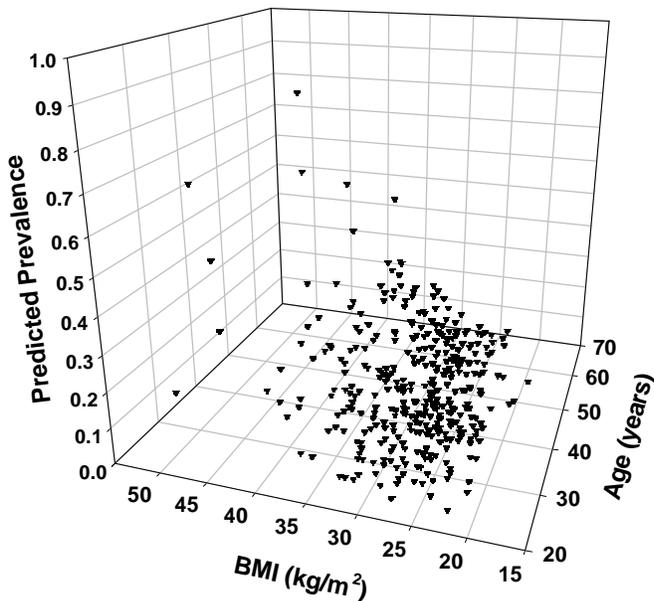
Figure 5.14. Prevalence density of AHI≥30 episodes/hour as function of degree of obesity and age. The points in the plot indicate the predicted probability for AHI≥30 episodes/hour from the non-linear logistic regression model summarized in Table 5.6. The individual points are model-based predictions for the probability that each male CDL holder in our sample has an AHI≥30. The points in the graph were derived from Equation 3 in Table 5.6.

In these figures (Figures 5.12 to 5.14) we can see that at any particular level of BMI, the prevalence of the different severities of apnea increase with age (particularly over 50 years of age). At any age, the prevalence of sleep apnea of the different severities increases with increasing BMI. We can also see, however, the public-health interaction previously alluded to.

The effect of increasing BMI on prevalence of sleep apnea is more pronounced at the older ages (e.g., above 50 years of age) than it is at the younger age groups as a consequence of the 'amplifying effect' produced by the multiplicative relationship.

The models in Table 5.6 assume a specific parametric form for the relationship between the log odds of sleep apnea and how this increases with age and with BMI. Specifically, the model assumes that for each age, the log odds increases as a quadratic function of BMI and that for each value of BMI, the log odds increases as quadratic function of age.

There were, however, a sufficient number of subjects with AHI>=5 at each of the nominal categories of age and BMI described in Table 5.5 to estimate the effect of differences in age and BMI categories with no such structural assumptions, that is using a non-parametric model. There are not sufficient numbers to do so for AHI>15 or AHI>30. The reference categories are the 20-29 years age group and the BMI <25 kg/m^2 group. The relative risks for each of the other age and BMI categories are compared to these reference groups. Results are displayed in Table 5.7. As can be seen, an age greater than 60 years is associated with approximately a three-fold increased risk of having at least mild apnea (OR=3.1) relative to an age between 20 and 29. Similarly, a BMI >=35 has an increased relative risk of more than 30 fold (OR=30.1) relative to subjects with a BMI of less than 25 kg/m^2. The results of this non-parametric model for AHI>=5 are shown in Figure 5.15. As with the figures describing the results of the other model, we see the following: (a) at any given BMI, prevalence increases with age; (b) at any age, prevalence increases with BMI; and (c) the effect of BMI on prevalence of apnea is more marked in individuals in the older age group.

Table 5.7. Non-parametric model for AHI≥5 in males.

	b	**OR**	**95% CI**
Intercept	-3.1377	-----	
Age 30-39†	-0.1376	0.9	(0.2 - 3.1)
Age 40-49†	0.4070	1.5	(0.4 - 5.2)
Age 50-59†	1.0081	2.7	(0.8 – 9.8)
Age >=60†	1.1328	3.1	(0.8 - 11.5)
BMI 25-<30‡	1.3931	4.0	(1.3 - 12.5)
BMI 30-<35‡	1.8559	6.4	(2.0 - 20.1)
BMI >=35‡	3.4276	30.1	(9.1 – 104.7)
Notes:			
† Indicator variable for comparison with age 20-29 years			
‡ Indicator variable for comparison with BMI < 25 kg/m^2			

The logistic regression model was estimated using weighted data to account for sampling design. The key column is that labeled "OR" which shows the increased relative risk of having an AHI>5 for different age categories and degrees of obesity. The confidence intervals (CI) are conservative since they ignore the design effect for AHI≥5 resulting from the stratified sampling design. The design effect for AHI≥5 is 0.787. That is, the estimated variance for the prevalence estimate of AHI≥5 from our stratified sampling design is only 78.7% as large as would have been obtained based on a simple random sample with the same sample size.
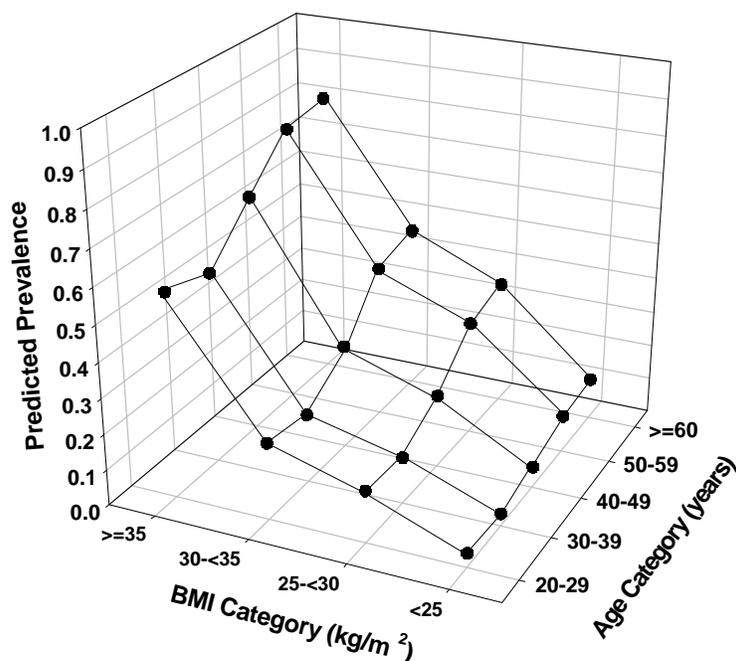
Figure 5.15. Non-parametric model for prevalence density of AHI≥5 episodes/hour as function of degree of obesity category and age category. The points in the plot indicate the predicted probability of AHI≥5 episodes/hour from the model summarized in Table 5.7. The effects of increasing age and increasing BMI on predicted prevalence are seen. Increasing BMI has a more pronounced effect on prevalence at older ages.

C. Neck Size and Sleep Apnea Prevalence

During physical assessment, neck size measurements were obtained. Among males in the higher and lower risk groups, the Pearson correlation coefficients for the association between neck size and BMI were r=0.738 (p<0.0001) and r=0.520 (p<0.0001), respectively. The partial correlation controlling for risk group was r=0.687 (p<0.0001). Spearman rank correlations were similar. Given the high correlation with BMI, it is not surprising that when BMI is replaced by neck size in the logistic regression models described above, predictive value of the models were comparable.   For example, in models containing age, age-squared, neck size and neck size squared, the discrimination index values for predicting AHI>=5, 15, and 30 episodes per hour were 0.704, 0.754, and 0.835. These values are similar to those presented in Table 5.6. Thus, the results of this analysis support the findings of Flemons et al [1994] who found that although neck circumference, when measured, is the most significant predictor among candidate anthropomorphic measurements, "The diagnostic information contributed by neck circumference appears to be replaceable by combinations of other variables that are covariates, such as the body mass index, age, or gender". Thus, our data indicate that while measurement of neck size is useful, it is not essential as part of the routine physical examination.

D. Alcohol Intake, Smoking and Sleep Apnea Prevalence

Table 5.8 summarizes the percentages of drivers with AHI≥5, ≥15, and ≥30 episodes per hour in relationship to smoking history and self-reported alcohol intake. Estimates were obtained as weighed averages from the higher and lower risk groups. As can be seen from examination of the data in this table, we do not find that drivers who currently smoked were at higher risk for sleep apnea. Similarly, drivers who drank >=2 alcoholic beverages per day were not at higher risk for sleep apnea. The small increases in the prevalence of AHI≥5 and AHI≥15 events/hour for those with any smoking did not appear statistically significant.

Table 5.8. Prevalence of different severities of apnea in individuals currently and not currently smoking, who ever versus never smoked, and with average daily drinks per day equal to >=2, 1-<2, and <1 drink per day. The data shown are the weighted prevalences from the higher and lower risk groups.

		AHI³5†	AHI³15†	AHI³30†
Current Smoking	**Yes**	28.1%	9.7%	2.3%
	No	28.0%	10.2%	5.2%
Any Smoking	**Yes**	30.8%	11.7%	4.8%
	No	24.2%	8.3%	4.0%
Alcohol Drinks / Day	**>=2**	30.5%	5.0%	1.1%
	1-<2	23.8%	4.3%	2.9%
	<1	31.3%	14.1%	6.8%
† Prevalence estimates obtained as weighted averages from higher and lower risk groups.				

E. Sleep Duration (from actigraphy)

There are some data, albeit far from comprehensive, that sleep deprivation makes sleep apnea worse [Stoohs and Dement, 1993]. It is postulated that sleep deprivation leads to greater reductions during sleep in the activity of the key upper airway dilator muscles maintaining airway patency, than would normally occur, thereby worsening the degree of sleep-disordered breathing. While this effect has been demonstrated in small studies, it has never been assessed in an epidemiological study of this type. We show in Tables 5.9a and 5.9b the prevalence of different severities of sleep apnea in individuals within different ranges of durations of the main bout of relative inactivity and the cumulative duration of inactivity in main sleep bout as measured from wrist activity monitoring at home. The data shown are for the weighted prevalence of the higher and lower risk groups.

Table 5.9a. Prevalence of different severities of apnea in individuals with different nightly durations of the main bout of relative inactivity. The latter is from the average of the measurements of sleep at home by actigraphy. The data shown are the weighted prevalences from the higher and lower risk groups. N (%) are of sample weights. The percentages represent estimates of the percentages of CDL holders from our population with the sleep durations equal to the indicated values. There is increasing prevalence of sleep apnea with shorter durations of the main bout of relative inactivity, particularly for moderate and severe sleep apnea (correlation χ^2=4.43, df=1, p=0.035).

Duration of Main Bout of Relative Inactivity	N (% of sample weights)	No Apnea (AHI<5)	Severity of Sleep Apnea		
			Mild (AHI≥5&<15)	Moderate (AHI≥15&<30)	Severe (AHI>30)
<6 hours/night	33.1 (9.8%)	68.2%	12.8%	12.8%	6.2%
6-<7 hours/night	72.0 (21.4%)	63.4%	20.4%	10.4%	5.7%
7-8 hours/night	153.4 (45.6%)	73.1%	19.8%	3.1%	4.0%
> 8 hours/night	78.4 (23.3%)	75.8%	17.2%	4.4%	2.6%

Table 5.9b. Prevalence of different severities of apnea in individuals with different nightly amounts of cumulative duration of inactivity in main sleep bout. The latter is from the average of the measurements of sleep at home by actigraphy. The data shown are the weighted prevalences from the higerh and lower risk groups. N (%) equals the sample weights. The percentages represent estimates of the percentages of CDL holders from our population with the sleep durations equal to the indicated values. There is increasing prevalence of sleep apnea with shorter amounts of cumulative duration of inactivity during the main bout, particularly for moderate and severe sleep apnea (correlation χ^2=32.5, df=1, p<0.001).

Duration of Cumulative Inactivity	N (% of sample weights)	No Apnea (AHI<5)	Severity of Sleep Apnea		
			Mild (AHI≥5&<15)	Moderate (AHI≥15&<30)	Severe (AHI>30)
<5 hours/night	45.7 (13.6%)	53.4%	15.2%	13.5%	18.0%
5-<6 hours/night	73.2 (21.7%)	60.7%	25.2%	8.9%	5.6%
6-<7 hours/night	101.0 (30.0%)	70.8%	23.1%	4.7%	1.4%
7-8 hours/night	95.7 (28.4%)	85.3%	11.8%	2.1%	0.7%
> 8 hours/night	21.3 (6.3%)	83.4%	13.4%	3.2%	0.0%

The data in Tables 5.9a and 5.9b are consistent with the hypothesis that shorter sleep durations are associated with a higher prevalence of sleep apnea. However, after also controlling for risk group using a Mantel Haenszel test, the correlation between apnea severity category and the duration of the main bout of relative inactivity lost its statistical significance (χ^2=1.5, df=1, p=0.215) while the correlation between sleep apnea severity and the cumulative duration of

inactivity during the main sleep bout retained its statistical significance (χ^2=12.9, df=1, p<0.001).

Table 5.10a demonstrates that for AHI≥5 and AHI≥15, the magnitude of association with the duration of the main bout of relative inactivity becomes slightly stronger when age and BMI (and gender) are controlled. The relative risk of AHI≥15 is approximately 3.9 (95% CI=0.9 to 17.4) times greater for those with main bouts less than 6 hours compared to those with more than 8 hours.

Risk of AHI≥5 and AHI≥15 was significantly associated with cumulative duration of inactivity during the main sleep bout (Table 5.10b). These associations are reduced but not eliminated when age and BMI (and gender) are controlled. The increased relative risk of AHI≥15 is still approximately eight times for cumulative durations less than 5 hours as compared to that between 7-8 hours even after controlling for age and BMI. An analogous approach could not be applied for AHI≥30 because of the small number of cases.

In conclusion, analyses of sleep duration using the conservative estimate of sleep duration (duration of the main bout of relative inactivity) are suggestive while analyses of sleep duration using the cumulative duration of inactivity during the main sleep bout revealed strong associations sleep apnea prevalence. While these data suggest that there is an increase in prevalence of sleep apnea in those with shorter sleep durations, we cannot determine whether this is likely to be a causative relationship as has been proposed [Stoohs and Dement, 1993], or rather a consequence of the disorder. It is conceivable that those with sleep apnea sleep less as a result of the sleep interruption that the disorder leads to.

Table 5.10a. Odds ratios and approximate 95% confidence intervals for any sleep apnea (AHI≥5) and for at least moderate sleep apnea (AHI≥15) for average nightly sleep duration as determined by actigraphy. The logistic regression model for the unadjusted models were for males only, were estimated based on sampling weights and contained only indicator variables for sleep duration using duration of major bout of relative inactivity. The adjusted estimates are from models that added age, age^2, BMI, and BMI^2. The confidence intervals are conservative since they ignore the stratified sampling which resulted in a design effects of 0.787 and 0.657, respectively for AHI≥5 and AHI≥15.

Average Nightly Duration of Major Bout of Relative Inactivity	AHI≥5		AHI≥15	
	Unadjusted	Adjusted	Unadjusted	Adjusted
<6 hours/night	1.4 (0.6-3.6)	1.5 (0.6-4.3)	3.5 (0.9-13.3)	3.9 (0.9-16.4)
6-<7 hours/night	1.7 (0.8-3.6)	1.8 (0.8-4.2)	2.7 (0.8-9.0)	2.7 (0.7- 9.7)
7-8 hours/night	1.2 (0.6-2.4)	1.3 (0.6-2.8)	1.2 (0.4-4.0)	1.3 (0.4- 4.4)
> 8 hours/night	Reference	Reference	Reference	Reference

Table 5.10b. Odds ratios and approximate 95% confidence intervals for any sleep apnea (AHI≥5) and for at least moderate sleep apnea (AHI≥15) for average nightly sleep duration as determined by actigraphy. The logistic regression model for the unadjusted models were for males only, were estimated based on sampling weights, and contained only indicator variables for duration of cumulative inactivity. The adjusted estimates are from models that added age, age^2, BMI, and BMI^2. The confidence intervals are conservative since they ignore the stratified sampling which resulted in a design effects of 0.787 and 0.657, respectively for AHI≥5 and AHI≥15.

Ave. Nightly Duration of Cumulative Inactivity	AHI≥5		AHI≥15	
	Unadjusted	Adjusted	Unadjusted	Adjusted
<5 hours/night	4.6 (2.0-10.4)	2.5 (1.004-6.1)	14.0 (3.6-54.7)	7.4 (1.8-31.0)
5-<6 hours/night	3.6 (1.7-7.5)	3.2 (1.4-7.3)	5.1 (1.3-20.2)	4.0 (0.9-17.0)
6-<7 hours/night	2.3 (1.1-4.7)	2.0 (0.9-4.4)	2.0 (0.4-8.7)	1.5 (0.3- 7.0)
7-8 hours/night	Reference	Reference	Reference	Reference
> 8 hours/night	0.9 (0.2-4.6)	1.2 (0.2-6.3)	1.7 (0.1-25.1)	2.5 (0.2-39.2)

E. Employment Status and Miles Driven/Year

Drivers who have sleep apnea may appreciate this and either seek other employment that does not require as high a level of vigilance or limit the amount of driving they perform. To assess this, we looked at prevalence of apnea in those currently employed at the time of our study as full-time commercial truck drivers, those currently employed driving full- or part-time any commercial vehicle, and those who had CDLs but no longer drove for a living. These data are shown in Table 5.11.

Table 5.11. Prevalence of different severities of sleep apnea in drivers with different employment statuses.

Employment	No Apnea (AHI<5)	Severity of Sleep Apnea		
		Mild (AHI≥5&<15)	Moderate (AHI ≥15&<30)	Severe (AHI ≥30)
Currently employed as full-time commercial truck driver	74.1%	16.9%	4.7%	4.3%
Currently employed as part-time or full-time as a driver of any commercial vehicle	73.2%	17.7%	5.0%	4.2%
Not currently employed as a commercial driver	66.2%	17.2%	9.3%	7.3%

As can be seen, the prevalence of severe and moderate sleep apnea is nearly twice as high in those who are no longer currently employed as commercial drivers but who have CDLs. The most parsimonious hypothesis to explain this result is that these drivers might have appreciated the difficulties they had and sought other employment. To assess this, we estimated the odds ratio and approximate 95% confidence interval for AHI≥15 using logistic regression on weighted

data comparing CDL holders currently not employed driving a commercial vehicle to those with current full- or part-time employment driving a commercial vehicle. The design effect resulting from the stratified sampling for AHI$\geq$15 is 0.657. Therefore, the confidence intervals are conservative. The unadjusted odds ratio (95% confidence intervals) was 2.0 (0.97 to 4.0) reflecting the doubling of the odds of at least moderate sleep apnea among unemployed CDL holders compared to those who were currently employed. A logistic regression model containing age, age^2, BMI, and BMI^2 resulted in an adjusted odds ratio (approximate 95% CI) of 1.7 (0.7-3.8). Thus, increased risk of at least moderate apnea appeared to be only partially explained by differences in age and BMI between unemployed and employed CDL holders. In contrast, we find no relationship between number of miles driven/year and sleep apnea prevalence in those who indicated that they were driving full-time as commercial truck drivers. These data are shown in Table 5.12.

Table 5.12. Prevalence of different severities of sleep apnea in drivers with numbers of miles driven among full-time commercial truck drivers.

Number of miles driver per year	No Apnea (AHI<5)	Severity of Sleep Apnea		
		Mild (AHI$\geq$5&<15)	Moderate (AHI $\geq$15&<30)	Severe (AHI>30)
70k-130k	73.9%	15.5%	5.3%	5.3%
30-<70k	69.3%	20.1%	6.4%	4.3%
15-<30k	75.0%	15.2%	6.7%	3.2%

For completeness, we provide prevalence estimates for different severities of sleep apnea for drivers with different employment status and driving history in Tables 5.13, 5.14, and 5.15. These provide prevalence estimates for AHI>=5, AHI>=15, and AHI>=30 episodes per hour for each of the driving status variables with and without stratification by the dominant risk factor variable, weight status (overweight versus not overweight). Overweight is defined by a BMI>=25 kg/m^2. In these tables, prevalence estimates for driving status variables are limited to full- or part-time drivers.

Thus, the numbers in these tables are prevalence estimates for sleep apnea at 3 severity levels broken up by four driving status variables. Prevalence estimates are provided for the higher risk group, the lower risk group, the weighted population estimate, the weighted population estimate restricted to overweight CDL holders, and the weighted population estimate restricted to non-overweight CDL holders.

Table 5.13. Prevalence of AHI>=5 episodes per hour by driving status and stratified whether or not the respondent was overweight (BMI>25 kg/m^2).

Status	Higher risk (n=247) %	Lower risk (n=159) %	Weighted Prevalence %	Overweight (n=274) n	Overweight (n=274) %	Not Overweight (n=130) n	Not Overweight (n=130) %
All	22.4	0.6	9.6				
Employment Status							
Currently employed full-time as a commercial truck driver	53.3	7.0	26.2	58	38.3	7	7.1
Currently employed part-time as a commercial truck driver	43.8	20.0	29.9	15	36.1	8	22.8
Currently not employed - full commercial truck driver	54.2	17.9	32.9	23	50.3	2	7.2
Currently employed as commercial vehicle driver	51.3	9.9	27.1	74	38.1	15	11.0
Currently not employed as commercial vehicle driver	53.2	18.5	32.9	22	49.6	2	7.6
Type of Driving %							
Over the road	46.2	0.0	19.2	3	19.1	1	8.8
Local	53.3	9.1	27.4	46	38.3	10	10.7
Both	46.7	18.5	30.2	23	41.6	4	14.0
Driving schedule %							
Only days	53.5	10.2	28.1	30	38.5	10	14.7
Only nights	40.0	16.7	26.4	4	42.5	0	0.0
Both	50.5	9.4	26.4	38	37.0	5	8.5
Number of miles driving a truck %							
70k-130k	50.8	10.3	27.1	26	36.2	2	7.2
30k-<70k	56.6	5.7	26.8	27	40.0	6	15.0
15-<30k	61.0	12.5	32.6	19	44.8	4	12.8
<15k	40.0	15.2	25.5	18	38.8	5	9.4
Notes							

% Includes only CDL holders indicating current employment as a commercial truck driver

Figure 5.14. Prevalence of AHI>=15 episodes per hour by driving status and stratified whether not the respondent was overweight (BMI>25 kg/m²).

Prevalence of Sleep Apnea by Employment and Driving Status Overall and by Weight Category

Status	Higher risk (n=247) %	Lower risk (n=159) %	Weighted Prevalence %	Overweight (n=274) n	%	Not Overweight (n=130) n	%
All	22.4	0.6	9.6				
Employment Status							
Currently employed full-time as a commercial truck driver	20.7	1.0	9.2	21	14.0	1	1.5
Currently employed part-time as a commercial truck driver	22.9	0.0	9.5	8	17.7	0	0.0
Currently not employed - full commercial truck driver	33.3	3.6	15.9	11	24.0	1	4.8
Currently employed as commercial vehicle driver	21.1	0.8	9.2	29	14.8	1	1.1
Currently not employed as commercial vehicle driver	34.0	3.7	16.3	11	24.4	1	5.1
Type of Driving %							
Over the road	15.4	0.0	6.4	1	9.6	0	0.0
Local	20.5	1.1	9.2	17	14.1	1	1.6
Both	23.3	0.0	9.7	10	17.4	0	0.0
Driving schedule %							
Only days	16.3	0.0	6.8	10	12.1	0	0.0
Only nights	20.0	0.0	8.3	1	13.7	0	0.0
Both	26.3	1.6	11.8	18	17.3	1	2.5
Number of miles driving a truck %							
70k-130k	27.0	2.6	12.7	13	18.5	0	0.0
30k-<70k	22.4	0.0	9.3	12	17.4	0	0.0
15-<30k	34.1	3.1	16.0	9	21.3	2	6.4
<15k	11.1	0.0	4.6	3	6.0	1	1.3

Table 5.15.   Prevalence of AHI>=30 episodes per hour by driving status and stratified whether or not the respondent was overweight (BMI>25 kg/m^2).

Status		Higher risk (n=247) %	Lower risk (n=159) %	Weighted Prevalence %	Overweight (n=274) n	Overweight (n=274) %	Not Overweight (n=130) n	Not Overweight (n=130) %
All		22.4	0.6	9.6				
Employment Status	Currently employed full-time as a commercial truck driver	10.7	0.0	4.4	11	7.2	0	0.0
	Currently employed part-time as a commercial truck driver	8.3	0.0	3.5	3	6.4	0	0.0
	Currently not employed - full commercial truck driver	16.7	0.0	6.9	5	10.4	1	2.4
	Currently employed as commercial vehicle driver	10.1	0.0	4.2	14	7.0	0	0.0
	Currently not employed as commercial vehicle driver	17.0	0.0	7.1	5	10.5	1	2.5
Type of Driving %	Over the road	7.7	0.0	3.2	1	4.8	0	0.0
	Local	6.6	0.0	2.7	5	4.5	0	0.0
	Both	16.7	0.0	6.9	7	12.4	0	0.0
Driving schedule %	Only days	4.7	0.0	1.9	3	3.5	0	0.0
	Only nights	10.0	0.0	4.2	1	6.9	0	0.0
	Both	15.2	0.0	6.3	10	10.0	0	0.0
Number of miles driving a truck %	70k-130k	12.7	0.0	5.3	5	7.7	0	0.0
	30k-<70k	9.2	0.0	3.8	5	7.1	0	0.0
	15-<30k	14.6	0.0	6.1	3	8.2	1	2.0
	<15k	6.7	0.0	2.8	2	4.5	0	0.0
Notes	% Includes only CDL holders indicating current employment as a commercial truck driver							

5.5 Sleep Apnea Symptoms and Presence of Sleep Apnea

An important issue for policy makers is how predictive the presence of sleep apnea symptoms are for presence of sleep apnea of different severities since the presence of these symptoms can be evaluated at routine medical examinations of drivers. As described in Section 5.1 for the population survey respondents, the sleep questionnaire included questions concerning the frequency of symptoms known to be associated with sleep apnea [Maislin et al, 1995]. These are loud snoring, any snoring, snorting and gasping, and breathing stops and/or struggling for breath. Tables 5.16 to 5.19 provide the estimated prevalence of severe, moderate, and mild sleep apnea with respect to subjects with different complaints for those drivers who had in-lab studies. Prevalence estimates were determined using sample weights. The percentage of total sample weights with each symptom frequency is also provided. These percentages may be compared with those in Section 5.1 to assess the representativeness of the in-lab sample compared to the population survey sample. Inspection revealed very similar apnea prevalence for subjects reporting "never" and "rarely, less than once per week" and so these categories were combined for ease of presentation. We see a marked increase in the prevalence of severe sleep apnea (i.e., AHI$\geq$30 episodes/hours) for subjects with loud snoring 3-4 and 5-7 times per week. The prevalence values were between 2.5 and 3.5 times larger than subjects reporting frequencies of loud snoring <1/week or 1-2 times per week. A similar finding was seen for reports of any snoring. Reports of very frequent (i.e. $\geq$3 times per week) snorting and gasping and breathing stops, struggling for breath were associated with even larger increased risk for severe sleep apnea, but the overall prevalence of these symptoms was substantially less.

Table 5.16. Prevalence of different severities of sleep apnea in drivers with average weekly frequency of apnea symptoms – Loud snoring. Estimated percentages are based on sample weights.

Frequency	% †	AHI (Episodes/hour)			
		³30	15-<30	5-<15	<5
< 1 / week	51.9	3.9%	6.6%	15.3%	74.3%
1-2 / week	17.5	2.9%	4.8%	16.0%	76.4%
3-4 / week	9.9	10.2%	8.5%	31.2%	50.1%
5-7 / week	12.5	9.8%	4.2%	23.9%	62.1%
Don't know/missing	8.2	0.0%	2.0%	10.2%	87.7%
† Percentage of total sample weights reporting frequency of symptoms					

Table 5.17. Prevalence of different severities of sleep apnea in drivers with average weekly frequency of apnea symptoms – Any snoring. Estimated percentages are based on sample weights.

Frequency	%†	AHI (Episodes/hour)			
		³30	15-<30	5-<15	<5
< 1 / week	59.0	3.1%	5.5%	14.6%	76.2%
1-2 / week	15.2	2.1%	6.6%	18.1%	73.1%
3-4 / week	8.8	13.4%	3.8%	32.9%	49.9%
5-7 / week	9.1	9.8%	9.8%	27.8%	52.7%
Don't know/missing	7.9	6.4%	4.3%	10.7%	78.7%
† Percentage of total sample weights reporting frequency of symptoms					

Table 5.18. Prevalence of different severities of sleep apnea in drivers with average weekly frequency of apnea symptoms – Snorting or gasping. Estimated percentages are based on sample weights.

Frequency	%†	AHI (Episodes/hour)			
		³30	15-<30	5-<15	<5
< 1 / week	77.8	4.1%	5.7%	15.1%	75.1%
1-2 / week	6.7	2.5%	5.0%	36.3%	56.2%
3-4 / week	4.6	14.7%	11.0%	22.1%	52.2%
5-7 / week	2.0	25.0%	8.3%	25.0%	41.7%
Don't know/missing	8.9	2.0%	4.0%	22.1%	71.9%
† Percentage of total sample weights reporting frequency of symptoms					

Table 5.19. Prevalence of different severities of sleep apnea in drivers with average weekly frequency of apnea symptoms – Breathing stops or struggle for breath. Estimated percentages are based on sample weights.

Frequency	%†	AHI (Episodes/hour)			
		³30	15-<30	5-<15	<5
< 1 / week	86.9	4.5%	5.1%	15.9%	74.6%
1-2 / week	3.6	4.7%	14.0%	33.6%	47.7%
3-4 / week	1.0	16.7%	50.0%	16.7%	16.7%
5-7 / week	0.7	75.0%	0.0%	25.0%	0.0%
Don't know/missing	7.9	0.0%	4.5%	29.2%	66.3%
† Percentage of total sample weights reporting frequency of symptoms					

5.6 Relationship Between Calculated Likelihood of Sleep Apnea and Presence of Sleep Apnea

Equally important for policy makers is whether the calculated likelihood of sleep apnea by the multivariable apnea prediction (MAP) index [Maislin et al, 1995] provides useful information that could be used by physicians evaluating commercial drivers for fitness for employment. We show in Table 5.20 the prevalence of sleep apnea of different severities in categories defined by decile of the multivariable apnea prediction (MAP).

Table 5.20. Prevalence of different severities of sleep apnea as a function of the Multivariable Apnea Prediction (MAP) index. Estimated percentages are based on sample weights.

MAP Score	Frequency %†	AHI (Episodes/hour)			
		$\geq$30	15-<30	5-<15	<5
0.0 to < 0.1	2.6	0.0%	0.0	0.0%	100.0%
0.1 to < 0.2	11.2	0.0%	0.0%	0.0%	100.0%
0.2 to < 0.3	16.4	0.0%	3.3%	6.7%	90.0%
0.3 to < 0.4	15.4	1.1%	1.1%	12.8%	85.0%
0.4 to < 0.5	15.5	4.4%	4.4%	15.1%	76.2%
0.5 to < 0.6	15.2	2.3%	7.0%	31.3%	59.4%
0.6 to < 0.7	10.7	6.3%	15.8%	31.9%	46.1%
0.7 to < 0.8	8.6	19.9%	14.0%	31.9%	34.2%
0.8 to < 0.9	3.9	17.4%	8.7%	34.8%	39.1%
0.9 to ≤1.0	0.7	75.0%	25.0%	0.0%	0.0%
† Percentage of total sample weights reporting frequency of symptoms					

As can be seen, there is a strong relationship between the value of the multivariable apnea prediction and the prevalence of sleep apnea. In particular, the instrument has high negative predictive value. No driver with a MAP of under 0.215 had any degree of sleep apnea. Based on in-lab sample weights, we estimate that 16.3% of our population had MAP values below 0.215. The relationship to more severe sleep apnea is shifted to even higher values of MAP. No subject with a MAP value of less than 0.323 (34.9% of the total population) had an AHI$\geq$30 episodes/hour while only 0.59% with a MAP value of less than 0.47 (57.2% of the population) had severe sleep apnea (i.e., the negative predictive value for MAP<0.47 predicting no severe sleep apnea was 99.4%, a prediction applicable to 57.2% of our population).

This property of the MAP is generalizeable, i.e., not specific to this population of commercial drivers. We have performed a similar analysis on a population of 359 patients seeking evaluation at our sleep disorders center. The results are shown in Figure 5.16. The results are almost identical to those reflected by Table 5.20 for the commercial driver population. Thus, we propose that the MAP will have utility in evaluating commercial drivers at routine medical examinations. If the value of MAP is less than 0.215, we can exclude with reasonable certainty any sleep apnea; if the value is less than 0.323, we can exclude severe sleep apnea, and a value less than 0.47 indicates that it is very unlikely that the individual has severe sleep apnea.
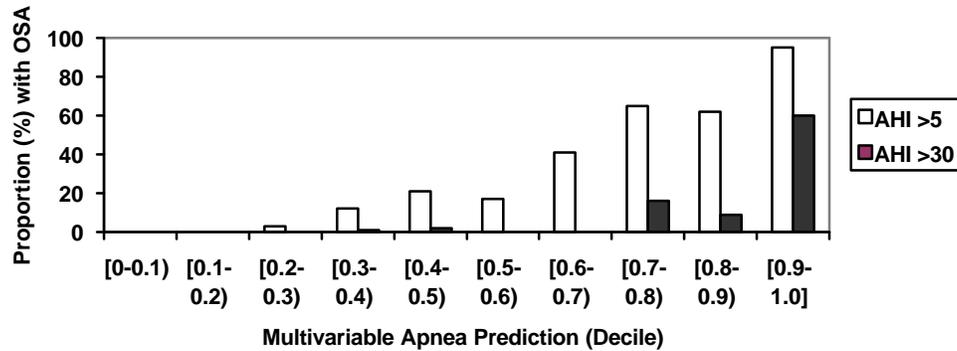
Figure 5.16. Prevalence of sleep apnea of different severities in relationship to the multivariable apnea prediction. The data are for a population of 359 patients being evaluated at our sleep disorders center for clinical purposes.

5.7 **Prevalence of Sleep Apnea with Different Definitions of Respiratory Events**

As outlined in Chapter Two, prevalence of different degrees of sleep apnea depends on the definitions used to define a respiratory event [Redline et al, 2000]. For studies utilizing, as we did, thermistors to measure airflow, the major difference between different definitions depends on the degree of desaturation (2 or 3 or 4 percent drops in oxygen saturation) and whether identification of an arousal on the EEG etc. is, or is not, included in the definition. While the American Academy of Sleep Medicine has proposed one definition [1999] that we have used throughout our analyses, we provide in Appendix Tables H.1 to H.6, prevalence of different levels of severity when different definitions of hypopnea are used. We provide this information for completeness and to allow comparison of our results on prevalence to those reported in other epidemiological studies. The data shown are for the higher risk group, the lower risk group and for the weighted population prevalence estimate. The standard errors (SE) and 95% confidence intervals taking into account the sampling design are also provided. Tables with hypopneas defined using as decreased breathing associated with either a 4%, 3%, and 2% desaturation or an arousal are provided, respectively, in Tables H.1, H.2, and H.3. Tables H.4, H.5, and H.6 give similar data but where we excluded those hypopneas solely defined by the presence of arousals and give data where events were associated with different degrees of desaturation (drops in oxygen by 2%, 3% and 4%).

5.8 **Prevalence of Sleep Apnea Syndrome of Different Severities**

As indicated in Chapter Two, not all subjects with increased levels of respiratory disturbance during sleep have complaints of excessive daytime sleepiness. This has led to the concept of the sleep apnea/hypopnea syndrome, i.e., when individuals have both respiratory disturbance during sleep and are excessively sleepy (see Section 2.3, Chapter Two). While this concept is established, the precise methodology to operationalize this syndrome is unclear. There are issues as to whether this should be based on self-reported sleepiness or objectively measured sleepiness. As it transpires (see Chapters Six and Seven), this is a particular issue for commercial drivers since there is no relationship between self-reported sleepiness and degree of respiratory disturbance during sleep, while there is a clear relationship with objectively measured sleepiness. Nevertheless, current convention is to base the definition of the syndrome on self-

reported sleepiness [American Academy of Sleep Medicine, 1999]. Thus, we report here prevalence of the syndrome based on this to allow comparison to previous reports. We utilized the Epworth Sleepiness Scale [Johns, 1991] to assess self-reported excessive sleepiness (see Appendix J.4). This scale is based on subjects indicating how likely they are to doze in eight defined situations. The overall range of the score is from zero (never likely to doze in any of the eight situations) to 24 (very likely to doze in all eight situations). We defined, based on previous literature, excessive sleepiness to be present when the score on the Epworth Sleepiness Scale is 10 or above [Johns, 1991]. This definition led to the prevalence estimates of sleep apnea/hypopnea syndrome for our sample shown in Table 5.21.

Table 5.21. Prevalence of sleep apnea/hypopnea syndrome of different levels of severity in our sample.

AHI[&]	Higher Risk (n=229)					Lower Risk (N=156)					Weighted			
	N	Rate	SE	(95%CI)		N	Rate	SE	(95%CI)		Rate	SE^	(95%C.I.)	
None	181	0.790	0.011	0.768	0.812	148	0.949	0.010	0.929	0.969	0.883	0.015	0.853	0.913
5 - <15	24	0.105	0.008	0.089	0.121	7	0.045	0.010	0.026	0.064	0.070	0.013	0.045	0.095
15 - <30	14	0.061	0.007	0.048	0.074	1	0.006	0.004	-0.001	0.013	0.029	0.007	0.014	0.044
>= 30	10	0.044	0.006	0.033	0.055	0	0.000	0.000	0.000	0.000	0.018	0.006	0.007	0.029

[&] Apnea/hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.
  The category indicated as 'None' includes subjects where AHI<5 OR Epworth Sleepiness Scale <10.
[&] Weighted percent computed as (0.415*Higher risk percent) + (0.585*Lower risk percent).
  Weights determined at the population level.
^ Weighted standard error (SE) computed as square root of $(0.415)^2 * (Higher\ risk\ SE)^2 + (0.585)^2 * (Lower\ risk\ SE)^2$.

AHI computed as total apneas plus hypopneas per hour of sleep. Hypopneas were defined as >=3% desaturation + airflow arousal hypopneas. Syndrome definition required AHI above the specified level plus Epworth Sleepiness Scale >=10. Data are shown for both higher and lower risk groups and for the weighted average.

The prevalence of the syndrome is, as expected, considerably lower than that of the respiratory disturbance during sleep alone, without consideration of whether self-reported sleepiness is present or not (compare data in Table 5.21 with that in Table 5.5). We found that 1.8% of all CDL holders had severe sleep apnea syndrome; 2.9% moderate or severe; and 7.0% at least mild sleep apnea syndrome. As with sleep apnea, the prevalence of the syndrome was dependent on age and BMI. Prevalence increases with age and with increasing obesity. Table 5.22 provides estimates of the prevalence of sleep apnea syndrome stratified by age and overweight status. Data in Table 5.22 should be compared to that in Table 5.5 where similar data are shown for sleep apnea itself, i.e., irrespective of whether the CDL holders complained of excessive sleepiness.

Table 5.22. Estimated prevalence of sleep apnea syndrome[%&] by age and BMI category for various apnea severities. Syndrome definition required AHI above the specified level plus Epworth Sleepiness Scale ≥10.

Age	Overweight	Yes	No	Total
20-29	PSG's[&]	14	6	20
	AHI>=5	0.000	0.185[%]	
	AHI>=10	0.000	0.000	
	AHI>=15	0.000	0.000	
	AHI>=30	0.000	0.000	
30-39	PSG's	56	39	95
	AHI>=5	0.145	0.009	
	AHI>=10	0.093	0.000	
	AHI>=15	0.080	0.000	
	AHI>=30	0.027	0.000	
40-49	PSG's	91	38	129
	AHI>=5	0.163	0.000	
	AHI>=10	0.097	0.000	
	AHI>=15	0.043	0.000	
	AHI>=30	0.022	0.000	
50-59	PSG's	62	22	84
	AHI>=5	0.292	0.166	
	AHI>=10	0.211	0.050	
	AHI>=15	0.164	0.050	
	AHI>=30	0.049	0.000	
>=60	PSG's	39	17	56
	AHI>=5	0.189	0.043	
	AHI>=10	0.135	0.000	
	AHI>=15	0.108	0.000	
	AHI>=30	0.081	0.000	
	Total	262	122	384

Notes:

[%] Prevalence estimated as weighted average of higher and lower risk group prevalences. Weights were determined within age and BMI category from population-based screen (n=1329).

[&] Number of subjects (pooling higher and lower risk groups) with polysomnography within age and BMI strata.

5.9 Determinants of the Prevalence of the Sleep Apnea/Hypopnea Syndrome

We carried out identical analyses for risk factors for the prevalence of the syndrome as we reported above for sleep apnea (Section 5.4). We show the results of these analyses for the interested reader in Appendix I. The results parallel those described above for sleep apnea itself. As we reported above, the prevalence of the syndrome is greater in males than females; in males, the prevalence increases with increasing age and increasing body mass index. There is a 'public-health' interaction (i.e., non-additive association) between the latter such that the effect of increasing BMI is more marked in older subjects because of the increased risk inherent in older individuals regardless of weight. The data allow us to provide policy makers and owners of truck companies with analogous prediction equations to those described above for apnea for the

prevalence in males of sleep apnea/hypopnea syndrome of different severities (see Appendix I Tables I.1 and I.2).

5.10 Prevalence of Sleep Apnea/Hypopnea Syndrome in Relationship to Likelihood of Apnea

An important result that we reported above was the prevalence of sleep apnea at various levels of the multivariable apnea prediction (MAP) (see Table 5.20). We found a similar result when we examined the relationship between prevalence of the syndrome at different levels of severity and MAP (see Table 5.23). Thus, although the prevalence of sleep apnea syndrome is lower in this population compared to sleep apnea defined on the basis of disturbed breathing during sleep alone, the general associations between prognostic factors and presence of disease are similar. However, the much lower prevalence of sleep apnea syndrome defined on the basis of self-report of sleepiness must be interpreted in light of the findings in Chapter Six. There, we report that, unlike in other populations, individuals in our population with sleep apnea do not report increased sleepiness compared to those with no sleep apnea (notwithstanding that subjects with sleep apnea did have reduced function when measured objectively).

Table 5.23. Prevalence of different severities of sleep apnea syndrome as a function of the multivariable apnea prediction (MAP) index. Estimated percentages are based on adjusted sample weights (Total of weights is 389).

MAP Score	Frequency % †	AHI (episodes/hour)			
		³30	15-<30	5-<15	<5
0.0 to < 0.1	2.6	0.0%	0.0%	0.0%	100.0%
0.1 to < 0.2	10.9	0.0%	0.0%	0.0%	100.0%
0.2 to < 0.3	16.6	0.0%	2.3%	4.5%	93.2%
0.3 to < 0.4	14.9	0.0%	0.0%	2.5%	97.5%
0.4 to < 0.5	15.6	0.0%	3.5%	6.0%	90.5%
0.5 to < 0.6	15.4	0.0%	2.4%	8.3%	89.3%
0.6 to < 0.7	10.7	3.4%	8.5%	13.6%	74.5%
0.7 to < 0.8	8.7	8.3%	6.2%	16.6%	68.8%
0.8 to < 0.9	4.0	13.6%	4.6%	18.2%	63.6%
0.9 to ≤1.0	0.5	33.0%	0.0%	0.0%	66.7%
† Percentage of total sample weights reporting frequency of symptoms					

5.11 Summary and Discussion

Our studies provide estimates of the prevalence of sleep apnea and sleep apnea/hypopnea syndrome of different severities. We report that in our population of holders of commercial drivers licenses mild sleep apnea (AHI≥5&<15 episodes/hour) occurs in 17.6% (95% confidence interval (CI): 7.9% to 27.3%); moderate sleep apnea (AHI≥15&<30 episodes/hour) in 5.8% (95% CI: 0.0% to 16.1%); and severe sleep apnea (AHI≥30) in 4.7% (95% CI: 0.0% to 9.6%).

Sleep apnea/hypopnea syndrome is less common, i.e., when we also require complaints of excessive daytime sleepiness. The comparable prevalence numbers are: mild sleep

apnea/hypopnea syndrome (7.0%, 95% CI: 4.5% to 9.5%); moderate sleep apnea/hypopnea syndrome (2.9%, 95% CI: 1.4% to 4.4%); and severe sleep apnea/hypopnea syndrome (1.8%, 95% CI: 0.7% to 2.9%).

Both sleep apnea and sleep apnea/hypopnea are more common in males than females. With males, which make up most of this sample, the prevalence of both sleep apnea and the syndrome increase with age and with increasing obesity. These effects appear to be multiplicative, that is, the effect of increasing obesity has more marked effects on prevalence rates in older as compared to younger individuals. We provide prediction equations that allow calculation of estimates of prevalence of sleep apnea of different severities, and of the sleep apnea syndrome based on knowledge of age and body mass index (BMI). These equations could be used by owners of trucking companies to estimate how many drivers they will have with different levels of severity of sleep apnea if they know the ages and BMI values of their drivers.

We also report suggestive evidence for a relationship between average sleep duration at home and the demonstrated prevalence of sleep apnea or sleep apnea/hypopnea syndrome. Drivers with shorter duration of "sleep" have a higher prevalence of the disorder. This might be related to the hypothesized effects of sleep deprivation in augmenting the degree of sleep apnea [Stoohs and Dement, 1993]. Alternatively, sleep apnea might result in shorter durations of sleep. Whatever the explanation, shorter sleep durations in those with apnea will augment the decrements in daytime performance that result from the sleep fragmentation caused by sleep apnea as we discuss fully in Chapter Seven.

The prevalence of at least moderate sleep apnea was higher in those holders of CDLs who no longer drove for a living, compared to those still employed full-time as commercial drivers. While the explanation of this is unknown, we hypothesize that some drivers with sleep apnea may realize that they have a problem and seek other employment. This means that the prevalence of sleep apnea is slightly lower in those who continue to drive for a living than in the total population of CDL holders. In full-time commercial truck drivers we report the following prevalence rates: mild sleep apnea/hypopnea (16.9%); moderate sleep apnea/hypopnea (4.7%); and severe sleep apnea/hypopnea (4.3%). Among those not currently employed, these values were 17.2%, 9.3%, and 7.3%, respectively.

These estimates of prevalence, both for all CDL holders and for full-time commercial truck drivers, are related to the likelihood of apnea as detected by the multivariable apnea prediction (MAP), based on BMI, the presence and frequency of symptoms of apnea, as well as the age and gender. This likelihood has relative values between zero and one. Our results show that one can exclude with certainty any sleep-disordered breathing in individuals with a MAP under 0.215. This is 16.3% of our population of CDL holders and 18.5% of subpopulation of full-time commercial truck drivers. Similarly, one can exclude severe sleep apnea with a MAP value of less than 0.323, which includes 34.9% of our population of CDL holders and 36.3% of subpopulation of full-time commercial truck drivers. Furthermore, the negative predictive value for severe sleep apnea was 99.4% for MAP values less than 0.47, a range of values possessed by 57.2% of the total population. Among those with full-time employment, the negative predictive value for MAP values less than 0.47 was 99.5% with 58.2% of those employed full-time having

MAP values less than 0.47. This instrument has utility, therefore, in being able to determine that many drivers will not have sleep apnea and, in particular, not have severe sleep apnea.

Finally, the estimates of prevalence of sleep apnea and sleep apnea/hypopnea syndrome we report are of a similar magnitude to that reported in other large epidemiological studies in subjects in this age range [Young et al, 1993; Bearpark et al, 1995; Bixler et al, 1998]. The prevalence is slightly higher in holders of CDLs which is, we believe, the result of the high prevalence of obesity in holders of CDLs. The data we report in Table 5.3 and Table 5.4 should be compared to that in Table 2.3. We do not replicate the extremely high prevalence of sleep apnea reported by Stoohs et al [1995]. The reason for the high prevalence in the study of Stoohs et al [1995], which is a clear outlier, is unexplained. Thus, our results show that sleep apnea is a significant issue for commercial drivers but it does not affect as many drivers as was reported in the study of Stoohs et al [1995].

CHAPTER SIX

Subjective Measures of Sleepiness

6.1 Introduction

All in-laboratory subjects were assessed using multiple instruments with regard to their subjective perception of sleepiness and functional impairment, as well as with regard to objective measures of sleepiness, lack of attention, and other potential functional consequences of sleepiness. Chapter Six reports on results from the subjective assessments and introduces statistical methods used in analyses of both subjective and objective measures. Chapter Seven reports on results from the objective assessments. In general, we found no associations between sleep apnea severity and 'subjective sleepiness' in our population of CDL holders. These results differ from several studies that have found significant differences between individuals with sleep apnea and normal controls relative to subjectively perceived sleepiness. These studies include those that employed the Epworth Sleepiness Scale [Johns, 1991, 1993] and the Functional Outcomes of Sleep Questionnaire [Weaver et al, 1997], both of which were used in our studies. In contrast, in our study, clinically important and statistically significant associations were found between apnea severity and objective measurements, even after controlling for the mean **main bout length of relative inactivity** category or the mean **cumulative duration of inactivity in main sleep bout** category plus demographic characteristics. These contrary results have potentially very important implications with regard to commercial vehicle drivers' ability to recognize their own functional impairment resulting from severe sleep apnea as well as from reduced sleep durations. These results will be discussed in more detail in Chapter Seven.

Subjective assessments were made either once during intake interviews (e.g., for the Epworth Sleepiness Scale [Johns, 1991, 1993] and the Functional Outcomes of Sleep Questionnaire [Weaver et al, 1997]) or multiple times during a day of testing following the overnight polysomnography (e.g., Karolinska Sleepiness Scale [Akerstedt and Gillberg, 1990] and Stanford Sleepiness Scale [Hoddes et al, 1973]. (These various instruments are in Appendix J.)

In the next four sections, the associations between the Epworth Sleepiness Scale, the Karolinska Sleepiness Scale, the Stanford Sleepiness Scale, and the Functional Outcomes of Sleep Total Score with both sleep apnea severity and indices related to mean sleep duration are examined. Copies of these various instruments are given in Appendix J. Two-way analyses of variance (ANOVA) were used in order to account for sampling by risk group. Overall differences between mean values as well as linear and quadratic trends in mean values were assessed for both sleep apnea severity and for the sleep duration related indices (i.e., main bout length of relative inactivity and cumulative duration of inactivity). Following these analyses, multivariable models are introduced in which the effects of sleep apnea severity, sleep duration related indices as well as age, gender, and obesity are simultaneously assessed. Sampling weights were used in these analyses so that results better reflect populations associations. Specifically, the *a priori* model included age, BMI, female gender, apnea severity (>=30 vs. <5; 15-<30 vs. <5; and 5-<15 vs. 5 episodes per hour), either mean main bout length of relative inactivity (<6 hours vs. >8; 6-<7 vs. >8; and 7-8 hours vs. >8 hours) or mean cumulative

duration of inactivity during the main bout (<5 hours vs. 7-8; 5-<6 vs. 7-8; 6-<7 vs. 7-8; and >8 vs. 7-8 hours) plus a simple self-report health-related quality of life (health-related QoL) score measured on a 6 point scale (1=Perfect health to 6=Miserable). Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a male subject who is 45.5 years old, has a body mass Index of 29.9 kg/m^2, a health-related QoL index value of 2.44 points (scale range 1 to 6), without sleep apnea and without reduced mean duration of main sleep bout (or reduced mean cumulative duration of inactivity in main sleep bout). The health-related QoL measure was included in the model to better distinguish between general perceptions related to overall health and those specifically related to sleepiness and sleepiness-induced reduced function. These analyses also included assessment of the impact of frequent snoring, having more than 2 alcoholic beverages per day, and current smoking status. To facilitate comparisons between the effects of these factors on subjective and objective sleepiness and functional measurement scales, results from analogous multiple linear regression models are presented in Chapter Seven for the objective measures. Finally, for selected parameters, analyses were performed in the subsample of subjects without significant sleep apnea (i.e., AHI<5 events/hr). These analyses focused on assessing the association between subjective sleepiness and shortened cumulative durations of inactivity during the main sleep bout.

6.2. Epworth Sleepiness Scale

A. Description of Scale

The Epworth Sleepiness Scale (ESS) is a simple self-administered eight-item questionnaire measuring the general level of daytime sleepiness, or average sleep propensity, in adults. Subjects are asked how likely they are to doze in a number of situations [Johns, 1991]. A cumulative score is calculated. A score above 10 is considered indicative of pathological sleepiness [Johns, 1994]. The conceptual basis of the ESS involves a process model of sleep and wakefulness. The sleep propensity at any particular time is a function of the ratio of the total sleep drive to the total wake drive with which it competes. In the seminal study describing the ESS, Johns [1991] provided results based on 180 adults including 30 normal men and women as controls and 150 patients with a range of sleep disorders. They rated the chances that they would doze off or fall asleep when in eight different situations commonly encountered in daily life. Total ESS scores significantly distinguished normal subjects from patients in various diagnostic groups of disorders that lead to excessive sleepiness, including obstructive sleep apnea syndrome, narcolepsy and idiopathic hypersomnia. ESS scores were significantly correlated with sleep latency measured during the multiple sleep latency test and during overnight polysomnography. In patients with obstructive sleep apnea syndrome, ESS scores were significantly correlated with the respiratory disturbance index and the minimum oxygen level recorded during overnight polysomnography. ESS scores of patients who simply snored did not differ from controls. The ESS, unlike the Stanford Sleepiness Scale and the Karolinska Sleepiness Scale, seeks to determine how sleepy individuals are on a frequent basis over a prolonged period of time.

The ESS has been shown to have good reliability and internal consistency [Johns, 1992]. Johns [1992] presented results demonstrating: (1) that when 87 healthy medical students were

tested and retested 5 months later, their paired ESS scores did not change significantly and were highly correlated (r = 0.82). By contrast, ESS scores that were initially high in 54 patients suffering from obstructive sleep apnea syndrome returned to more normal levels, as expected, after 3-9 months' treatment with nasal continuous positive airway pressure; (2) ESS had a high level of internal consistency as measured by Cronbach's alpha (0.88); and (3) factor analysis of item scores showed that the ESS had only one factor, i.e., excessive sleepiness, when data from 104 medical students or from 150 patients with various sleep disorders were analyzed.

Johns [1993] also showed that ESS scores significantly distinguished patients with primary snoring from those with obstructive sleep apnea syndrome (OSAS), and that ESS scores increased with the severity of OSAS. Multiple regression analysis showed that ESS scores were more closely related to the frequency of apneas than to the degree of hypoxemia in OSAS.

More recently, Stutts et al [1999] used Epworth Sleep Scale categories defined as 0-5 (none or mild), 5-10 (moderate sleepiness), 11-15 (heavy sleepiness), and 16+ (extreme sleepiness) to demonstrate associations between the level of daytime sleepiness as measured by the Epworth Sleepiness Scale as a factor in sleep-related crashes, adjusted for driver age and gender. This study involved comparisons of Epworth Sleepiness scores among three groups of drivers: (a) those judged by the police to have a fall-sleep crash (sleepy crash drivers); (b) drivers who had a crash but not judged to be due to falling asleep driving (non-sleepy crash drivers); and (c) drivers renewing their licenses who had not had a crash in the previous three years (non-crash drivers). Odds ratios (95% confidence intervals) for comparing ESS in sleepy crash drivers vs. non-sleepy crash drivers of 1.43 (1.08 to 1.91), 2.95 (1.97 to 4.42), and 5.79 (2.27 to 14.72) for Epworth values of 6-10, 11-15, and 16+ were observed, respectively, relative to that in the reference category defined as ESS<5. Similarly, comparing sleepy crash drivers vs. non-crash drivers, odds ratios (95% CI) of 1.34 (0.92 to 1.96), 4.20 (2.38 to 7.43), and 15.18 (3.17, 72.78) for Epworth values of 6-10, 11-15, and 16+, respectively compared to <5 were reported. Thus, an Epworth Sleepiness Score above 10 is associated with a large increase in the risk of a fall-sleep crash.

B. Distribution of ESS Values in the Higher and Lower Risk Groups

Figure 6.1 displays the mean and standard deviation of ESS values for the higher and lower risk groups. The difference in mean values between risk groups was not statistically significant (t=1.26, df=384, p=0.209). The population mean value was estimated as the weighted average of within group means and was found to be equal to 8.78. The usual pooled estimate of the standard deviation (i.e., square-root of weighted variances with weights determined by degrees-of-freedom) was 4.45. Thus, our estimates for the mean and SD Epworth Sleepiness Score in our population of CDL holders were 8.78 and 4.45, respectively. Epworth total scores greater than 10 are considered indicative of excess sleepiness (see above). In the higher risk group, 77 of 229 (33.6%) had values greater than 10 while in the lower risk group, 50 of 157 (31.9%) had values greater than 10. These percentages were not statistically significantly different (χ^2=0.133, df=1, p=0.715). The weighted estimate for the percentage of individuals in our population of CDL holders who had an Epworth Sleepiness Score value greater than 10 is 32.6%. As noted above, this score in the self-report measure of sleepiness has been shown in

passenger car drivers to be associated with a three- to four-fold increased risk of a fall-asleep crash.
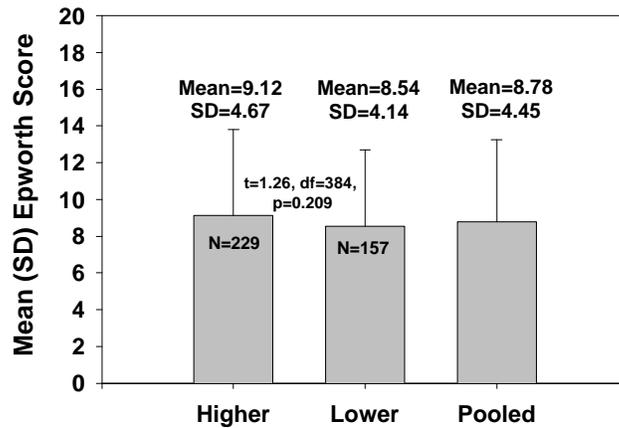
Figure 6.1. Epworth Sleepiness Scale mean and standard deviation for the higher and lower risk groups. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal true within risk group SD's and is computed using the usual formula for computing a pooled SD estimate.

C. Association with Sleep Apnea Severity

Table 6.1 summarizes the mean (SD) Epworth Sleepiness Scale values by sleep apnea severity for subjects in the higher and lower risk groups. The mean values were similar across sleep apnea severity levels in both risk groups.

Table 6.1. Epworth sleepiness scale values by severity of sleep apnea for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	110	8.9	4.7	0	20	138	8.6	4.2	0	20
5 - <15	63	8.8	4.0	2	17	16	8.0	3.9	0	13
15 - <30	30	10.0	4.2	3	19	2	8.5	2.1	7	10
>= 30	25	9.8	6.3	1	23					
Total N	228					156				

Two-way analysis of variance (ANOVA) was used to assess if there were significant differences among the sleep apnea severity levels for mean Epworth Sleepiness Scale controlling for risk group. There were no significant differences (F=0.76, df=3,379, p=0.518). A more statistically powerful (single degree-of-freedom) test was obtained by specifying a linear trend in mean values in the alternative hypothesis. There was no evidence of a linear trend in Epworth values as a function of sleep apnea severity (F=1.45, df=1,357, p=0.378). Results from the

ANOVA are summarized in Table 6.2. There were no significant differences among mean values by apnea severity or by risk group.

Table 6.2. Two-way ANOVA (apnea severity and risk group) for Epworth Sleepiness Scale.

Test	F value	df		P value	Pair-wise t-test p<0.05
AHI (Overall)	0.76	3	379	0.518	
Risk Group	0.78	1	379	0.378	
AHI Linear Trend	1.45	1	379	0.229	
AHI Quadratic Trend	0.01	1	379	0.925	
Risk Group * AHI Category Interaction	0.12	2	377	0.888	

D. Association with Prior Week's Mean Sleep Duration

Chapter Four describes procedures for collecting indices related to the prior week's mean sleep duration using wrist actigraphy. As described in that chapter, there is concern that actigraphy-derived estimates of sleep time are sensitive to artifactual signals of wakefulness associated with sleep-disturbed breathing. The main bout length of relative inactivity is less sensitive to such artifacts compared to the cumulative duration of inactivity in the main sleep bout because it depends only on identifying the time of sleep onset and the time of final sleep termination. However, the main bout length of relative inactivity may overestimate duration of sleep because it fails to exclude true wakefulness after sleep onset. In contrast, the cumulative duration of inactivity in the main sleep bout may underestimate the duration of sleep in subjects with apnea if movements during sleep associated with respiratory events are defined as wakefulness by actigraphic scoring. The homeostatic drive for sleep is increased by cumulative partial sleep deprivation [Dinges et al, 1997], a value thought to be approximated by the sum of daily differences between sleep needed and sleep obtained. This latter value was approximated using the two indices described above. Subjects with relatively diminished amounts of sleep during the week prior to neurobehavioral testing are likely to have increased sleep pressure, which is hypothesized to produce both a subjective perception of sleepiness and reduced functional capacity. To investigate this, while accounting for the sampling by risk group, two-way analyses of variance were performed.

Table 6.3 provides descriptive statistics for the Epworth sleepiness scale for different durations of the mean main bout length of relative inactivity for the higher and lower risk groups (all subjects). In both groups, there appeared to be increased mean Epworth values for subjects with a mean of <6 hours of duration of the main sleep bout per day. Table 6.4 summarizes the two-way analysis of variance for the Epworth Sleepiness Scale. To assess these differences more completely, we performed a two-way analysis of variance for the Epworth Sleepiness Scale (see Table 6.4).

Table 6.3. Epworth Sleepiness Scale values by mean main bout length of relative inactivity category.

	Higher Risk					Lower Risk				
Hours	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6	20	11.2	5.6	3	23	13	9.8	4.6	3	19
6 - <7	47	10.3	4.1	2	20	23	8.8	3.5	2	14
7 - 8	81	8.1	4.5	0	22	63	8.4	4.0	0	20
>8	45	8.3	4.8	1	23	32	7.8	4.6	0	18
Total N	193					131				

Table 6.4. Two-way ANOVA (mean main bout length of relative inactivity category by risk group) for Epworth Sleepiness Scale. There were significant differences among mean values by mean main bout length of relative inactivity.

Test	F value		df	P value	Pair-wise t-test p<0.05	P value
Duration (Overall)	4.48	3	319	0.004	<6 vs. >8	0.005
Risk Group	0.76	1	319	0.382	<6 vs. 7-8	0.006
Duration Linear Trend	10.75	1	319	0.001	6-<7 vs. >8	0.018
Duration Quadratic Trend	0.33	1	319	0.567	6-<7 vs. 7-8	0.020
Risk Group * Duration Interaction	0.70	3	316	0.552		

Controlling for risk group, there were significant differences in mean Epworth Sleepiness Scale among the mean main bout length of relative inactivity categories (p=0.004). The linear trend in mean Epworth scores was significant (p=0.001) and there was no evidence of a quadratic trend (p=0.567). Specific pair-wise contrasts with the reference duration of 7 to 8 hours were highly significant, <6 vs. >8 (p=0.005) and <6 vs. 7-8 (p=0.006) and also significant between 6-<7 vs. >8 (p=0.018) and between 6-<7 vs. 7-8 (p=0.020). Least squares adjusted mean values of ESS controlling for risk group were 10.6 for a mean duration of <6 hours, 9.8 for a mean duration of 6-7 hours, 8.2 for a mean duration of 7-8 hours and 8.0 for a mean duration of >8 hours. The pooled estimate of the residual SD was 4.43.

Further analysis was performed examining the association of Epworth Sleepiness Scale values with mean main bout length of relative inactivity using the categories defined in Stutts et al [1999] (Figure 6.2). As is evident, greater proportions of subjects with heavy and extreme sleepiness were observed for subjects with shorter mean main bout lengths of relative inactivity. Among subjects with a mean duration of less than 6 hours, the percentage of subjects with extreme sleepiness (ESS=16+) was 21.4%. This compares to only 8.2%, 7.2%, and 7.5% for subjects with mean durations of 6-<7 hours, 7-8 hours, and >8 hours, respectively.
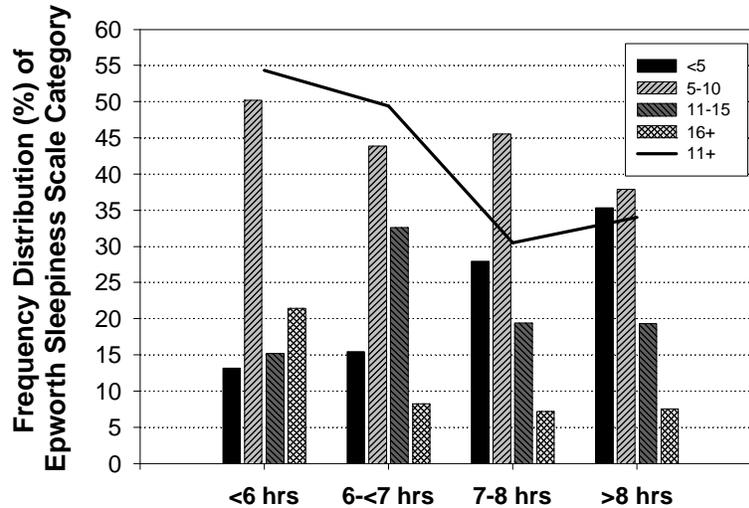
Figure 6.2 Percentages of subjects with Epworth Sleepiness Scale according to criteria used in Stutts et al [1999] (<5 none or minimal sleepiness), 5-10 (moderate sleepiness) 11-15 (heavy sleepiness), and 16+ (extreme sleepiness) by mean main bout length of relative inactivity during the prior week as determined by wrist actigraphy. The solid line shows the percentages of subjects with values equal to 11+. Percentages were determined based on weighted data.

These analyses were repeated replacing main bout length of relative inactivity category with cumulative duration of inactivity in main sleep bout category. Table 6.5 provides descriptive statistics for the Epworth sleepiness scale by mean duration of sleep category for the higher and lower risk groups. In both groups, there appeared to be increased mean Epworth values for subjects with a mean of <5 and 5-<6 hours of cumulative duration of inactivity during the main sleep bout per day. Interestingly, the mean values for subjects with >8 hours mean duration were slightly larger, rather than smaller, relative to those with 7-8 hours, potentially suggesting that the extended sleep durations were in response to perceived need for greater sleep.

Table 6.5. Epworth Sleepiness Scale values by mean cumulative duration of inactivity in main sleep bout category during the prior week in the two different risk groups.

Hours	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	41	10.3	4.9	3	23	10	9.4	4.4	5	19
5 - <6	44	10.0	5.3	2	22	27	9.5	3.8	3	17
6 - <7	61	8.4	4.2	0	19	37	8.5	3.7	0	20
7 - 8	42	7.8	4.4	1	17	45	7.6	4.6	0	18
>8	5	8.2	4.3	4	14	12	8.6	4.4	2	18
Total N	193					131				

As above, we assessed these differences using a two-way ANOVA (see Table 6.6).

Table 6.6. Two-way ANOVA (mean cumulative duration of inactivity in main sleep bout category and risk group) for Epworth Sleepiness Scale. There were significant differences among mean values by category of cumulative duration of inactivity during main sleep bout but not by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05	P value
Duration (Overall)	3.33	4	318	0.011	<5 vs. 7-8	0.004
Risk Group	0.18	1	318	0.668	5-<6 vs. 7-8	0.004
Duration Linear Trend	3.73	1	318	0.054	<5 vs. 6-<7	0.035
Duration Quadratic Trend	1.05	1	318	0.304	5-<6 vs. 6-<7	0.044
Risk Group * Duration Interaction	0.12	4	314	0.981		

Table 6.6 summarizes the two-way analysis of variance for the Epworth Sleepiness Scale. Controlling for risk group, there were significant differences in mean Epworth Sleepiness Scale among the mean cumulative duration of inactivity in main sleep bout categories (p=0.011). Although the linear trend just missed statistical significance at the α=0.05 level, specific pair-wise contrasts with the reference duration of 7 to 8 hours were highly significant, <5 vs. 7-8 (p=0.004) and 5-<6 vs. 7-8 (p=0.004) and also significant between <5 vs. 6-<7 (p=0.035) and 5-<6 vs. 6-<7 (p=0.044). Least squares adjusted mean values of ESS controlling for risk group were 10.0 for a mean duration of <5 hours, 9.8 for a mean duration of 5-<6 hours, 8.4 for a mean duration of 6-7 hours, 7.7 for a mean duration of 7-8 hours and 8.5 for a mean duration of >8 hours. The pooled estimate of the residual SD was 4.43.

Among subjects with a mean cumulative duration of inactivity in main sleep bout category of less than 5 hours, the percentage of subjects with extreme sleepiness (ESS=16+), was 16.2%. This compares to only 5.1%, 7.7%, and 7.0% for subjects with mean durations of 6-<7 hours, 7-8 hours, and >8 hours, respectively (see Figure 6.3).
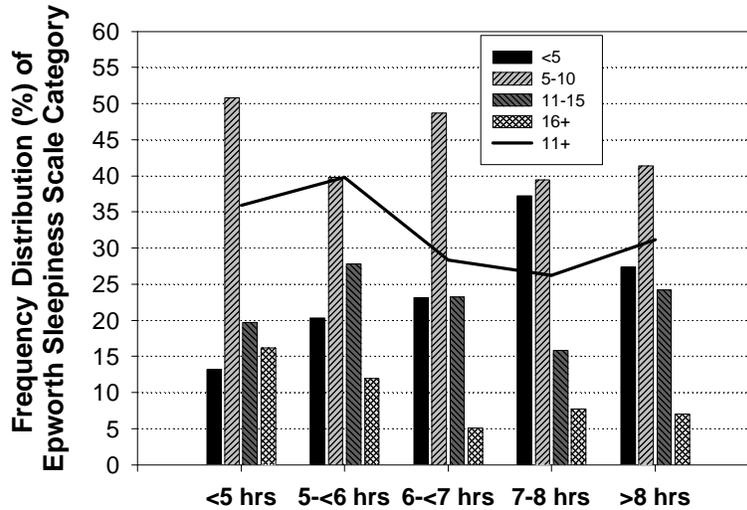
Figure 6.3. Percentages of subjects with Epworth Sleepiness Scale according to criteria used in Stutts et al [1999] (<5 none or minimal sleepiness), 5-10 (moderate sleepiness) 11-15 (heavy sleepiness), and 16+ (extreme sleepiness) by mean cumulative duration of inactivity in main sleep bout category during the prior week as determined by wrist actigraphy. The solid line shows the percentages of subjects with values equal to 11+. Percentages were determined based on weighted data.

Finally, these analyses were repeated restricting attention to subjects with no significant sleep apnea (AHI<5 events/hr). Tables 6.7 and 6.8 provide descriptive statistics and results from the two-way ANOVA, respectively. Comparison between Tables 6.5 and 6.7 reveal similar patterns of mean Epworth scores across the cumulative duration of inactivity categories. Thus, shorter sleep durations appear to be associated with perceived sleepiness even among subjects with no significant sleep apnea. The differences among mean values and the linear trend were not conclusively statistically significant in this smaller subsample. However, the significance levels of pairwise contrasts were comparable to those found in the larger sample (see Table 6.8).

Table 6.7. Epworth Sleepiness Scale values by mean cumulative duration of inactivity in main sleep bout category during the prior week in the two risk groups among subjects with AHI<5/hr.

Hours	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	14	10.5	4.6	3	20	9	9.9	4.3	5	19
5 - <6	15	9.8	6.4	2	20	21	9.4	4.0	3	17
6 - <7	35	7.8	4.2	0	18	31	8.7	3.8	0	20
7 - 8	29	8.1	4.5	1	17	41	7.6	4.5	0	18
>8	2	9.0	7.1	4	14	11	8.5	4.6	2	18
Total N	95					113				

Table 6.8. Two-way ANOVA (mean cumulative duration of inactivity in main sleep bout category and risk group) for Epworth Sleepiness Scale among subjects with AHI<5/hr. There were no significant differences among mean values by apnea severity or by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	1.91	4	202	0.111	<5 vs. 7-8 (p=0.024)
Risk Group	0.01	1	202	0.924	5-<6 vs. 7-8 (0.061)
Duration Linear Trend	2.49	1	202	0.116	<5 vs. 6-<7 (p=0.052)
Duration Quadratic Trend	1.24	1	202	0.267	
Risk Group * Duration Interaction	0.27	4	198	0.898	

In conclusion, these analyses provide evidence that subjective sleepiness is increased by reduced sleep duration.

E. Determinants of Epworth Sleepiness

As described above, although mean ESS did not significantly vary by sleep apnea severity, subjects with reduced mean main bout length of relative inactivity did have a significantly elevated mean ESS. Thus, we wished to assess the impact of sleep apnea severity and mean main bout length of relative inactivity simultaneously while also controlling for other variables, in particular for the potential effect of age, gender, obesity, and general perception of health-related quality-of-life. To do this, we performed a multiple linear regression analysis using the Epworth Sleepiness Scale value as the outcome variable. Results are detailed in Table 6.9.

Age, BMI, and female gender were not significant predictors of Epworth Sleepiness Scale. Consistent with the two-way ANOVA results, apnea severity was not related to Epworth Sleepiness Scale even after controlling for these other variables (p=0.667). The association

between mean main bout length of relative inactivity and ESS retained a trend toward significance controlling for the other variables in the model (p=0.061). A mean main bout length of relative inactivity <6 hours was associated with a significant increase in the expected Epworth score of 2.1 points relative to those with >8 hours (p=0.019). Similarly, a mean sleep bout duration between 6 and 7 hours compared to >8 hours was associated with an increase in the expected Epworth score of 1.5 points relative to those with >8 hours (p=0.039). Perceptions of increased sleepiness were significantly associated with perceptions of generally lower health-related QoL (p<0.0001). Each 1-point increase (worsening) in the 6-point QoL scale was associated with an expected increase in ESS of 1.2 points (SE=0.3). When alcohol use and current smoking were added to the model, they were not significant. However, when frequent snoring (>=3 times per week) was added to the model its incremental explanatory power was statistically significant (controlling for sleep apnea, mean sleep duration, etc.) at p=0.003. On average, controlling for the other variables in the model, frequent snoring increased the expected Epworth total score by almost 2 points (SE=0.6). This result suggests that the presence of frequent snoring itself may be an independent risk factor for excessive daytime sleepiness as measured by self-report. Frequent snoring will occur not only in individuals with obvious sleep apnea, but also in individuals who have the upper airway resistance syndrome [Guilleminault et al, 1993]. Thus, this result supports the concept that upper airway resistance syndrome can manifest itself in self-reported excessive sleepiness.

Table 6.9. Assessment of variables related to Epworth Sleepiness Scale including duration of main bout of relative inactivity

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	7.9386	0.5353	0.0000	---
Age - 45.5	-0.0154	0.0217	0.4786	---
BMI - 29.9	0.0733	0.0523	0.1618	---
Female gender	-0.1787	0.9771	0.8550	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.6674
Linear trend	---	---	---	0.3489
Quadratic trend	---	---	---	0.7822
AHI Model parameters	---	---	---	---
>30 vs. <5	0.9126	1.3208	0.4901	---
15-30 vs. <5	0.8062	1.0811	0.4564	---
5-<15 vs. <5	-0.3613	0.6539	0.5810	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.0612
Linear trend	---	---	---	0.0103
Quadratic trend	---	---	---	0.9529
Bout Duration model parameters				
<6 vs. >8 hrs.	2.1413	0.9092	0.0191	---
6-<7 vs. >8 hrs.	1.5376	0.7405	0.0386	---
7-8 vs. >8 hrs.	0.6689	0.6093	0.2731	---
Health related QoL Score - 2.44	1.1859	0.2911	0.0001	---
Model Summary	---	---	---	---
R-square	0.103	---	---	---
Root MSE	4.2642	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	1.9416	0.6366	0.0025	0.7999
Current Smoking	-0.0187	0.2930	0.9491	0.7009
Alcohol >2 drinks / day	-0.1331	0.3508	0.7046	0.4967

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m^2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

Multiple linear regression model for Epworth Sleepiness Scale. Explanatory variables included sleep apnea severity, mean main bout length of relative inactivity during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator. The top row indicates that the expected ESS is 7.94 for a 45.5 year-old male subject who has a BMI of 29.9 kg/m^2 and a QoL index value equal to 2.44 (scale range 1 to 6). The parameter estimates indicate the change in Epworth Sleepiness Score (either positive or negative) for the variable shown.

Estimation of the regression model was repeated replacing main bout length of relative inactivity with cumulative duration of inactivity during main sleep bout. Results are summarized in Table 6.10. In general the results were similar.

Table 6.10. Assessment of variables related to Epworth Sleepiness Scale including mean duration of cumulative inactivity

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	7.9901	0.4683	0.0000	---
Age - 45.5	-0.0185	0.0219	0.3993	---
BMI - 29.9	0.0559	0.0524	0.2868	---
Female gender	-0.2546	0.9756	0.7943	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.6537
Linear trend	---	---	---	0.5038
Quadratic trend	---	---	---	0.8519
AHI Model parameters	---	---	---	---
>30 vs. <5	0.5321	1.3508	0.6939	---
15-30 vs. <5	0.7579	1.0817	0.4840	---
5-<15 vs. <5	-0.5201	0.6591	0.4306	---
Cum. Duration of Inactivity	---	---	---	---
Overall difference in means	---	---	---	0.0893
Linear trend	---	---	---	0.1827
Quadratic trend	---	---	---	0.3962
Model parameters	---	---	---	---
<5 vs. 7-8 hrs.	1.7294	0.8454	0.0416	---
5-<6 vs. 7-8 hrs.	1.6755	0.6995	0.0172	---
6-<7 vs. 7-8 hrs.	0.6662	0.6233	0.2860	---
>8 vs. 7-8 hrs.	0.7890	1.0621	0.4581	---
Health related QoL Score - 2.44	1.1456	0.2896	0.0001	---
Model Summary	---	---	---	---
R-square	0.0998	---	---	---
Root MSE	4.2727	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	1.9201	0.6244	0.0023	0.6805
Current Smoking	0.0235	0.2949	0.9366	0.6846
Alcohol >2 drinks / day	-0.1367	0.3512	0.6974	0.6686

Multiple linear regression model for Epworth Sleepiness Scale. Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in main sleep bout during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator. The top row indicates that the expected ESS is 7.94 for a 45.5 year-old male subject who has a BMI of 29.9 kg/m^2 and a QoL index value equal to 2.44 (scale range 1 to 6). The parameter estimates indicate the change in Epworth Sleepiness Score (either positive or negative) for the variable shown.

6.3. Karolinska Sleepiness Scale

A. Description of Scale

The Karolinska Sleepiness Scale (KSS) is a highly sensitive subjective measurement scale for sleepiness [Akerstedt and Gillberg, 1990] (see Appendix J). Values of seven and above occur when the subject is so sleepy that there is intrusion of sleep into wakefulness as measured by the electroencephalogram (micro-sleeps) while a value of 5-6 represents an intermediate level of sleepiness seen when sleep duration is shortened [Akerstedt, 2000]. The KSS was obtained as part of our Neurobehavioral Assessment Battery (NAB). The NAB was administered four times the day after each subject underwent overnight polysomnography. For purposes of data analysis, the mean of the four test-day KSS scores were computed. These summary measures were subjected to further analysis.

B. Distribution of KSS Values in the Higher and Lower Risk Groups

Figure 6.4 displays the mean and standard deviation of (the mean) KSS values for the higher and lower risk groups. The difference in mean values between risk groups was not statistically significant (t=-1.35, df=401, p=0.177). The population mean value was estimated as the weighted average of within group means and was found to be equal to 4.11. The pooled estimate of the SD (weighted according to degrees-of-freedom) was 1.28. As described in the previous section, KSS values greater than or equal to 7 are considered indicative of excess sleepiness while values of 5-6 are considerate intermediate values. In the higher risk group, 7 of 247 (2.8%) had values greater than or equal to 7 and 50 of 247 (20.2%) had values less than 7 but greater than or equal to 5. The remaining 190 of 247 (76.2%) had values less than 5. The same numbers (%) in the lower risk group were 4 of 156 (2.6%), 35 of 156 (22.4%), and 117 of 156 (75.0%). These percentages were not statistically significantly different (Kruskal-Wallis χ^2=0.170, df=1, p=0.690) from those in the higher risk group. The weighted estimate for the percentage of individuals in our population of CDL holders who have a Karolinska Sleepiness value greater than or equal to 7 or greater than or equal to 5 are 2.7% and 24.2%, respectively.
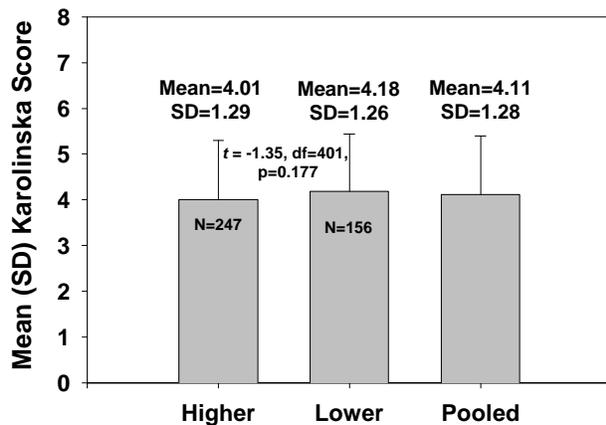
Figure 6.4. Karolinska Sleepiness Scale mean and standard deviation for the higher and lower risk groups. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal true within risk

C. Association with Sleep Apnea Severity

Table 6.11 summarizes the mean (SD) Karolinska Sleepiness Scale values by sleep apnea severity for subjects in the higher and lower risk groups. The mean values were similar across sleep apnea severity levels in both risk groups.

Table 6.11. Karolinska Sleepiness Scale values by severity of sleep apnea for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	4.04	1.24	1.40	7.60	137	4.18	1.31	0.80	8.00
5 - <15	69	3.98	1.38	1.80	8.00	16	4.15	0.91	2.80	5.80
15 - <30	30	4.13	1.42	1.40	7.00	2	4.20	0.00	4.20	4.20
>= 30	28	3.87	1.18	2.20	6.80					
Total N	246					155				

Two-way analysis of variance (ANOVA) was used to assess if there were significant differences among the sleep apnea severity levels for mean Karolinska Sleepiness Scale controlling for risk group. There were no significant differences (F=0.23, df=3,396, p=0.877) between risk groups. Nor was there evidence of a linear trend in mean values as a function of sleep apnea severity (F=0.16, df=1,396, p=0.686). Results from the ANOVA are summarized in Table 6.12.

Table 6.12. Two-way ANOVA (apnea severity and risk group) for Karolinska Sleepiness Scale. There were no significant differences among mean values by apnea severity or by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
AHI (Overall)	0.23	3	396	0.877	
Risk Group	1.01	1	396	0.317	
AHI Linear Trend	0.16	1	396	0.686	
AHI Quadratic Trend	0.30	1	396	0.584	
Risk Group * AHI Category Interaction	0.01	2	394	0.993	

D. Association with Prior Week's Mean Sleep Duration

Tables 6.13 and 6.14 summarize the analysis of the relationship between the Karolinska Sleepiness Scale and mean main bout length of relative inactivity during the prior week as determined by wrist actigraphy.

Table 6.13. Karolinska Sleepiness Scale values by mean main bout length of relative inactivity during the prior week and risk group.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	4.36	1.34	2.50	7.20	13	4.69	1.69	2.60	8.00
6<=hr< 7	53	4.24	1.25	2.00	8.00	23	3.77	1.29	0.80	5.40
7<=hr<= 8	88	3.86	1.26	1.40	7.80	61	3.95	1.19	1.20	7.20
hr>8	47	4.01	1.32	2.00	6.80	32	4.44	1.00	1.60	6.60
Total N	208					129				

Table 6.14. Two-way ANOVA (mean main bout length of relative inactivity and risk group) for Karolinska Sleepiness Scale. There was a statistically significant quadratic trend among mean KSS values. There were no significant differences among mean values by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05	P value
Duration (Overall)	2.42	3	332	0.066	< 6 vs. 7-8	0.015
Risk Group	0.33	1	332	0.564		
Duration Linear Trend	1.92	1	332	0.167		
Duration Quadratic Trend	4.57	1	332	0.033		
Risk Group * Duration Category Interaction	1.64	3	329	0.180		

Least squares adjusted mean values controlling for risk group were 4.5 for a mean duration of <6 hours, 4.1 for a mean duration of 6-<7 hours, 3.9 for a mean duration of 7-8 hours and 4.2 for a mean duration of >8 hours. The mean KSS value was significantly larger for subjects with less than a 6-hour main bout length of relative inactivity compared to these with between 7 and 8 hours (p=0.015). The non-linear trend emerged because subjects with duration greater than 8 hours appeared to have elevated values relative to those with values between 7 and 8 hours. The pooled estimate of the residual SD was 1.26. Since it is difficult to interpret the magnitude of numeric differences in terms of clinical significance we computed the percentage difference and the standardized effect size [Cohen, 1988] for this difference. The percentage difference in adjusted means between subjects with less than a 6 hour long main bout length compared to those with between 7 and 8 hours was 15.4%. Relative to the between subject residual standard deviation, this represents a standardize effect size (4.5-3.9)/1.26, or 0.48. Thus, this difference is at most, only moderately large.

These analyses were repeated for mean cumulative duration of inactivity during the main sleep bout. Results are summarized in Tables 6.15 and 6.16. Results were comparable to those for the duration of the major bout of relative inactivity.

Table 6.15. Karolinska Sleepiness Scale values by mean cumulative duration of inactivity during the main sleep bout during the prior week in the two risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	45	4.24	1.41	2.00	8.00	10	4.48	1.51	2.40	7.20
5<=hr< 6	47	4.36	1.27	2.40	7.60	28	4.10	1.23	1.00	8.00
6<=hr< 7	67	3.91	1.21	1.80	7.80	35	3.94	1.20	0.80	6.60
7-8 hrs	44	3.63	1.22	1.40	6.80	44	4.06	1.34	1.20	7.20
hr> 8	5	4.52	0.87	3.80	6.00	12	4.52	0.80	3.00	5.80
Total N	208					129				

Table 6.16. Two-way ANOVA (mean cumulative duration of inactivity during the main sleep bout and risk group) for Karolinska Sleepiness Scale. Statistically significant differences among mean values were observed among the sleep duration categories. There were no significant differences among mean values by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	2.49	4	331	0.043	<5 vs. 7-8
Risk Group	0.51	1	331	0.477	5-<6 vs. 7-8
Duration Linear Trend	0.01	1	331	0.922	
Duration Quadratic Trend	4.51	1	331	0.034	
Risk Group * Duration Category Interaction	0.78	4	327	0.539	

E. Determinants of Karolinska Sleepiness

The same model as described previously in Section 6.2.E (Determinants of Epworth Sleepiness) was used to predict Karolinska Sleepiness (KSS). Results are summarized in Table 6.17 using main bout length of relative inactivity. Results using cumulative duration of inactivity during the main sleep bout were similar. Only the health-related QoL measure was independently associated with mean KSS although the quadratic trend for sleep duration was significant reflecting the relatively larger value for subjects with >8 hours mean durations of sleep.

Table 6.17. Multiple linear regression model for Karolinska Sleepiness Scale.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	4.2793	0.1509	0.0000	---
Age - 45.5	-0.0093	0.0061	0.1256	---
BMI - 29.9	0.0117	0.0146	0.4231	---
Female gender	0.0073	0.2778	0.9790	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.8116
Linear trend	---	---	---	0.9899
Quadratic trend	---	---	---	0.7777
AHI Model parameters	---	---	---	---
>30 vs. <5	-0.0991	0.3520	0.7785	---
15-30 vs. <5	0.1573	0.3075	0.6093	---
5-<15 vs. <5	-0.1260	0.1830	0.4917	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.1254
Linear trend	---	---	---	0.4392
Quadratic trend	---	---	---	0.0192
Bout Duration model parameters				
<6 vs. >8 hrs.	0.2082	0.2580	0.4204	---
6-<7 vs. >8 hrs.	-0.2649	0.2077	0.2030	---
7-8 vs. >8 hrs.	-0.2605	0.1721	0.1311	---
Health related QoL Score - 2.44	0.3860	0.0814	0.0000	---
Model Summary	---	---	---	---
R-square	0.105	---	---	---
Root MSE	1.2138	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	0.2124	0.1758	0.2279	0.8104
Current Smoking	-0.0867	0.0825	0.2944	0.7690
Alcohol >2 drinks / day	0.1460	0.1074	0.1748	0.7269

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m^2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

6.4. Stanford Sleepiness Scale

A. Description of Scale

The Stanford Sleepiness Scale (SSS) is a single seven-item ordinal subjective measure scale for sleepiness [Hoddes et al, 1973]. The SSS asks subjects to circle the statement that best describes how sleepy they feel at the moment they are answering the question. The categories are: (1) Feeling active and vital; alert; wide awake; (2) Functioning at a high level, but not at peak; able to concentrate; (3) Relaxed; awake not at full alertness; responsive; (4) A little foggy; not at peak; let down; (5) Fogginess; beginning to lose interest in remaining awake; slowed

down; (6) Sleepiness; prefer to be lying down; fighting sleep; woozy; and (7) Almost in reverie; sleep onset soon; lost struggle to remain awake. The SSS was also obtained as part of the NAB. It was obtained at the beginning and at the end of each NAB session. For purposes of data analysis, the mean of the four SSS scores were computed separately for the assessment obtained at the beginning and at the end of the NAB session. These summary measures were subjected to further analysis. The results for the values obtained at the end of the NAB are described below.

B. <u>Distribution of SSS Values in the Higher and Lower Risk Groups</u>

Figure 6.5 displays the mean and standard deviation of (the mean) SSS values for the higher and lower risk groups. The difference in mean values between risk groups was not statistically significant (t=-1.32, df=401, p=0.187). The population mean value was estimated as the weighted average of within group means and was found to be equal to 2.76. The pooled estimate of the SD (weighted according to degrees-of-freedom) was 0.81. As the scale above indicates, SSS values greater than or equal to 4 reflect a self-perception of being "a little foggy" or worse. In the higher risk group, 38 of 247 (15.4%) had values greater than or equal to 4 while 8 of 247 (3.2%) had values greater than or equal to 5. Only 1 subject had a value of 6 (0.4%) and there were no values equal to 7. In the lower risk group, 23 of 156 (14.8%) had values greater than or equal to 4 and 5 of 156 (3.2%) had values equal to 5. There were no subjects with mean SSS values equal to 6 or 7. The distributions of mean SSS values did not significantly differ between the risk groups (Kruskal-Wallis χ^2=1.46, df=1, p=0.227).
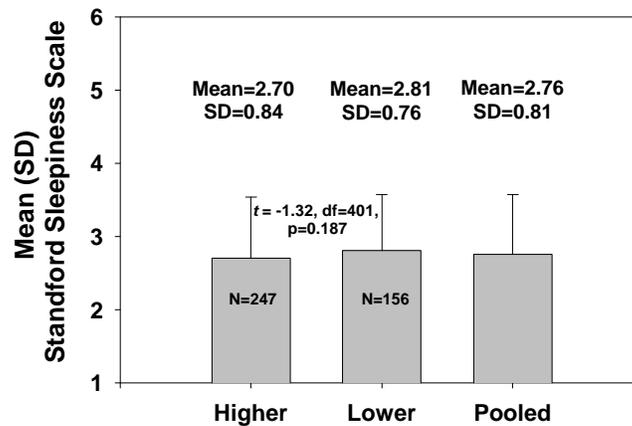
Figure 6.5. Stanford Sleepiness Scale mean and standard deviation for the higher and lower risk groups. The values summarized are mean values over four test sessions on the day after the overnight polysomnography. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal true within risk group SD's and is computed using the usual formula for computing a pooled SD estimate.

C. Association with Sleep Apnea Severity

Table 6.18 summarizes the mean (SD) Stanford Sleepiness Scale values by sleep apnea severity for subjects in the higher and lower risk groups. The mean values were similar across sleep apnea severity levels in both risk groups.

Table 6.18. Stanford Sleepiness Scale values by severity of sleep apnea for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
AHI	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	2.76	0.86	1.00	5.80	137	2.82	0.79	1.00	4.80
5 - <15	69	2.70	0.89	1.20	5.40	16	2.68	0.57	1.80	3.60
15 - <30	30	2.57	0.85	1.20	4.00	2	2.80	0.85	2.20	3.40
>= 30	28	2.61	0.63	1.80	4.60					
Total N	246					155				

Two-way analysis of variance (ANOVA) was used to assess if there were significant differences among the sleep apnea severity levels for mean Stanford Sleepiness Scale controlling for risk group. There were no significant differences (F=0.64, df=3,396, p=0.587). There was no evidence of a linear trend in mean SSS values as a function of sleep apnea severity (F=1.07, df=1,396, p=0.301). Results from the ANOVA are summarized in Table 6.19.

Table 6.19. Two-way ANOVA (apnea severity and risk group) for Stanford Sleepiness Scale. There were no significant differences among mean values by apnea severity or by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
AHI (Overall)	0.64	3	396	0.587	
Risk Group	0.30	1	396	0.585	
AHI Linear Trend	1.07	1	396	0.301	
AHI Quadratic Trend	0.22	1	396	0.639	
Risk Group * AHI Category Interaction	0.11	2	394	0.895	

D. Association with Prior Week's Mean Sleep Duration

Tables 6.20 and 6.21 summarize the analysis of the relationship between the Stanford Sleepiness Scale and mean main bout length of relative inactivity. The quadratic trend resulting from the relatively larger mean value for the hour >8 was also present (p=0.053). The adjusted mean SSS for those with <6 hours versus 7 to 8 hours appeared to differ (2.9 vs. 2.6, p=0.056). However, when the analyses were repeated for mean cumulative duration of inactivity in the main sleep bout, the quadratic trend lost its statistical significance (results not shown, p=0.123).

Table 6.20. Stanford Sleepiness Scale values by mean main bout length of relative inactivity during the prior week and risk group.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	2.84	0.99	1.40	4.80	13	3.02	0.92	1.80	4.60
6<=hr< 7	53	2.75	0.80	1.40	5.40	23	2.60	0.68	1.00	3.80
7<=hr<= 8	88	2.58	0.80	1.00	5.20	61	2.65	0.74	1.00	4.50
hr>8	47	2.77	0.96	1.20	5.80	32	2.85	0.59	1.20	4.00
Total N	208					129				

Table 6.21. Two-way ANOVA (mean main bout length of relative inactivity and risk group) for Stanford Sleepiness Scale. There were no significant differences among mean values by duration of main bout of relative inactivity or by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	1.78	3	332	0.151	
Risk Group	0.17	1	332	0.681	
Duration Linear Trend	0.64	1	332	0.426	
Duration Quadratic Trend	3.76	1	332	0.053	
Risk Group * Duration Category Interaction	0.42	3	329	0.740	

E. Determinants of Stanford Sleepiness Scale

We again used the multiple regression approach described previously to assess the role of different variables in determining the Stanford Sleepiness Scale. The results for the Stanford Sleepiness Scale are summarized in Table 6.22. These results were generally similar to those obtained for the Karolinska Sleepiness Scale.

Table 6.22. Multiple linear regression model for Stanford Sleepiness Scale.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	2.8492	0.0939	0.0000	---
Age - 45.5	-0.0044	0.0038	0.2418	---
BMI - 29.9	0.0043	0.0091	0.6379	---
Female gender	-0.0172	0.1728	0.9210	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.6735
Linear trend	---	---	---	0.4847
Quadratic trend	---	---	---	0.6478
AHI Model parameters	---	---	---	---
>30 vs. <5	-0.1429	0.2190	0.5145	---
15-30 vs. <5	-0.1653	0.1913	0.3881	---
5-<15 vs. <5	-0.1090	0.1138	0.3391	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.2904
Linear trend	---	---	---	0.5422
Quadratic trend	---	---	---	0.0633
Bout Duration model parameters				
<6 vs. >8 hrs.	0.0966	0.1605	0.5478	---
6-<7 vs. >8 hrs.	-0.1254	0.1292	0.3327	---
7-8 vs. >8 hrs.	-0.1395	0.1071	0.1935	---
Health related QoL Score - 2.44	0.2284	0.0507	0.0000	---
Model Summary	---	---	---	---
R-square	0.088	---	---	---
Root MSE	0.7552	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	0.2036	0.1099	0.0649	0.5352
Current Smoking	-0.0360	0.0515	0.4855	0.6713
Alcohol >2 drinks / day	0.0223	0.0674	0.7409	0.9766

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m^2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

6.5. Functional Outcomes of Sleep Questionnaire

A. Description of Scale

The Functional Outcomes of Sleep Questionnaire (FOSQ) was the first self-report measure designed to assess the impact of disorders of excessive sleepiness on multiple activities of everyday living [Weaver et al, 1997] (see questionnaire in Appendix J). Weaver et al [1997] described its development and performed psychometric analyses in three samples: Sample 1 (n=153) consisted of individuals seeking medical attention for a sleep problem and persons of similar age and gender having no sleep disorder; Samples 2 (n=24) and 3 (n=51) were composed

of patients from two medical centers diagnosed with obstructive sleep apnea. Factor analysis of the FOSQ yielded five factors: activity level, vigilance, intimacy and sexual relationships, general productivity, and social outcome. Internal reliability was excellent for both the subscales (alpha=0.86 to alpha=0.91) and the total scale (alpha=0.95). Test-retest reliability of the FOSQ yielded coefficients ranging from r=0.81 to r=0.90 for the five subscales and r=0.90 for the total measure. The FOSQ successfully discriminated between normal subjects and those seeking medical attention for a sleep problem (t=-5.88, df=157, p=0.0001). This psychometric evaluation of the FOSQ demonstrated parameters acceptable for its application in research and in clinical practice to measure functional status outcomes for persons with excessive daytime sleepiness. The conclusions of this study were that the FOSQ could be used to determine how disorders of excessive sleepiness affect patients' abilities to conduct normal activities and the extent to which these abilities are improved by effective treatment of excessive daytime sleepiness. In the analyses described below, results for the FOSQ Total Score are summarized.

B. <u>Distribution of FOSQ Total Scores in the Higher and Lower Risk Groups</u>

Figure 6.6 displays the mean and standard deviation of the FOSQ Total Score values for the higher and lower risk groups. The difference in mean values between risk groups was not statistically significant (t=-0.52, df=401, p=0.605). The population mean value was estimated as the weighted average of within group means and was found to be equal to 17.6. The pooled estimate of the SD (weighted according to degrees-of-freedom) was 2.22. There is currently no widely accepted criterion for defining pathological sleepiness based on FOSQ total score or subscale values.
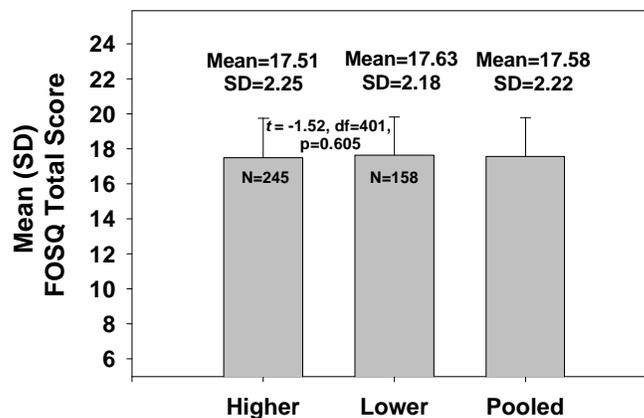
Figure 6.6. Functional Outcomes of Sleep Questionnaire (FOSQ) mean and standard deviation for the higher and lower risk groups. The values summarized are mean values over four test sessions on the day after the overnight polysomnography. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal true within risk group SD's and is computed using the usual formula for computing a pooled SD estimate.

C. <u>Association with Sleep Apnea Severity</u>

Table 6.23 summarizes the mean (SD) FOSQ total score values by sleep apnea severity for subjects in the higher and lower risk groups. The mean values were similar across sleep apnea severity levels in both risk groups.

Table 6.23. Functional Outcomes of Sleep Questionnaire (FOSQ) mean Total Scores for categories of severity of sleep apnea in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	117	17.66	2.37	8.88	20.00	139	17.69	2.14	7.86	22.44
5 - <15	69	17.50	2.10	11.15	20.00	16	16.96	2.61	11.29	19.43
15 - <30	30	17.07	2.08	12.91	20.00	2	18.81	0.48	18.47	19.15
>= 30	28	17.47	2.34	11.46	20.00					
Total N	244					157				

Two-way analysis of variance (ANOVA) was used to assess if there were significant differences among the sleep apnea severity levels for FOSQ total score controlling for risk group. There were no significant differences (F=0.62, df=3,396, p=0.602). There was no evidence of a linear trend in mean FOSQ total score as a function of sleep apnea severity (F=0.35, df=396, p=0.553). Results from the ANOVA are summarized in Table 6.24.

Table 6.24. Two-way ANOVA (apnea severity and risk group) for FOSQ total score. There were no significant differences among mean values by apnea severity or by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
AHI (Overall)	0.62	3	396	0.602	
Risk Group	0.01	1	396	0.939	
AHI Linear Trend	0.35	1	396	0.553	
AHI Quadratic Trend	0.78	1	396	0.377	
Risk Group * AHI Category Interaction	0.96	2	394	0.385	

D. Association with Prior Week's Mean Sleep Duration

Tables 6.25 and 6.26 summarize the analysis of the relationship between the FOSQ Total Score and mean main bout length of relative inactivity.

Table 6.25. FOSQ Total Score values by mean main bout length of relative inactivity during the prior week and risk.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	19	17.57	2.17	11.63	19.65	13	17.04	2.27	11.78	19.67
6<=hr< 7	52	17.14	2.16	9.81	19.89	24	17.91	2.31	11.29	22.44
7<=hr<= 8	88	17.70	2.11	11.15	20.00	62	17.78	1.97	10.00	20.00
hr>8	47	17.49	2.70	8.88	20.00	32	17.39	2.64	7.86	20.00
Total N	206					131				

Table 6.26. Two-way ANOVA (mean main bout length of relative inactivity and risk group) for FOSQ Total Score. There were no significant differences among mean values by main bout length category or by risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	0.59	3	332	0.623	
Risk Group	0.22	1	332	0.639	
Duration Linear Trend	0.19	1	332	0.667	
Duration Quadratic Trend	0.32	1	332	0.572	
Risk Group * Duration Category Interaction	0.73	3	329	0.538	

There was no evidence of differences among the mean values for FOSQ Total Score among the various categories for mean sleep duration.

E. Determinants of Functional Outcomes of Sleep Questionnaire Total Score

Finally, we assessed the determinants of FOSQ using the multiple regression approach described in previous sections. Results for the Functional Outcomes of Sleep Questionnaire Total Score are given in Table 6.27. Several demographic factors were significantly associated with FOSQ Total Score. Increases in age were associated with larger, i.e., improved, scores (p=0.035). Increases in BMI were marginally associated with smaller scores (p=0.063). Female gender was significantly associated with a FOSQ Total Score value that was more than one point smaller relative to that expected in males (p=0.023). As with Epworth Total Score, there was a large and highly statistically significant (p=0.0002) reduction in function among subjects reporting frequent snoring. There was a reduction in the expected FOSQ Total Score of more than 1 point even after controlling for sleep apnea severity, mean sleep duration, age, BMI, gender, and health-relate QoL. Again, the health-related QoL measure was significantly associated with subjectively reported function. The results for the model replacing mean main bout length of relative inactivity with mean cumulative duration of inactivity during the main sleep bout were very similar (data not shown).

Table 6.27. Multiple linear regression model for Functional Outcomes of Sleep Questionnaire Total Score. Explanatory variables included sleep apnea severity, mean main bout length of relative inactivity during the prior week as determined by actigraphy, age, gender, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	17.6175	0.2610	0.0000	---
Age - 45.5	0.0223	0.0105	0.0351	---
BMI - 29.9	-0.0472	0.0253	0.0628	---
Female gender	-1.0956	0.4802	0.0232	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.8421
Linear trend	---	---	---	0.9920
Quadratic trend	---	---	---	0.8271
AHI Model parameters	---	---	---	---
>30 vs. <5	-0.0919	0.6086	0.8801	---
15-30 vs. <5	0.0143	0.5316	0.9785	---
5-<15 vs. <5	-0.2807	0.3160	0.3751	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.8891
Linear trend	---	---	---	0.6114
Quadratic trend	---	---	---	0.5789
Bout Duration model parameters				
<6 vs. >8 hrs.	-0.1948	0.4493	0.6650	---
6-<7 vs. >8 hrs.	-0.0095	0.3568	0.9789	---
7-8 vs. >8 hrs.	0.1146	0.2970	0.6999	---
Health related QoL Score - 2.44	-0.8299	0.1405	0.0000	---
Model Summary	---	---	---	---
R-square	0.160	---	---	---
Root MSE	2.0986	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	-1.1883	0.3110	0.0002	0.8143
Current Smoking	0.0187	0.1419	0.8952	0.8442
Alcohol >2 drinks / day	0.1515	0.1558	0.3315	0.2822

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m^2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

6.6. Relationships Among Subjective Measures

Since the results of our analysis of all of these subjective measures of sleepiness were somewhat similar, we next questioned whether the results for these different measurement tools were related. Specifically, an analysis was performed to determine the extent to which the Epworth Sleepiness Scale, the (mean) Karolinska Sleepiness Scale, the (mean) Stanford Sleepiness Scale, and the FOSQ Total Score were measuring similar perceptions. First, Pearson and Spearman pair-wise correlations were determined separately for the higher and lower risk groups. In both groups the Pearson and Spearman correlations were similar and overall, the correlations in the higher risk group were similar to those in the lower risk group. Therefore, a Pearson partial correlation matrix was computed controlling for risk group in order to simplify the presentations. Table 6.28 contains the pair-wise Pearson partial correlations among these four subjective sleepiness measures.

Table 6.28. Pearson partial correlations among four subjective assessments of sleepiness controlling for risk group. All p-values were <0.0001. For ESS, KSS, and SSS larger values reflect greater sleepiness. For FOSQ Total Score, larger values reflect better functioning, hence the negative partial correlation.

	ESS	KSS	SSS	FOSQ
Epworth Sleepiness Scale (ESS)	1	+0.27	+0.25	-0.40
Karolinska Sleepiness (KSS)		1	+0.89	-0.29
Stanford Sleepiness (SSS)			1	-0.28
Functional Outcomes of Sleep Total Score (FOSQ)				1

There was a very large correlation between the (mean) Karolinska Sleepiness Scale and (mean) Stanford Sleepiness Scale (partial r=0.89). These measures were both obtained four times during the neurobehavioral test battery. The correlations are between the mean values. The high correlation is not surprising since both of these tools measure how sleepy an individual is at a given moment in time. The large correlation indicates stability within day and provides evidence for (concurrent) validity. The FOSQ total score and the Epworth Sleepiness Scale were obtained during the day of testing in our laboratory. There was a moderately large correlation between the results for these two instruments, that both measure the effects of sleepiness on daytime activity over a period of time. Correlations between the self-report measures of sleepiness at a moment in time (Karolinska, Stanford) and those over a period of time (Epworth, FOSQ) are, not surprisingly, lower (of the order of 0.25 to 0.29) but still significant. This internal consistency among these four measures provides evidence that we did, in fact, obtain valid and reliable assessments of *perceptions of sleepiness* through our self-report instruments. Therefore, the lack of association with sleep apnea severity does not appear to be caused by a lack of validity or reliability in the instruments themselves. As we show in Chapter Seven, there were significant associations between sleep apnea severity and function when assessed using objective measures. The inconsistency between subjectively reported perceptions of sleepiness and objectively measured function in their associations with sleep apnea severity is vitally important. It suggests that in our populations, there may be drivers of commercial vehicles with functional decrements resulting from sleep apnea that are not self-perceived. This will be discussed more fully in Chapter Seven.

6.10 <u>Discussion</u>

This component of our study reveals that a substantial proportion of our sample have evidence of self-reported sleepiness. Using, for example, a well-validated self-report measure of excessive sleepiness—the Epworth Sleepiness Score—we find that 32.6% overall of our sample have a score that is above the accepted norm for self-reported sleepiness (i.e., Epworth Sleepiness Score greater than 10). Car drivers with a score above the norm have a 3-4 fold increased risk of a crash due to falling asleep at the wheel [Stutts et al, 1999]. The very high prevalence of these elevated scores for excessive daytime sleepiness is alarming.

While a substantial proportion of drivers have elevated scores on self-reported sleepiness measures, the values of these scores do not relate to severity of sleep apnea measured in our laboratory. This result is in contrast to that reported in the next chapter, where we demonstrate that objective measures of sleepiness do relate to severity of sleep apnea. This implies that we cannot rely on self-report measures of sleepiness in this population to identify those most likely to be impaired from the presence of severe sleep apnea. We do not believe that this is simply explained by this population denying self-reported sleepiness. Indeed, as pointed out earlier, there is a very large fraction of all commercial drivers studied who do report excessive daytime sleepiness.

In contrast to the results for sleep apnea, we did find that for one of our key measures— the Epworth Sleepiness Scale (but not the other measures)—there was an association with sleep duration as measured at home no matter which method we chose to define our assessment of sleep duration. That we found this association with both ways we defined sleep duration indicates that the association is not simply related to difficulties we discussed in Chapter 4 in adequately measuring sleep at home in a population with a large number of individuals with sleep apnea. Thus, obtaining an Epworth Sleepiness Score is likely to be of value in this population.

We did find in our study that there were correlations between the results obtained from one self-report sleepiness measure and the others employed. This adds weight to our overall conclusions and suggests that in assessing self-report sleepiness in commercial drivers, only one of these instruments may be required. Based on our results, the Epworth Sleepiness Scale will provide the most valuable information when seeking to determine who is excessively sleepy as judged by self-report and it is associated with short sleep durations at home no matter how sleep duration is defined.

CHAPTER SEVEN

Objective Measures of Sleepiness

As described previously, all in-laboratory subjects were assessed using multiple instruments with regard to their subjective perception of sleepiness and functional impairment as well as with objective measures of sleepiness, lack of attention, and other potential functional neurobehavioral consequences of sleepiness. Chapter Seven reports on results from objective assessments with the primary focus being on the Mean Sleep Latency Test (MSLT), the Psychomotor Vigilance Test (PVT), the Divided Attention Driving Task (DADT), and the Digit Symbol Substitution Test (DSST). The nature of these various measures is described in the various individual sections that follow. Again, we first compare across sleep apnea severity categories and prior week's mean actigraphy results utilizing two putative measures of sleep duration, the mean main bout length of relative inactivity and mean cumulative duration of inactivity in main sleep bout (for detailed description, see Chapter Four). Then, the joint effect of both of these factors controlling for age, gender, BMI and other variables is assessed. In contrast to the lack of significant association with subjective sleepiness reported in Chapter Six, these objective measures were simultaneously associated with both sleep apnea severity and with mean main bout length of relative inactivity. Therefore, additional multivariable models were developed to graphically illustrate how the likelihood of impairment changes as AHI increases over its full range as assessed on an interval scale controlling for mean main bout length of relative inactivity and other variables in the model. The implications of these various results are discussed at the end of the chapter.

7.1 Multiple Sleep Latency Test (MSLT)

A. Description of Scale

The Multiple Sleep Latency Test (MSLT) is considered the primary method for objective assessment of physiological sleepiness [Carskadon and Dement, 1982; Carskadon et al, 1986] . In a sense, it measures the "pressure for sleep". The concept is simple: individuals who are sleepy and have a high pressure for sleep will fall asleep quicker in a darkened room when asked to do so. Thus, the MSLT measures sleep tendency by the rate at which individuals are able to fall asleep voluntarily. The MSLT test provides a series of 20 minute (typically, and in our study, 4) opportunities separated by 2 hours of enforced wakefulness during which a standard polysomnography electrode configuration for EEG recording etc. is maintained. The latency to sleep is averaged across these nap opportunities to give a mean sleep latency (MSLT) score taken as a measure of physiological sleepiness. By convention, albeit not backed by large bodies of epidemiological data, MSLT values from 10 to 20 minutes are considered normal and latencies of less than 5 minutes are considered 'pathological' sleepiness as they have been found to be associated with impaired performance in sleep deprivation experiments using normal controls [Carskadon et al, 1986]. Scores ranging from 6 to 10 minutes are considered to be in a "gray area" [Carskadon et al, 1986; Van den Hoed et al, 1981]. Sleep latency has been shown to be greatly reduced in patients with obstructive sleep apnea/hypopnea syndrome compared to normals [Kribbs et al, 1993b]. MSLT improves with therapy but often does not return to completely normal levels

[Kribbs et al, 1993b]. All of our in-laboratory subjects were assessed in our Clinical Research Center using standardized MSLT methods on the day following their overnight polysomnography.

B. Distribution of MSLT Values in the Higher and Lower Risk Groups

Figure 7.1 displays the mean and standard deviation MSLT values for the higher and lower risk groups. The difference in mean values between risk groups was not statistically significant (t=-1.57, df=404, p=0.112). In our population of holders of commercial vehicle drivers' licenses, the mean value was estimated (as the weighted average of within group means) to be equal to 8.30 minutes. The population standard deviation was estimated to be 4.60 minutes[3]. As indicated above, MSLT values less than 5 minutes are considered indicative of excessive sleepiness, values at least equal to 5 but less than 10 minutes are considered in the 'gray area', while values greater than or equal to 10 minutes are considered by convention to be normal. In the higher risk group, 76 of 247 (30.8%) subjects had values less than 5 minutes, 99 of 247 (40.1%) had values at least equal to 5 minutes but less than 10 minutes while the remaining 72 subjects (29.2%) had values greater than or equal to 10 minutes. In the lower risk group (n=159), 35 (22.0%), 72 (45.3%), and 52 (32.7%) subjects have values <5 minutes, 5 to <10, and ≥10 minutes, respectively. These percentages were not statistically significantly different (Kruskal-Wallis χ^2=2.53, df=1, p=0.112). The weighted estimate for the percentage of individuals in our population of CDL holders having MSLT values less than 5 minutes was 25.6% while the weighted estimate for having an MSLT at least equal to 5 but less than 10 minutes was 43.1%. Thus, overall, a substantial minority of our subjects—one in four—had levels of excessive sleepiness that would be of clinical concern.
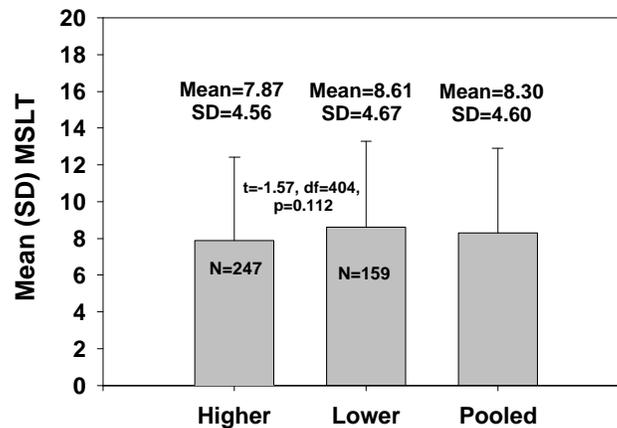
Figure 7.1.   Multiple Sleep Latency Test (MSLT) mean and standard deviation for the higher and lower risk groups. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal true within risk group SD's and is computed using the usual formula for computing a pooled SD estimate.

---

[3] The population SD was estimated as the square-root of a weighted average of sample variances with weights determined by degrees-of-freedom.

C. Association with Sleep Apnea Severity

Table 7.1 summarizes the mean (SD) MSLT values by sleep apnea severity for subjects in the higher and lower risk groups. Among higher risk subjects, there was a clear increase in physiological sleepiness as measured by the MSLT for subjects with severe sleep apnea (i.e., with AHI≥30 episodes per hour). In this group, the average value (4.8 minutes) was in the range usually considered indicative of pathological sleepiness (i.e., <5 minutes). Subjects with moderate sleep apnea (i.e., AHI at least equal to 15 but less than 30 episodes per hour) had a mean value between those with severe sleep apnea and those with mild sleep apnea (i.e., AHI at least equal to 5 but less than 15 episodes per hour) or with no sleep apnea (i.e., AHI less than 5 episodes per hour). In the lower risk group, the mean value for subjects with no sleep apnea and the mean value for subjects with mild sleep apnea were both very similar to those observed in the higher risk group. There were too few lower risk subjects with moderate sleep apnea (i.e., only 2 subjects) to allow for a meaningful comparison between risk groups for this apnea severity category and there were no subjects from the lower risk group having severe sleep apnea.

Table 7.1. Multiple Sleep Latency Test (MSLT) values by severity of sleep apnea for CDL holders in the higher and lower risk groups.

RDI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	8.23	4.42	0.52	20.00	140	8.60	4.70	1.25	20.00
5 - <15	69	8.93	4.79	1.13	20.00	16	8.69	4.64	1.53	17.75
15 - <30	30	6.84	4.78	0.92	18.55	2	10.68	0.26	10.50	10.87
>= 30	28	4.80	2.74	0.22	11.80					
Total N	246					158				

Two-way analysis of variance (ANOVA) was used to assess if there were significant differences among the sleep apnea severity levels in mean MSLT values controlling for risk group. The differences in mean values among subjects with severe, moderate, mild, and no sleep apnea were highly statistically significant (F=6.03, df=3,399, p=0.0005). Both linear and non-linear (i.e., quadratic) trends were statistically significant (linear trend F=16.32, df=1, 399, p<0.0001 and quadratic trend F=4.70, df=1, 399, p=0.031). Results from the ANOVA are summarized in Table 7.2. Controlling for risk group, subjects with severe sleep apnea (AHI≥30 episodes per hour) had significantly smaller mean MSLT values compared to subjects with no sleep apnea (AHI<5) (F=13.4, df=1,399, p=0.0003), smaller than subjects with mild sleep apnea (5≤AHI<15) (F=16.5, df=1,399, p<0.0001), and smaller than subjects with moderate sleep apnea (15≤AHI<30) (F=3.72, df=1,399, p=0.054). Thus, increasing severity of sleep apnea is associated with increasing sleepiness on daytime testing, i.e., reduced sleep latency. Least squares estimation of mean values was used to determine adjusted mean values for each level of apnea severity adjusting for risk group taking into account the imbalance in the distributions of apnea severity between the risk groups. The least squares estimated mean values for MSLT adjusting for risk group were 4.98, 7.24, 9.00, and 8.42 minutes for subjects with severe, moderate, mild, and no sleep apnea, respectively. The pooled estimate of the standard deviation from the ANOVA was 4.52.

Table 7.2. Two-way ANOVA (apnea severity and risk group) for Multiple Sleep Latency Test scores. Significant differences were found among mean values among sleep apnea severity levels for holders of commercial vehicle drivers licenses from our population.

Test	F value	df	P value	Pair-wise contrasts
AHI (Overall)	6.03	3 399	0.001	≥30 vs. <5: p=0.031
Risk Group	0.49	1 357	0.482	≥30 vs. 5-<15: p=0.0003
AHI Linear Trend	16.32	1 357	0.000	≥30 vs. 15-<30: p=0.054
AHI Quadratic Trend	4.70	1 357	0.031	15-<30 vs. 5-<15: p=0.062
Risk Group * AHI Category Interaction	0.67	2 397	0.514	

Further analysis was performed examining the association of sleep apnea severity with percentage of subjects in each of the described MSLT categories (<5 minutes; 5-<10 minutes; ≥10 minutes). Figure 7.2 illustrates the clear elevation in risk for pathological sleepiness (i.e., an MSLT value less than 5 minutes) for subjects with severe and moderate levels of sleep apnea. A significant linear trend between sleep apnea severity as measured by AHI and MSLT category was observed by assessment using weighted data (χ^2 = 7.6, df=1, p=0.006). In subjects with severe apnea (AHI>30 episodes/hour), more than 50% of them have an MSLT value of <5.0 minutes. It is noteworthy that even in those without apnea (AHI<5 episodes/hour), just over 20% of subjects have an MSLT value that is in the range of marked excessive sleepiness.
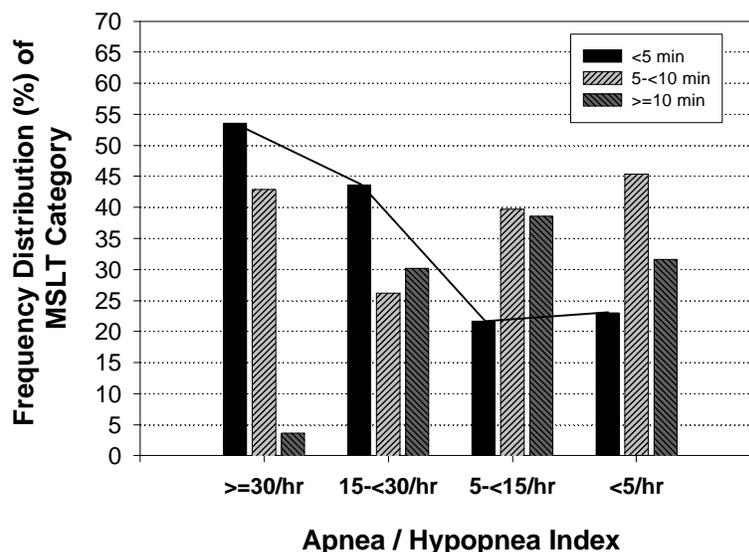
Figure 7.2. Percentages of subjects with Multiple Sleep Latency values <5, 5 to <10, and greater than or equal to 10 minutes by severity of sleep apnea. The solid line shows the clear elevation in risk for pathological sleepiness (i.e., MSLT<5 minutes) for subjects with severe and moderate severities of sleep apnea. Percentages were determined based on weighted data.

D. Association with Different Estimates of Prior Week's Mean Sleep Duration

As described in Chapter Six, as a result of concern that actigraphy-derived estimates of sleep time are sensitive to artifactual signals of wakefulness associated with sleep-disturbed breathing, we utilized two proxies for sleep duration. The main bout length of relative inactivity is less sensitive to such artifacts compared to the cumulative duration of inactivity in the main sleep bout because it depends only on identifying the time of sleep onset and the time of final sleep termination. However, the main bout length of relative inactivity may overestimate duration of sleep because it fails to exclude true wakefulness after sleep onset. In contrast, the cumulative duration of inactivity in the main sleep bout may underestimate the duration of sleep in subjects with apnea if movements during sleep associated with respiratory events are defined as wakefulness by actigraphic scoring. Subjects with reduced amounts of sleep during the week prior to neurobehavioral testing are hypothesized to have increased pressure for sleep. Reduced sleep durations across many nights lead to accumulating pressure for sleep and to increasing performance deficits across days [Dinges et al, 1997]. Thus, we hypothesized that subjects with diminished amounts of sleep in the week prior to MSLT testing as determined by the mean main bout length of relative inactivity or the mean cumulative duration of inactivity in the main sleep bout as determined by actigraphy would have reduced values of MSLT. To investigate this, while accounting for sampling by risk group, two-way analyses of variance were performed comparing mean MSLT values among sleep duration categories controlling for risk group. We first show data for the main bout length of relative inactivity. Descriptive statistics are shown in Table 7.3 and the related Analysis of Variance (ANOVA) in Table 7.4.

Table 7.3 provides descriptive statistics for mean MSLT by categories defined on the basis of mean main bout length of relative inactivity category. CDL holders in both the higher and lower risk groups with durations larger than 8 hours tended to have larger MSLT values.

Table 7.3. MSLT values by mean main bout length of relative inactivity category as determined by actigraphy during the prior week and by risk group.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	4.76	3.27	0.22	10.50	13	8.62	4.52	1.75	16.00
6<=hr< 7	53	8.10	4.53	0.75	17.87	24	8.25	5.27	2.12	20.00
7<=hr<= 8	88	8.23	4.66	0.75	20.00	63	8.39	4.10	1.25	20.00
hr>8	47	9.46	4.85	1.15	20.00	32	10.70	5.27	2.00	20.00
Total N	208					132				

Table 7.4 displays the two-way analysis of variance for the effect of mean main bout length of relative inactivity on MSLT. Differences in mean MSLT values among these durations of relative inactivity were statistically significant (p=0.001). Specifically, the linear trend in mean values was significant (p<0.001) but the non-linear component (i.e., quadratic trend) was not significant (p=0.822). Furthermore, on average, those with mean main bout length of relative

inactivity of less than 6 hours per week had significantly smaller mean MSLT values compared to those with any longer duration including those with average sleep durations between 7 and 8 hours (p=0.025).

Table 7.4. Two-way ANOVA (for mean sleep duration as determined by mean main bout length of relative inactivity category as determined by actigraphy during the prior week and by risk group for the MSLT. Significant reductions in mean MSLT values were observed in subjects with reduced mean sleep durations as determined by this variable, controlling for risk group.

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	5.38	3	335	0.001	<6 vs. >8; p<0.001
Risk Group	2.30	1	335	0.130	<6 vs. 7-8; p=0.025
Duration Linear Trend	14.24	1	335	0.000	<6 vs. 6-<7; p=0.045
Duration Quadratic Trend	0.05	1	335	0.822	6-<7 vs. >8; p=0.020
					7-8 vs. >8; p=0.009
Risk Group * Duration Category Interaction	1.56	3	332	0.198	

Table 7.5 provides descriptive statistics for the alternative proxy measure for sleep duration, namely the mean cumulative duration of inactivity in the main sleep bout. In both higher and lower risk groups, there were decreased mean latencies for subjects with mean durations of this variable of <5 hours and between 5 and 6 hours during the main sleep bout per day as compared to those whose mean cumulative duration of inactivity in the main sleep bout was 7 hours or more. As above, we carried out a two-way ANOVA to assess differences.

Table 7.5. MSLT values by mean cumulative duration of inactivity in the main sleep bout as determined by actigraphy during the prior week and by risk group.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	45	6.12	4.16	0.22	17.93	10	6.82	4.20	1.75	15.17
5<=hr< 6	47	8.04	4.89	0.75	20.00	28	9.23	5.40	1.25	20.00
6<=hr< 7	67	9.03	4.77	1.13	20.00	37	7.52	3.48	1.37	15.37
7-8(ref)	44	8.68	4.14	1.33	20.00	45	9.96	5.09	2.00	20.00
hr> 8	5	10.71	6.41	2.95	20.00	12	10.71	4.12	4.78	17.62
Total N	208					132				

Table 7.6 summarizes the two-way analysis of variance for the effect of mean cumulative duration of inactivity in the main sleep bout on MSLT. Differences in mean MSLT values among the sleep duration categories were statistically significant (p=0.002). Specifically, the linear trend in mean values was significant but the non-linear component (i.e., quadratic trend) was not significant. Furthermore, on average, those with mean sleep durations of less than 5

hours per week as determined by the mean cumulative duration of inactivity in the main sleep bout had significantly smaller MSLT values compared to those with any longer duration including those with average sleep durations between 5 and 6 hours (p=0.008).

Table 7.6. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout as determined by one week of actigraphy and risk group) for the MSLT. Significant reductions in mean MSLT values were observed in subjects with reduced mean sleep durations as determined by this measure, controlling for risk group.

Test	F value	df	P value	Pair-wise contrasts
Duration (Overall)	4.40	4 334	0.002	<5 vs. >8: p=0.001
Risk Group	0.23	1 334	0.632	<5 vs. 7-8: p=0.0002
Duration Linear Trend	12.03	1 334	0.001	<5 vs. 6-<7: p=0.005
Duration Quadratic Trend	0.11	1 334	0.746	**<5 vs. 5-<6: p=0.008**
				5-<6 vs. >8: p=0.087
Risk Group * Duration Category Interaction	1.36	4 330	0.248	6-<7 vs. >8: p=0.081

Further analysis was performed to examine the association between mean cumulative duration of inactivity in the main sleep bout and MSLT category (i.e., <5, 5-<10, and ≥10 minutes) (see Figure 3). The was a clear elevated risk for pathological sleepiness (i.e., an MSLT value less than 5 minutes) for subjects with mean cumulative durations of inactivity in the main sleep bout of less than 5 hours. A significant linear trend between sleep duration category and MSLT category was observed (χ^2 = 14.3, df=1, p<0.001).

Thus, data from both methods to assess sleep deprivation led to the same conclusion that chronic partial sleep deprivation occurs in commercial drivers, is relatively common, and is associated with increased physiologically measured sleepiness.
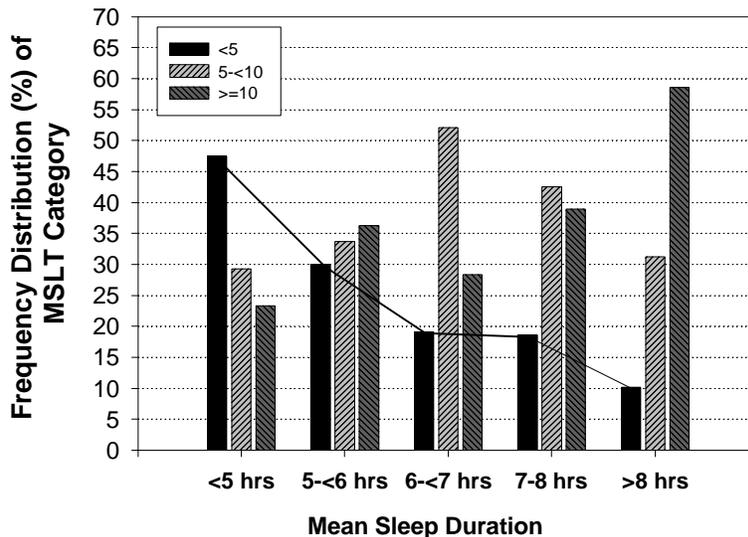
E. **Determinants of Multiple Sleep Latency**

To further evaluate the relative role of sleep apnea and sleep duration in an analysis analogous to that performed for the subjective outcomes (Chapter Six), multiple linear regression was used to simultaneously assess the effect of sleep apnea severity (>=30 vs. <5; 15-<30 vs. 5; and 5-<15 vs. 5 episodes per hour) and either mean main bout length of relative inactivity category (<6 hrs vs. >8 hrs; 6-<7 hrs vs. >8 hrs; and 7-<8 hrs vs. >8 hrs) or mean cumulative duration of inactivity in the main sleep bout category (<5 hrs vs. 7-8; 5-<6 vs. 7-8; 6-<7 vs. 7-8; and >8 vs. 7-8) on multiple sleep latency, while controlling for age, BMI, female gender, and a simple self-report health-related quality of life. The QoL was assessed on a 6-point scale (1=perfect health to 6=miserable). Sampling weights were used to account for sampling by risk group so that results better reflect populations associations. Age, BMI, and self-report quality of life were centered by subtracting their respective weighted mean value in order to make the intercept interpretable as the expected response value for a male with AHI<5 events per hour, sleep duration category set to its reference value (>8 hrs for main bout length of relative inactivity category and 7-8 hrs for mean cumulative duration of inactivity in the main sleep bout category), and with age, BMI, and QoL equal to the sample means (45.4 years, 29.9 kg/m^2, and 2.44 on the 6-point scale, respectively).

Table 7.7 presents the results for the model with use of the main bout length of relative inactivity category. The parameter estimate in the top row (labeled "Intercept") is the predicted MSLT value (minutes) for a male (age 45.5 years, BMI 29.9 kg/m^2) with AHI<5 events per hour and a mean main bout length of relative inactivity duration >8 hours. This value is 10.3 minutes, i.e., within the range that is typically considered 'normal'. The model suggests that an AHI of >30 episodes/hour reduces the average MSLT by 3.00 minutes as compared to that when AHI is less than 5 episodes/hour. Likewise, a mean main bout length of relative inactivity duration of <6 hours in duration reduces the expected value of MSLT by 2.95 minutes as compared to that when this duration is >8 hours. Thus, severe sleep apnea and chronic sleep durations of less than 6 hours (as determined by the mean bout length of relative inactivity) have roughly equivalent effects on reducing the latency to falling asleep.

Moreover, both the linear trend in MSLT values among apnea severity categories (p=0.011) and main bout length of relative inactivity (p=0.005) were simultaneously statistically significant reflecting their independent effect of daytime propensity for sleep. The linear trend for mean bout length of relative was statistically significantly (p=0.003) with no significant non-linear component (p=0.366). None of the control variables were significantly associated with differences in mean MSLT after controlling for apnea severity and sleep duration. However, adding frequent snoring (≥3 times per week) increased the overall p-value for AHI category from

0.060 to 0.099 and adding more than 2 alcoholic beverages per day increased the p-value for AHI category from 0.060 to 0.306.

Table 7.7. Multiple linear regression model for MSLT. Explanatory variables include sleep apnea severity, mean main bout length of relative inactivity category during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	10.2887	0.5683	0.0000	---
Age - 45.5 (yrs)	-0.0288	0.0228	0.2075	---
BMI - 29.9 (kg/m^2)	-0.0467	0.0551	0.3966	---
Female gender	0.8463	1.0457	0.4189	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.0600
Linear trend	---	---	---	0.0108
Quadratic trend	---	---	---	0.2053
AHI Model parameters	---	---	---	---
>30 vs. <5	-3.0039	1.3252	0.0241	---
15-30 vs. <5	-1.2715	1.1572	0.2727	---
5-<15 vs. <5	0.4735	0.6878	0.4917	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.0047
Linear trend	---	---	---	0.0031
Quadratic trend	---	---	---	0.3660
Bout Duration model parameters				
<6 vs. >8 hrs.	-2.9471	0.9716	0.0026	---
6-<7 vs. >8 hrs.	-2.0705	0.7771	0.0081	---
7-8 vs. >8 hrs.	-1.9354	0.6460	0.0029	---
Health related QoL Score	-0.0899	0.3059	0.7691	---
Model Summary	---	---	---	---
R-square	0.087	---	---	---
Root MSE	4.5707	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	-0.6053	0.6939	0.3837	0.0986
Current Smoking	0.3805	0.3079	0.2175	0.0406
Alcohol >2 drinks / day	0.6276	0.3926	0.1109	0.3055

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m^2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

The analysis was repeated replacing main bout length of relative inactivity category with mean cumulative duration of inactivity in the main sleep bout category. Results are displayed in Table 7.8. In this multivariable model, the linear trend in MSLT values among apnea severity categories remained statistically significant (p=0.035). The linear trend for mean duration of prior weeks sleep was also simultaneously significantly associated with differences among mean MSLT values (p=0.007). None of the control variables were significantly associated with differences in mean MSLT. Adding frequent snoring (≥3 times per week), current smoking, or having more than 2 alcoholic beverages per day did not change the results. The parameter estimate in the top row (labeled "Intercept") is the predicted MSLT value (minutes) for a male (age 45.5 years, BMI 29.9 kg/m^2) with AHI<5 events per hour and a mean sleep duration between 7 and 8 hours inclusive. This value is 9.6 minutes, or approximately 10 minutes, i.e., extremely close to the lower bound of what is typically considered 'normal'. The model results indicate that an AHI of >30 episodes/hour reduces the average MSLT by 2.52 minutes as compared to that when AHI is less than 5 episodes/hour. Likewise, a mean cumulative duration of inactivity in the main sleep bout of less than 5 hours reduces expected MSLT by 2.58 minutes as compared to that when sleep duration is between 7 and 8 hours. Thus, severe sleep apnea and chronic sleep durations of less than 5 hours had roughly equivalent effects on reducing the latency to falling asleep.

Table 7.8. Multiple linear regression model for MSLT. Explanatory variables include sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	9.563221	0.502209	0.0001	---
Age - 45.5 years	-0.017543	0.023059	0.4474	---
BMI - 29.9 kg/m^2	-0.018040	0.056062	0.7478	---
Female gender	1.011429	0.077812	0.3487	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	**0.1450**
Linear trend	---	---	---	**0.0349**
Quadratic trend	---	---	---	0.2698
AHI Model parameters	---	---	---	---
>30 vs. <5	-2.523041	1.365759	0.0656	---
15-30 vs. <5	-1.104668	1.161249	0.3422	---
5-<15 vs. <5	0.522730	0.696404	0.4534	---
Sleep Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	**0.0210**
Linear trend	---	---	---	**0.0073**
Quadratic trend	---	---	---	0.9920
Sleep Duration model parameters	---	---	---	---
<5 vs. 7-8 hrs.	-2.575045	0.889336	0.0040	---
5-<6 vs. 7-8 hrs.	-0.905932	0.744560	0.2246	---
6-<7 vs. 7-8 hrs.	-1.421839	0.664800	0.0332	---
>8 vs. 7-8 hrs.	0.686399	1.147406	0.5501	---
Health related QoL Score	-0.044810	0.305826	0.8836	---
Model Summary	---	---	---	---
R-square	0.083	---	---	---
Root MSE	4.587810	---	---	---
Added variables				AHI p-value
Frequent Snoring (>=3/wk)	-0.625205	0.690654	0.3661	0.2757
Current Smoking	0.283620	0.310656	0.3620	0.1017
Alcohol >2 drinks / day	0.696719	0.396915	0.0806	0.4295

F. Exploratory Models: Impairment Risk Model for Multiple Sleep Latency

The objective of the multivariable model described in the previous section was to control for potentially confounding variables including prior week's mean sleep duration, age, gender, obesity, and general perceived health-related quality of life as well as other factors for the association between AHI severity category and MSLT. That is, these analyses helped to validate as well as to extend the results from the two-way analyses of variance described in the prior

section of this chapter. In a new analysis, we turn our attention to exploratory modeling of the relationship between apnea severity as measured on interval scale with presence versus absence of impairment. For multiple sleep latency, we defined impaired as the presence of pathological excessive sleepiness, i.e., a multiple sleep latency test of an average of less than 5.0 minutes. The results of models are shown in Figures 7.4 and 7.5, when sleep duration was measured by mean main bout length of relative inactivity category (Figure 7.4) and cumulative duration of inactivity in the main sleep bout category (Figure 7.5). These graphs display the probability of impairment (MSLT <5 minutes) as a function of apnea/hypopnea index. Curves are shown for various durations of sleep duration (see legend). With increasing apnea severity, there is an increased probability of impairment at any duration of sleep. However, the probability of impairment is strongly dependent on duration of sleep. For example, at an apnea/hypopnea index of 30 episodes/hour the probability of impairment goes from the order of 0.25 to 0.60 as the mean main bout length of relative inactivity category falls from >8 hours/night to less than 6 hours/night. Similar results were obtained when sleep duration was measured by mean cumulative duration of inactivity in the main sleep bout (Figure 7.5).
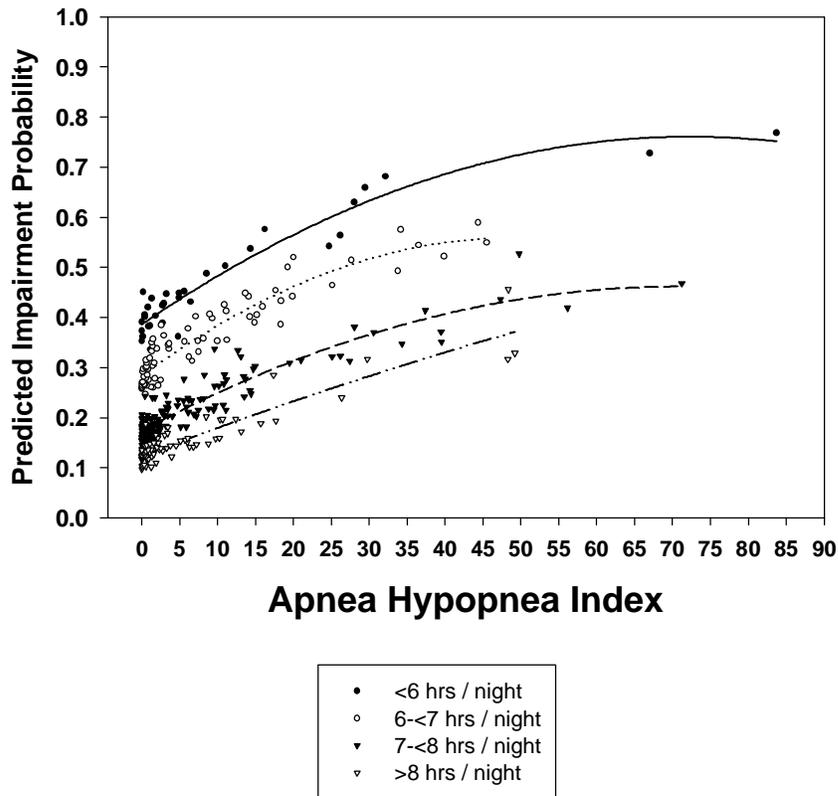
Figure 7.4. Predicted probability of MSLT <5 minutes by AHI and mean main bout length of relative inactivity category. Additional variables in the model include age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 24.4%. For other details see text.
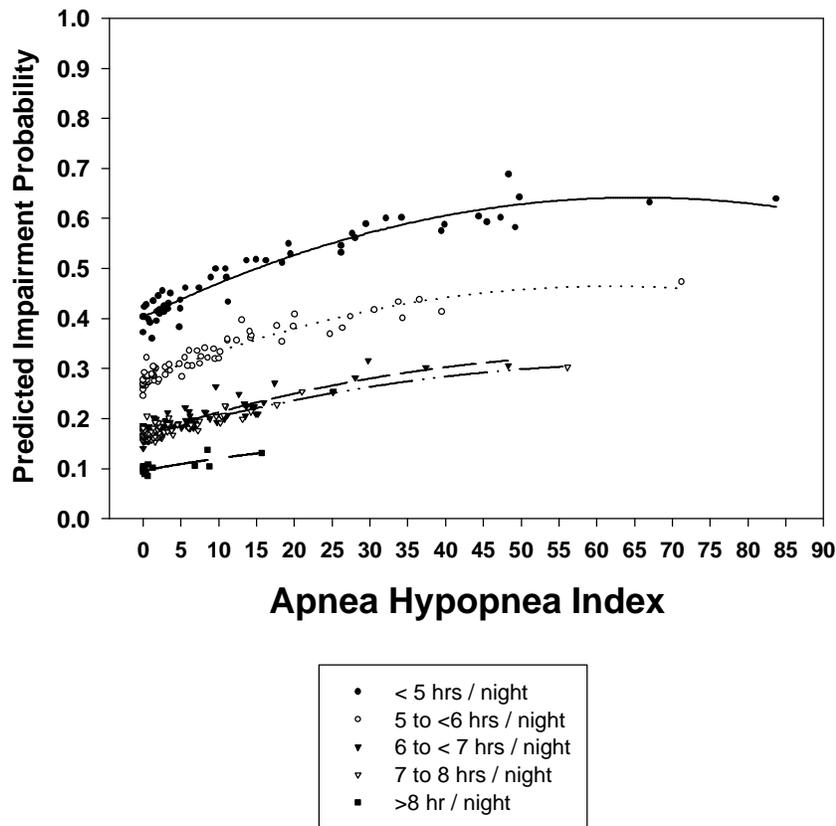
Figure 7.5. Predicted probability of MSLT <5 minutes by AHI and mean cumulative duration of inactivity in the main sleep bout category. Additional variables in the model include age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 24.4%. For other details see text.

7.2. Psychomotor Vigilance Task (PVT)

Description of Test

The psychomotor vigilance task (PVT) is a "simple" (as opposed to multiple task), portable reaction time (RT) test designed to evaluate the ability to sustain attention and respond in a timely manner to salient signals [Dinges and Powell, 1985]. The task involves measuring reaction time, i.e., speed to respond, to signals presented at random intervals over a brief period of testing. Sleepiness results in slowed response, further slowing of response over time on the task and frank performance lapses, i.e., failure to respond to a signal. These different aspects of PVT performance have been demonstrated to be highly sensitive to alertness associated with circadian phase, i.e., across the day [Dinges and Kribbs, 1991; Wyatt et al, 1997]; to be degraded by acute total sleep deprivation [Dinges et al, 1994], and cumulative partial sleep loss [Dinges et al, 1997]; PVT performance is also impaired in disorders of excessive sleepiness [Kribbs and Dinges, 1994]; sleepiness in the elderly [Samuel et al, 1996]; shift work [Rosekind et al, 1994];

the demands of medical house staff [Geer et al, 1995]; and other conditions that modulate the ability to sustain attention and respond efficiently [Dinges and Kribbs, 1991]. The PVT was designed to be simple to perform; free of a learning curve or influence from acquired skills (aptitude, education); and highly sensitive to an attentional process that is fundamental to normal alert functioning. The PVT task is contained in a small, programmable, portable, electronic box (20cm x 11cm x 6cm) that requires only a single switch to start. The task consists of responding to a small, bright red light stimulus (LED-digital counter) by pressing a response button as soon as the stimulus appears. When the response occurs, the stimulus is stopped and the counter displays the reaction time (RT) in milliseconds. The interval between adjacent stimuli varies randomly from 2 sec. to 10 sec., and the overall task duration is 10 minutes (which yields approximately 80 RTs per trial). The subject is instructed to press the button as soon as each stimulus appears, in order to keep the reaction time as low as possible, but not to press the button too soon (which yields a false start [FS] warning on the display).

Special software is used to extract multiple performance parameters from each PVT trial [Dinges and Powell, 1985; Dinges et al, 1987; Rosekind et al, 1994; Kribbs and Dinges, 1994]. Primary response parameters include: (1) the *median response time* (msec); (2) the *frequency of lapses* (which refer to the number of times the subject fails to respond to the signal or fails to respond in a timely manner); (3) the *duration of lapse domain*, which refers to shifts in lapse duration calculated from the 10% slowest RTs (a metric that reflects vigilance response slowing); (4) the *optimum response times*, computed as the average of the 10% fastest RTs per trial (reflecting the best performance an operator is capable of producing); and (5) the *fatigability function or slope*, which refers to the vigilance decrement function or the extent to which subjects maintained performance across time on task. These PVT performance variables can be compared either (a) to values from the same subjects at different times (e.g., pre- versus post-treatment); or (b) to values from known populations to obtain an estimate of performance capability relative to other occupations or work schedules; or (c) to effects in performance capability of subjects in controlled laboratory studies from impairment produced by different challenges, e.g., different durations of sleep deprivation. The compact, portable nature of the PVT, its simplicity as a task, and its documented sensitivity to the fundamental neurobehavioral function of sustained attention, make it an ideal technology for tracking the effects of disease and/or behavior on functional vigilance.

Subjects had 1 or 2 practice sessions with the PVT on the evening prior to their overnight polysomnography. The primary day of in-lab neurobehavioral testing took place on the day following the overnight polysomnography. PVT tests were conducted four times, at approximately 8:00 AM, 10:00 AM, 12:00 PM, and 2:00 PM. The mean response for each parameter across times-of-day were computed to simplify the analyses.

7.3 Psychomotor Vigilance Task (PVT) - Median Response Time

A. Distribution of Median Response Times in Higher and Lower Risk Groups

The first variable we extracted from the PVT studies was the median response time. Figure 7.6 displays the mean and standard deviation of PVT median response times (msec) for the higher and lower risk groups (i.e., averaging over the four trials). There was some evidence

that median response times differed by risk group (Welch's test for groups with unequal variances, $t=-1.76$, df=364.0, p=0.079). The population mean value was estimated as the weighted average of within group means and was found to be equal to 265.3 msec. The pooled estimate of the SD (weighted according to degrees-of-freedom) was 40.2 msec.
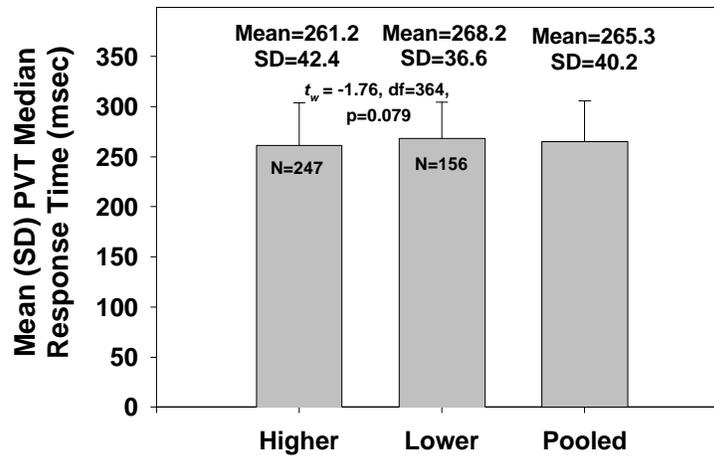
Figure 7.6. Psychomotor Vigilance Task (PVT) mean and SD values for the median reaction times by risk group and overall. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal within risk group SD's and is computed using the usual formula for computing a pooled SD estimate.

As noted above, analyses of PVT responses utilized the mean over 4 trials. This provided data for a formal reliability analysis. Therefore, random effects and mixed effects analyses of variance (ANOVA) were used to assess reliability by computing an intraclass correlation coefficient (ICC) [Fleiss, 1986; Dunn, 1992]. The ICC represents the proportion of total variance attributable to true differences among subjects (as opposed to differences within subject). The ICC is the reliability coefficient for a single measurement. By application of the Spearman-Brown prophecy formula [Carmines and Zeller, 1979], we can compute the reliability for the mean of *N* PVT median values as:

$$\text{Reliability} = (N * \text{ICC}) / \{(1 + (N-1)*\text{ICC}]$$

where *N* is equal to the number of PVT median values averaged together to produce a mean score.

However, since the first PVT assessment of the day is potentially affected by a sleep inertia effect, ICC values were computed with and without controlling for time-of-day as a fixed effect in the ANOVA. ICC values for PVT median were 0.544 and 0.536 controlling for and not controlling for a fixed effect of time of day. Thus, the reliability coefficient for the mean of 4 replications (with or without controlling for time of day) was about 0.82. Landis and Koch [1977] have characterized ICC values as follows: slight (0-0.20), fair (0.21-0.40), moderate (0.41-0.60), substantial (0.61-0.80), and almost perfect (0.81-1.00). Thus, the formal reliability analysis demonstrated that by conventional criteria, the sum of 4 PVT median values used to

assess function had excellent reliability. A single PVT median value appears of have 'moderate' reliability.

B. Association Between Median Response Times and Sleep Apnea Severity

We next assessed the association between median response time on the PVT and apnea severity as measured by AHI. Table 7.9 summarizes the mean (SD) PVT median response time values by sleep apnea severity separately for subjects in the higher and lower risk groups. In the higher risk group, the mean of the median RTs was 9.4% higher among subjects with severe sleep apnea compared to those with AHI values <5 episodes per hour and 12.9% higher compared to those with mild sleep apnea (i.e., at least 5 but less than 15 episodes per hour).

Table 7.9. PVT median response times (msec) by severity of sleep apnea category for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	260.0	35.8	183.6	400.0	137	268.8	38.0	196.0	439.6
5 - <15	69	251.9	25.4	206.0	328.4	16	256.6	18.4	229.8	302.4
15 - <30	29	266.1	27.7	222.2	334.6	2	307.4	25.5	289.4	325.4
>= 30	28	284.5	86.4	219.2	688.0					

Two-way analysis of variance (ANOVA) was used to assess if there were significant differences among the sleep apnea severity levels for PVT median response times controlling for risk group (see Table 7.10). Differences among the mean values were highly statistically significant (F=5.29, df=3,395, p=0.001) among the categories of sleep apnea severity. Both linear (p=0.0006) and non-linear (p=0.030) components of the trend in differences among the means were significant, reflecting the apparent "threshold" pattern observed in Table 7.9. That is, median response times tended to become only substantially larger for subjects with severe sleep apnea, i.e., AHI≥30 episodes/hour. Least squares estimates of mean values (accounting for risk group) were 264.4, 255.6, 272.6, and 288.9 msec, respectively, for subjects with AHI values <5, 5-<15, 15-<30, and ≥30 episodes per hour. Significant pairwise differences were observed between subjects with AHI≥30 vs. <5 (p=0.003) and between AHI≥30 vs. 5-<15 (p<0.0001) but not between AHI≥30 vs. 15-<30 (p=0.111) events per hour. In addition, a significant difference in median response times was observed between subjects with moderate sleep apnea compared to those with mild sleep apnea (AHI 15-<30 vs. AHI 5-<15; p=0.042). Figure 7.7 illustrates visually the least squares mean value estimates for PVT median responses controlling for risk group in the different categories of apnea severity. The increase in the median response time in those with severe sleep apnea is easily seen.

Table 7.10. Two-way ANOVA (apnea severity and risk group) for PVT median response times. Controlling for risk group, statistically significant increases in median response times were associated with severe sleep apnea compared to no sleep apnea or only mild sleep apnea.

Test	F value	df	P value	Pair-wise contrasts
AHI (Overall)	5.29	3 395	0.001	≥30 vs. <5: p=0.030
Risk Group	3.90	1 395	0.049	≥30 vs. 5-<15: p<0.0001
AHI Linear Trend	11.83	1 395	0.001	≥30 vs. 15-<30: p=0.111
AHI Quadratic Trend	4.74	1 395	0.030	15-<30 vs. 5-<15: p=0.042
Risk Group * AHI Category Interaction	0.70	2 393	0.499	
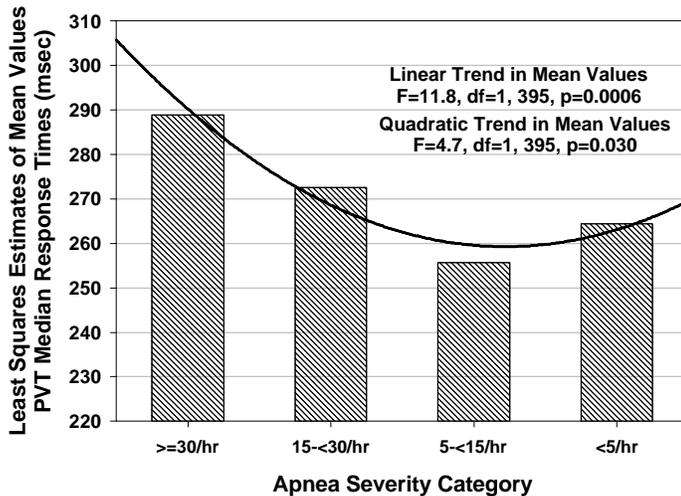
Figure 7.7. Least squares estimated mean values for PVT median response times (controlling for risk group) as a function of apnea severity category. After controlling for risk group, significant linear and quadratic components of the trend in means emerge. The solid line is from a least squares fit of the mean values that included a quadratic trend component.

C. Association Between PVT Median Response Time with Different Estimates of Mean Sleep Duration

We next assessed the association between this metric from the PVT (median response time) and mean sleep duration using our two measures, i.e. mean main bout length of relative inactivity category and mean cumulative duration of inactivity in the main sleep bout category as determined from actigraphy at home. Tables 7.11 and 7.13 summarizes the PVT median response times in different categories of mean sleep duration for these two measures in both

higher risk and lower risk groups. Tables 7.12 and 7.14 summarize the corresponding two-way analyses of variance used evaluate the relationship between these variables.

Table 7.11. Descriptive statistics for PVT median response times by mean main bout length of relative inactivity category during the prior week and risk group.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	253.8	39.5	221.0	400.0	13	272.2	63.5	205.8	439.6
6<=hr< 7	53	263.1	37.4	213.6	386.6	23	265.8	36.1	196.0	328.8
7<=hr<= 8	87	262.2	53.8	197.2	688.0	61	265.5	25.6	219.6	349.5
hr>8	47	258.9	33.2	183.6	364.2	32	265.4	35.1	210.8	351.8
Total N	207					129				

Table 7.12. Two-way ANOVA (mean sleep duration and risk group) for PVT median response times. There was not a significant association between PVT median response times and sleep duration when measured by mean main bout length of relative inactivity category.

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	0.10	3	331	0.961	
Risk Group	1.39	1	331	0.239	
Duration Linear Trend	0.00	1	331	0.990	
Duration Quadratic Trend	0.27	1	331	0.603	
Risk Group * Duration Category Interaction	0.32	3	328	0.810	

Table 7.13. Descriptive statistics for PVT median response times by mean cumulative duration of inactivity in the main sleep bout during the prior week in the two different risk groups (higher and lower risk).

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	44	274.4	74.8	206.0	688.0	10	261.5	35.9	205.8	326.6
5<=hr< 6	47	259.4	34.0	213.6	386.6	28	270.0	46.0	196.0	439.6
6<=hr< 7	67	260.3	26.2	214.4	328.4	35	267.3	30.3	219.6	327.6
7-8(ref)	44	251.5	34.3	183.6	364.2	44	266.1	31.1	210.8	351.8
hr> 8	5	245.3	13.0	231.8	260.2	12	258.6	33.6	222.4	334.4
Total N	207					129				

Table 7.14. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout and risk group) for PVT median response times. The linear trend in the mean values was statistically significant (p=0.045).

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	1.48	4	330	0.208	<5 vs. >8: p=0.064
Risk Group	2.99	1	330	0.085	<5 vs. 7-8: p=0.029
Duration Linear Trend	4.05	1	330	0.045	
Duration Quadratic Trend	0.03	1	330	0.870	
Risk Group * Duration Category Interaction	0.72	4	326	0.582	

Results differed between these two measures of sleep duration. There was no significant association between PVT median response times and mean main bout length of relative inactivity category (p=0.0961). In contrast, there was a significant linear trend in PVT median response times for shorter cumulative durations of inactivity in the main sleep bout (p=0.045). The least squares mean values for PVT median response times were 253.0, 258.8, 264.0, 264.4, and 274.6 msec for subjects with different durations of cumulative inactivity, i.e., >8 hours, 7-8 hours, 6-<7 hours, 5-<6 hours, and <5 hours, respectively. The pooled estimate of the standard deviation was 40.8 msec. The least squares mean values are illustrated Figure 7.8. As is evident, a clear linear trend emerges in the mean values when controlling for risk group.
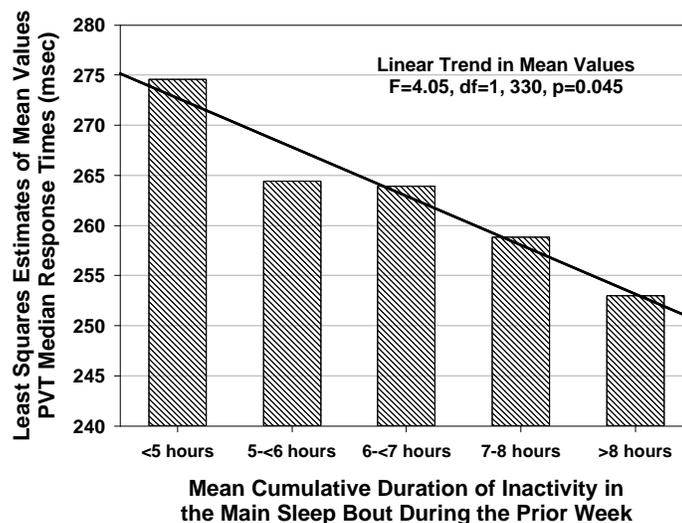
Figure 7.8. Psychomotor Vigilance Task (PVT) least squares estimated mean values for median response time controlling for risk group. After controlling for risk group, a clear significant linear trend representing differences among mean values emerge when sleep duration was assessed by mean cumulative duration of inactivity in the main sleep bout. The solid line is from a least squares fit of the mean values.

D. <u>Determinants of PVT Median Response Times</u>

The same type of multivariable model as we employed for MSLT was used to assess variables that determined PVT median response times.   Results are summarized in Tables 7.15 and 7.16.

For the model using mean main bout length of relative inactivity (see Table 7.15), significant differences in expected median response times were observed (Partial F=3.7, df=3,318, p=0.012) among sleep apnea severity groups controlling for the mean cumulative duration of inactivity in the main sleep bout in the week prior to testing, and the other control variables in the model. The (adjusted) expected increase in the median response time for an individual with AHI≥30 compared to an individual with AHI<5 was 23.4 msec. Similarly, the (adjusted) expected increase in median response time for an individual with AHI≥30 compared to an individual with an AHI at least equal to 5 but less than 15 was 35.1 msec.   The simultaneous effect of mean sleep duration as measured by the mean main bout length of relative inactivity category was not statistically significant in this model (p=0.823). Nor was there a significant linear trend for this variable (p=0.702). However, this model did expose statistically significant differences (p=0.016) between males and females in expected median response times. The expected median response time was 21.2 msec longer for females than for males controlling for sleep apnea severity, average sleep duration, age, BMI, and the QoL index.

Table 7.15. Multiple linear regression model for PVT median response time (msec). Explanatory variables included sleep apnea severity, mean main bout length of relative inactivity category during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	259.8719	4.7387	0.0000	---
Age - 45.5 (yrs)	0.1869	0.1904	0.3271	---
BMI - 29.9 (kg/m^2)	0.1365	0.4594	0.7666	---
Female gender	21.1579	8.7207	0.0158	---
<u>Apnea Hypopnea Index</u>	---	---	---	---
Overall difference in means	---	---	---	0.0124
Linear trend	---	---	---	0.0161
Quadratic trend	---	---	---	0.0283
<u>AHI Model parameters</u>	---	---	---	---
>30 vs. <5	23.4142	11.0510	0.0349	---
15-30 vs. <5	2.9085	9.8311	0.7675	---
5-<15 vs. <5	-11.6657	5.7433	0.0431	---
<u>Sleep Bout Duration (actigraphy)</u>	---	---	---	---
Overall difference in means	---	---	---	0.8283
Linear trend	---	---	---	0.7020
Quadratic trend	---	---	---	0.5327
<u>Bout Duration model parameters</u>				
<6 vs. >8 hrs.	3.3564	8.1018	0.6789	---
6-<7 vs. >8 hrs.	4.5170	6.5194	0.4889	---
7-8 vs. >8 hrs.	4.9547	5.4077	0.3602	---
Health related QoL Score	4.8685	2.5586	0.0580	---
<u>Model Summary</u>	---	---	---	---
R-square	0.069	---	---	---
Root MSE	38.1018	---	---	---
<u>Added variables</u>				**AHI p-value**
Frequent Snoring (>=3/wk)	-5.5882	5.5483	0.3146	0.0058
Current Smoking	5.1269	2.5628	0.0463	0.0207
Alcohol >2 drinks / day	-2.1712	3.3229	0.5140	0.1435

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

     The analyses described above were also performed using mean cumulative duration of inactivity in the main sleep bout. Results are summarized in Table 7.16. Significant differences in expected median response times were observed (Partial F=3.4, df=3,317, p=0.019) among sleep apnea severity groups controlling for prior weeks average sleep duration and the other control variables in the model. The (adjusted) expected increase in the median response time for an individual with AHI≥30 compared to an individual with AHI<5 was 22.1 msec. Similarly, the (adjusted) expected increase in median response time for an individual with AHI≥30 compared

to an individual with an AHI at least equal to 5 but less than 15 was 34.0 msec. The simultaneous effect of mean cumulative duration of inactivity, however, was not statistically significant in this model (p=0.592). Nor was there a significant linear trend (p=0.125). This model also exposed statistically significant differences (p=0.008) between males and females in expected median response times. The expected median response time was 24.1 msec larger for females than for males controlling for sleep apnea severity, average sleep duration, age, BMI, and the QoL index.

Table 7.16. Multiple linear regression model for PVT median response time (msec). Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout category during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	261.642655	4.195320	0.0000	---
Age - 45.5 years	0.162511	0.191384	0.3964	---
BMI - 29.9 kg/m^2	0.054902	0.465394	0.9062	---
Female gender	24.085206	8.942522	0.0074	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	**0.0187**
Linear trend	---	---	---	**0.0273**
Quadratic trend	---	---	---	0.0288
AHI Model parameters	---	---	---	---
>30 vs. <5	22.065960	11.328490	0.0523	---
15-30 vs. <5	1.722918	9.777020	0.8602	---
5-<15 vs. <5	-11.927990	5.784725	0.0400	---
Sleep Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	**0.5917**
Linear trend	---	---	---	**0.1251**
Quadratic trend	---	---	---	0.3105
Sleep Duration model parameters	---	---	---	---
<5 vs. 7-8 hrs.	4.537618	7.412547	0.5409	---
5-<6 vs. 7-8 hrs.	3.948214	6.192289	0.5242	---
6-<7 vs. 7-8 hrs.	3.179282	5.580734	0.5693	---
>8 vs. 7-8 hrs.	-11.042129	9.526249	0.2473	---
Health related QoL Score	4.529263	2.542812	0.0758	---
Model Summary	---	---	---	---
R-square	0.074727	---	---	---
Root MSE	38.047165	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	-6.093722	5.519613	0.2704	0.0161
Current Smoking	5.459533	2.564497	0.0340	0.0252
Alcohol >2 drinks / day	-2.214915	3.341390	0.5079	0.2273

E. <u>Exploratory Models: Risk of Impaired Function Defined by PVT Median Response Time > 275 msec</u>

Figure 7.9 displays the same exploratory model for impairment but uses a median response time of >275 msec as the criterion for impaired performance. For comparison purposes Powell et al [1999] provided mean PVT response data for 35 male subjects before and after consumption of alcohol. Mean breath alcohol (SD) at the start of different trials with different amounts of alcohol were 0 (0), 0.059 (0.0), 0.08 (0.01), and 0.084 (0.01) grams/200 liters (BrAC), respectively. This is equivalent to blood alcohol concentration of grams of alcohol per 100 milliliter of blood (BAC, g/dl). Values above 0.04 g/200 liters make it illegal to drive in California. Mean reaction times (SD) for these trials were 236 (19), 251 (26), 264 (36), and 271 (38), respectively. Thus, our criterion value for defining impairment (median RT>275) is similar to the mean value obtained in the Powell et al when alcohol consumption was over twice the legal limit. For this PVT parameter, is appears that the effect of reduced sleep duration as determined by mean cumulative duration of inactivity in the main sleep bout is not as pronounced relative to the effect of severe sleep apnea.

•	<5 hrs / night
○	5 to <6 hours / night
▼	6 to < 7 hrs / night
▽	7 to 8 hrs / night
■	>8 hr / night
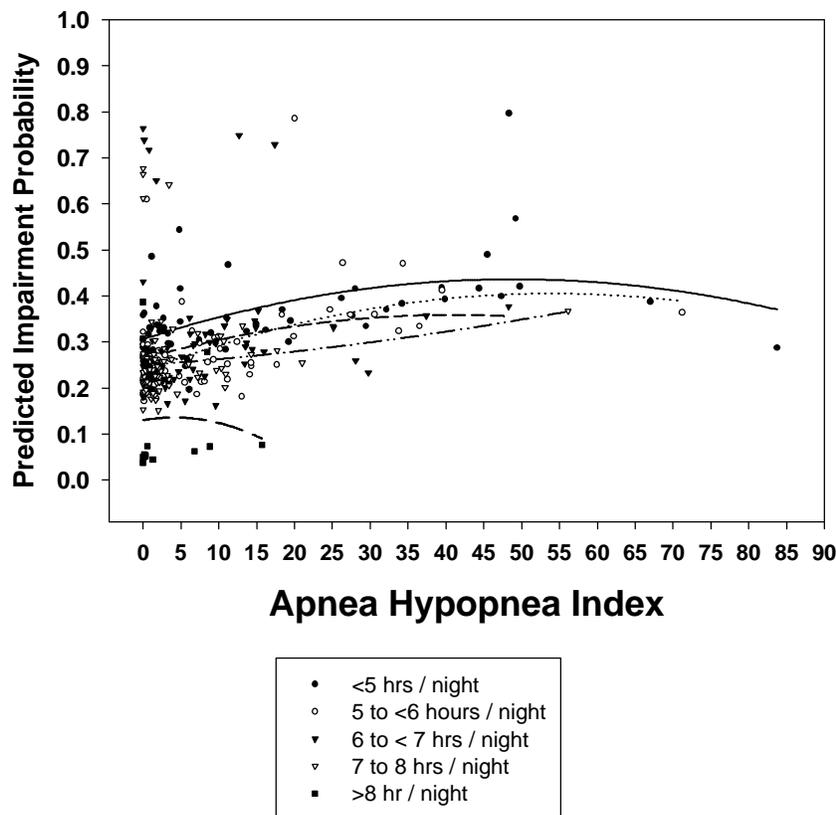
<u>Figure 7.9</u>. Predicted probability of PVT median response time per trial > 275 msec by AHI and mean cumulative duration of inactivity in the main sleep bout using different categories. Additional variables in the model include age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 30.6%.

7.4 Psychomotor Vigilance Task (PVT) - Vigilance Lapses

A. Distribution of Vigilance Lapses per Trial in the Higher and Lower Risk Groups

Figure 7.10 displays the mean and standard deviation for the mean total PVT vigilance lapses per trial for the higher and lower risk groups. As with median reaction times, there was some evidence of differences between risk groups (t=-1.95, df=400, p=0.051). The population mean value was estimated as the weighted average of within group means and was found to be equal to 2.92 lapses per trial. The pooled estimate of the SD was 3.96. The ICC analyses yielded reliability coefficient values of 0.529 and 0.521, respectively, with and without controlling for time-of-day. Thus, the reliability for the mean total lapses among the 4 trials was 0.81, a value very similar to that observed for median response time. Thus, as with median response time, reliability for our mean over four trials was 'almost perfect' using the criterion proposed by Landis and Koch [1977] while reliability would have been 'moderate' had a single measurement had been used to assess performance.
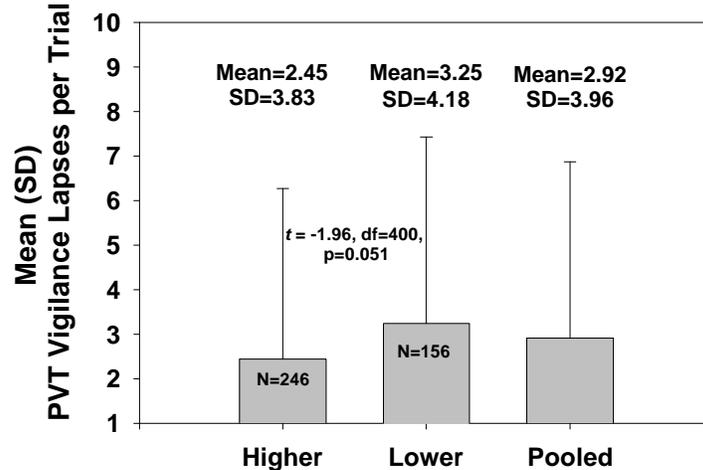
Figure 7.10. Psychomotor Vigilance Task (PVT) mean and SD values for the mean total vigilance lapses per trial by risk group and overall. The pooled mean is a weighted average computed as 0.415 times the higher risk mean and 0.585 times the lower risk mean. The pooled standard deviation assumes equal true within risk group SD's and is computed using the usual formula for computing a pooled SD estimate.

B. Association with Between Total Vigilance Lapses per Trial and Sleep Apnea Severity

Table 7.17 summarizes the mean (SD) PVT total vigilance lapses per trial by sleep apnea severity for subjects in the higher and lower risk groups. In the higher risk group, the mean value was 87.4% higher among subjects with severe sleep apnea compared to those with AHI value <5 episodes per hour and 215.2% higher compared to those with mild sleep apnea (i.e., at least 5 but less than 15 episodes per hour).

Table 7.17. PVT mean total vigilance lapses per trial by severity of sleep apnea for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	2.54	3.43	0.00	25.00	137	3.44	4.46	0.00	27.00
5 - <15	69	1.51	1.40	0.00	7.20	16	1.43	0.95	0.20	3.00
15 - <30	29	2.21	1.83	0.00	5.80	2	5.40	1.13	4.60	6.20
>= 30	28	4.76	8.06	0.00	39.50					
Total N	245					155				

Two-way analysis of variance (ANOVA) was used to assess if the observed differences for mean PVT total lapses per trial were statistically significant among the sleep apnea severity levels, controlling for risk group (see Table 7.18). Differences among the mean values were highly statistically significant ($F=5.59$, df=3,395, $p=0.0009$) among apnea severity categories. Both linear and non-linear components of the trend in differences among the means were significant, reflecting the apparent "threshold" pattern. That is, the mean value was substantially larger for subjects with severe sleep apnea (AHI≥30 episodes/hour). Least squares estimates of mean values (accounting for risk group) were 2.99, 1.74, 2.76, and 5.16 lapses per trial for subjects with AHI value <5, 5-<15, 15-<30, and ≥30 episodes per hour, respectively, with a residual SD of 3.93. Significant pairwise contrasts were observed between subjects with AHI≥30 vs. <5 ($p=0.009$) and between AHI≥30 vs. 5-<15 ($p<0.0001$) episodes per hour and in contrast to results for median response time, between AHI≥30 vs. 15-<30 ($p=0.020$). Standardized effect sizes comparing subjects with AHI values ≥30 vs. <5 and ≥30 vs. 5-<15 were 0.55 and 0.87, respectively. The latter value (i.e., 0.87) is typically considered very large. The apparent 'U-shape' in the profile of mean values appeared significant with a statistically significant increase for subjects with AHI<5 relative to those with AHI at least equal to 5 but less than 15 episodes ($p=0.016$). Figure 7.11 illustrates the pattern of mean total lapses per trial as a function of severity of sleep apnea. As can be seen, performance lapses increase in subjects with severe sleep apnea (AHI≥30 episodes/hour). There was no evidence that the relationship between PVT lapses and apnea severity differed by risk group (interaction $p=0.503$).

Table 7.18. Two-way ANOVA (apnea severity and risk group) for PVT mean total vigilance lapses per trial. Controlling for risk group, statistically significant increases in total vigilance lapses were associated with severe sleep apnea.

Test	F value	df		P value	Pair-wise Contrast
AHI (Overall)	5.59	3	395	0.001	≥30 vs. <5: p=0.009
Risk Group	3.19	1	395	0.075	≥30 vs. 5-<15: p=0.0001
AHI Linear Trend	8.33	1	395	0.004	≥30 vs.15-<30: p=0.020
AHI Quadratic Trend	10.16	1	395	0.001	5-<15 vs. <5: p=0.016
Risk Group * AHI Category Interaction	0.69	2	393	0.503	
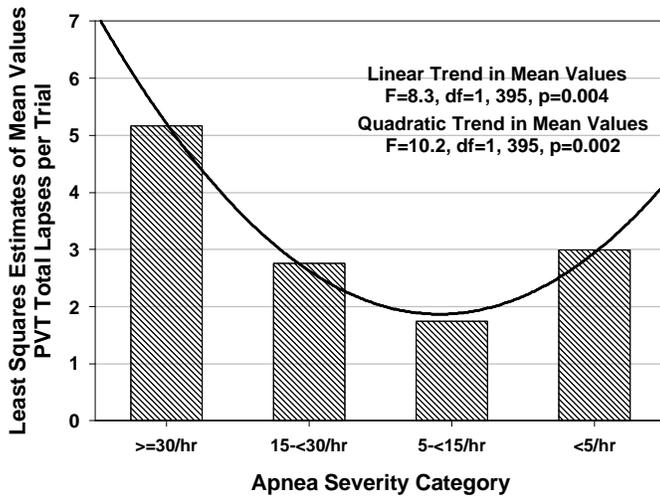
Figure 7.11. Least squares estimated mean values for PVT total vigilance lapses per trial (controlling for risk group) as a function of apnea severity category. After controlling for risk group, significant linear and quadratic components of the trend in means emerge. The solid line is from a least squares fit of the mean values that included a quadratic trend component.

C. Association Between Total Vigilance Lapses per Trial and Different Estimates of Mean Sleep Duration

Tables 7.19 and 7.21 provide descriptive statistics describing the relationships between the PVT vigilance lapses per trial with the mean sleep duration during the prior week as determined by our two approaches for assessment. The ANOVA analysis for the two measures of duration are shown in Tables 7.20 and 7.22, respectively.

Table 7.19.   Descriptive statistics for PVT mean total lapses per trial by mean main bout length of relative inactivity category during the prior week and risk group.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	3.05	5.36	0.00	25.00	13	5.54	7.55	0.20	26.20
6<=hr< 7	53	2.90	4.16	0.20	23.20	23	2.89	2.76	0.00	11.60
7<=hr<= 8	87	2.18	4.35	0.00	39.50	61	2.54	2.79	0.00	14.25
hr>8	47	2.26	2.58	0.00	13.80	32	3.45	4.30	0.00	20.40
Total N	207					129				

When categorized according to mean main bout length of relative inactivity both risk groups demonstrated an increase in lapses when the average duration was less than 6 hours. The analysis of variance is summarized in Table 7.20.

Table 7.20. Two-way ANOVA (mean bout length of relative inactivity category sleep duration and risk group) for mean PVT total lapses per trial. There was no significant interaction between PVT lapses and mean bout length of relative inactivity. There was a trend toward statistical significance for a linear trend in the mean values (p=0.076).

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	1.77	3	331	0.153	
Risk Group	2.40	1	331	0.122	
Duration Linear Trend	3.16	1	331	0.076	
Duration Quadratic Trend	2.18	1	331	0.141	
Risk Group * Duration Category Interaction	0.88	3	328	0.454	

The least squares estimated mean values controlling for risk group were 4.10, 3.04, 2.39, and 2.81 lapses per trial, respectively, for subjects with <6 hours, 6-<7 hours, 7-8 hours, and >8 hours average duration of their main bout length of relative inactivity, respectively. The pooled estimate of the standard deviation was 4.01. The linear trend in mean values approached statistical significance (F=3.16, df=1,331, p=0.076) while the non-linear component did not appear statistically significant (p=0.141). The standardized effect size comparing subjects with a mean duration of <6 hours to those with a mean duration of 7-<8 hours was (4.10-2.39)/4.01 or 0.43, a value usually considered to be at most, of moderate magnitude.

A stronger association between PVT lapses and duration was found when sleep duration was characterized by mean cumulative duration of inactivity in the main sleep bout category (Tables 7.21 and 7.22).

Table 7.21. Descriptive statistics for PVT mean total lapses per trial by mean cumulative duration of inactivity in the main sleep bout in different categories as measured during the week prior to testing and risk group (higher and low risk groups).

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	44	4.49	7.55	0.00	39.50	10	3.06	1.53	0.20	5.00
5<=hr< 6	47	2.13	2.37	0.00	14.20	28	3.64	5.52	0.20	25.00
6<=hr< 7	67	1.84	1.52	0.00	6.40	35	3.21	3.19	0.00	13.40
7-8(ref)	44	1.87	2.66	0.00	13.80	44	3.11	4.12	0.00	20.40
hr> 8	5	1.52	1.35	0.20	3.60	12	1.87	2.03	0.00	7.20
Total N	207					129				

As is evident in the higher risk group, there was a marked increase in lapses when the average duration of cumulative inactivity was less than 5 hours. The analysis of variance supporting this is summarized in Table 7.22.

Table 7.22. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout and risk group) for PVT mean total lapses per trial. The linear trend in the association between PVT lapses and mean values of cumulative duration of inactivity was statistically significant.

Test	F value	df		P value	Pairwise Contrasts
Duration (Overall)	3.17	4	330	0.014	<5 vs. > 8: p=0.009
Risk Group	4.79	1	330	0.029	<5 vs. 7-8: p=0.004
Duration Linear Trend	7.15	1	330	0.008	<5 vs. 6-<7: p=0.002
Duration Quadratic Trend	0.65	1	330	0.422	<5 vs. 5-<6: p=0.016
Risk Group * Duration Category Interaction	0.93	4	326	0.448	

The least squares estimated mean values controlling for risk group were 4.54, 2.82, 2.46, 2.49, and 1.56 lapses per trial, respectively, for subject with <5 hours, 5-<6 hours, 6-<7 hours, 7-8 hours, and >8 hours of sleep on average, respectively. The pooled estimate of the standard deviation was 3.97. Thus, the standardized effect size comparing subjects with a mean duration of <5 hours to those with a mean duration of 7-8 hours was (4.54-2.49)/3.97 or 0.52, a value usually considered to be of substantive magnitude. The linear trend was statistically significant (F=7.15, df=1,330, p=0.008) with no significant non-linear component (p=0.422). There were significant pairwise differences between durations of <5 hours and all of the other duration categories but not among the other categories of mean sleep duration. Thus, we find important

performance lapses increases in subjects with average durations of cumulative inactivity of less than 5.0 hours.

The least squares mean values for mean cumulative duration of inactivity in the main sleep bout category are illustrated Figure 7.12. Again, there was a significant linear trend in PVT lapses per trial (p=0.008), with particularly high values in individuals in whom the duration of this variable was under 5 hours.
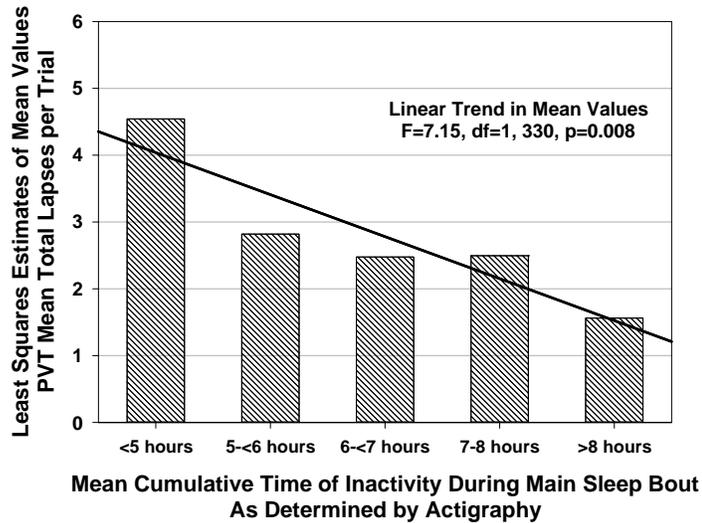
Figure 7.12. Psychomotor Vigilance Task (PVT) least squares estimated mean values for total lapses per trial controlling for risk group across categories of mean cumulative durations of inactivity during the main sleep bout. After controlling for risk group, a clear linear trend emerges (p=0.008) in the number of vigilance lapses. The solid line is from a least squares fit of the mean values.

D. Determinants of PVT Lapses

To assess the variables that determined lapses on the PVT, we used the same multivariable multiple linear regression model as in previous analyses using our two different methods of assessing sleep duration (see Tables 7.23 and 7.24).

Mean PVT lapses significantly differed according to sleep apnea severity controlling for mean cumulative duration of inactivity in the main sleep bout category (Table 7.23, p=0.003). Linear trends were significant for sleep apnea severity (p=0.030) and for at-home duration of major bout of relative inactivity (p=0.0499). The estimated difference (SE) in mean PVT lapses between drivers with severe sleep apnea relative to no sleep apnea was +2.3 (1.1). The estimated difference (SE) in mean PVT lapses between drivers with <6 hr mean at-home duration of major bout of relative inactivity compared to >8 hr was +1.5 (0.6). Thus, the effects of sleep apnea and sleep duration are similar in magnitude. There was also a significant association with health-related quality of life (p=0.024). For each one point worsening on the 6-point scale, the expected number of PVT lapses per trial increased by 0.59. Thus, worse perceptions were associated with worse performance.

Table 7.23. Multiple linear regression model for PVT lapses. Explanatory variables included sleep apnea severity, mean main bout length of relative inactivity category during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	2.9824	0.4827	0.0000	---
Age - 45.5 (yrs)	0.0124	0.0194	0.5225	---
BMI - 29.9 (kg/m^2)	-0.0381	0.0468	0.4157	---
Female gender	0.8801	0.8883	0.3226	---
<u>**Apnea Hypopnea Index**</u>	---	---	---	---
Overall difference in means	---	---	---	0.0034
Linear trend	---	---	---	0.0302
Quadratic trend	---	---	---	0.0031
<u>**AHI Model parameters**</u>	---	---	---	---
>30 vs. <5	2.2879	1.1257	0.0429	---
15-30 vs. <5	-0.6176	1.0014	0.5379	---
5-<15 vs. <5	-1.5254	0.5850	0.0095	---
<u>**Sleep Bout Duration (actigraphy)**</u>	---	---	---	---
Overall difference in means	---	---	---	0.1134
Linear trend	---	---	---	0.0499
Quadratic trend	---	---	---	0.0796
<u>**Bout Duration model parameters**</u>				
<6 vs. >8 hrs.	1.5320	0.8253	0.0643	---
6-<7 vs. >8 hrs.	0.1117	0.6641	0.8665	---
7-8 vs. >8 hrs.	-0.3337	0.5508	0.5450	---
Health related QoL Score	0.5928	0.2606	0.0236	---
<u>**Model Summary**</u>	---	---	---	---
R-square	0.083	---	---	---
Root MSE	3.8811	---	---	---
<u>**Added variables**</u>				**AHI p-value**
Frequent Snoring (>=3/wk)	-0.6803	0.5514	0.2182	0.0009
Current Smoking	0.2828	0.2570	0.2720	0.0092
Alcohol >2 drinks / day	-0.2446	0.3025	0.4193	0.0533

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

Similar results were obtained when at-home sleep was characterized by mean cumulative duration of inactivity in the main sleep bout. Mean PVT lapses significantly differed according to sleep apnea severity (p=0.008). The (adjusted) expected increase in the number of PVT lapses for an individual with AHI≥30 compared to an individual with AHI<5 was 1.87. The simultaneous effect of a linear increase in mean PVT lapses for decreases in mean duration of cumulative inactivity was also statistically significant (p=0.026). There was a significant association with health-related quality of life (p=0.015). For each one point worsening on the 6-

point scale, the expected number of PVT lapses per trial increased by 0.63. Thus, worse perceptions were associated with worse performance.

Table 7.24. Multiple linear regression model for PVT lapses. Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	2.802127	0.429136	0.0000	---
Age – 45.5 years	0.007722	0.019577	0.6935	---
BMI – 29.9 kg/m^2	-0.056074	0.047605	0.2397	---
Female gender	1.262324	0.914724	0.1685	---
<u>**Apnea Hypopnea Index**</u>	**---**	**---**	**---**	---
Overall difference in means	---	---	---	**0.0082**
Linear trend	---	---	---	**0.0707**
Quadratic trend	---	---	---	0.0087
<u>AHI Model parameters</u>	---	---	---	---
>30 vs. <5	1.874608	1.158783	0.1067	---
15-30 vs. <5	-0.517324	1.000085	0.6053	---
5-<15 vs. <5	-1.574417	0.591715	0.0082	---
<u>**Sleep Duration (actigraphy)**</u>	---	---	---	---
Overall difference in means	---	---	---	**0.2665**
Linear trend	---	---	---	**0.0261**
Quadratic trend	---	---	---	0.8201
<u>Sleep Duration model parameters</u>	---	---	---	---
<5 vs. 7-8 hrs.	1.213630	0.758224	0.1104	---
5-<6 vs. 7-8 hrs.	0.413877	0.633405	0.5140	---
6-<7 vs. 7-8 hrs.	0.084525	0.570849	0.8824	---
>8 vs. 7-8 hrs.	-1.189357	0.974433	0.2231	---
Health related QoL Score	0.634310	0.260102	0.0153	---
<u>**Model Summary**</u>	---	---	---	---
R-square	0.080709	---	---	---
Root MSE	3.891818	---	---	---
<u>**Added variables**</u>				**AHI p-value**
Frequent Snoring (>=3/wk)	-0.712564	0.545314	0.1922	0.0051
Current Smoking	0.282288	0.259096	0.2767	0.0143
Alcohol >2 drinks / day	-0.242781	0.301727	0.4216	0.0570

E. <u>Exploratory Model: Risk of Impaired Function Defined by PVT Lapses per Trial >3</u>

We next looked at the relationship between impairment as defined by having more than 3 lapses per trial and apnea/hypopnea index and our different methods of estimating sleep duration (see Figures 7.13 and 7.14). Powell et al [1999] provides mean PVT lapse data for 35 male subjects before and after consumption of alcohol. Mean breath alcohol (SD) at the various trials at different levels of alcohol were 0 (0), 0.059 (0.0), 0.08 (0.01), and 0.084 (0.01) grams/200 liters, respectively. Values above 0.04 g/200 liters make it illegal to drive in California. The number of PVT lapses (SD) for these trials were 0.4 (0.6), 0.49 (0.8), 0.91 (1.2), and 1.2 (1.6), respectively. Thus, our criterion value for defining impairment (> 3 lapses per trial) is well above the mean value obtained in the Powell et al study when alcohol consumption was over twice the legal limit.
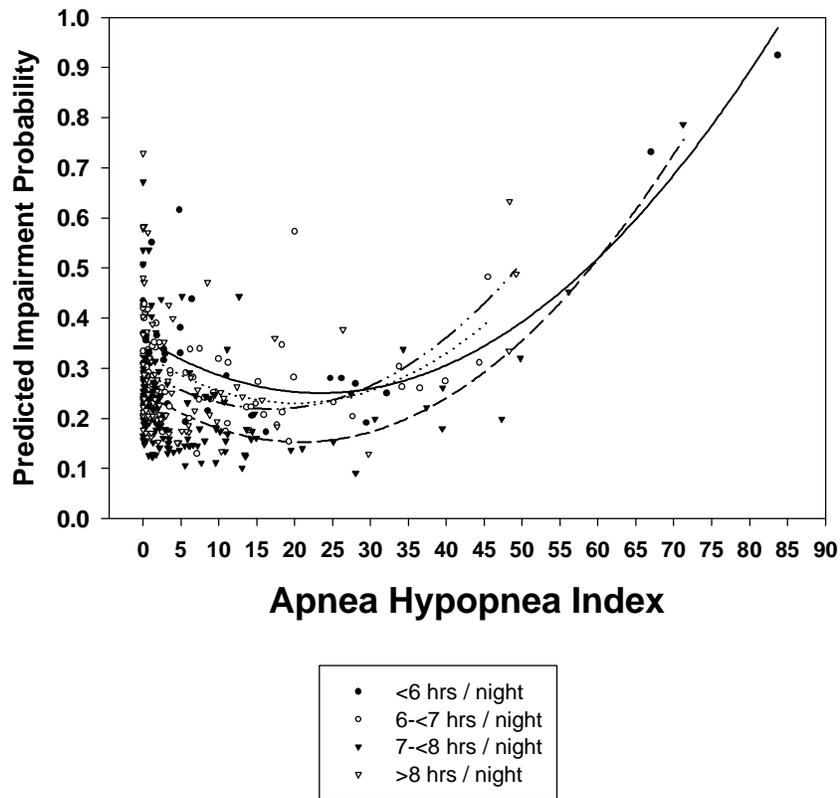
Figure 7.13. Predicted probability of PVT lapses per trial >3 by AHI and mean main bout length of relative inactivity in different categories. Additional variables included in this model were age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 29.2%.
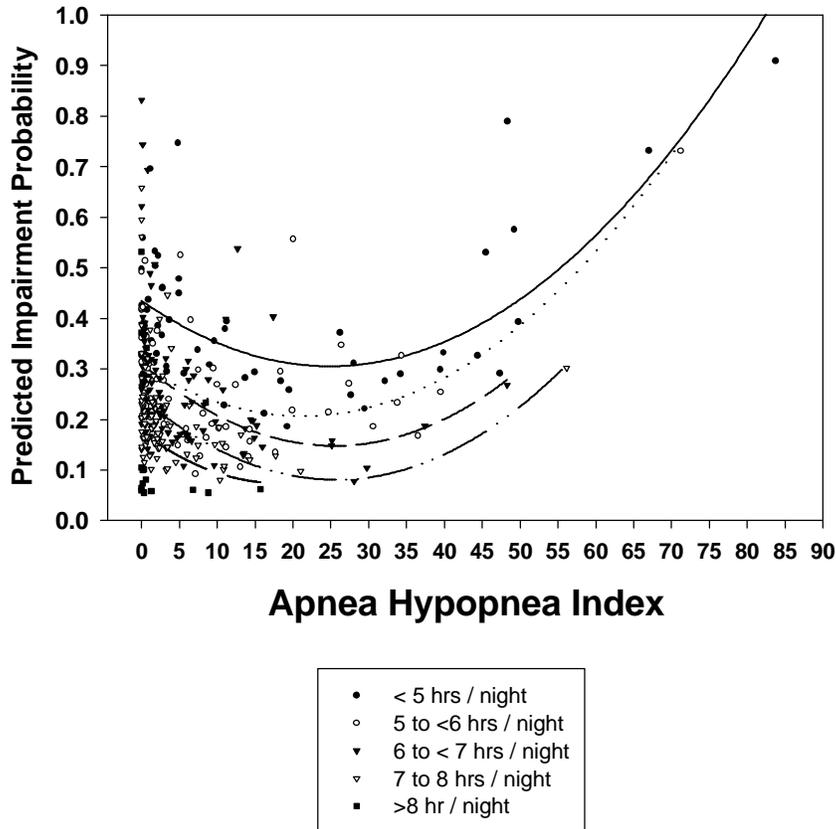
Figure 7.14.   Predicted probability of PVT lapses per trial >3 by AHI and mean cumulative duration of inactivity in the main sleep bout in different categories of hours/night. Additional variables in the model include age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 27.8%.

For both types of models we see that duration of sleep has some effect on predicted prevalence of impairment. This is not, surprisingly, more evident when the variable defining sleep was cumulative duration of relative inactivity (Figure 7.14) as compared to duration of main bout of relative inactivity (Figure 7.13). For both models we see a strong effect of sleep apnea severity with clear increases in predicted prevalence of impairment at apnea/hypopnea indices above 30 episodes/hour.

F. Analyses of PVT Mean Total Vigilance Lapses per Trial after Transformation

Sample distributions of total lapses during PVT trials typically have substantial right skewness, that is, they often possess some large extreme values. This is because many populations of interest typically consist of many individuals with negligible, minor, or mild functional deficits (reflected by relatively few vigilance lapses) plus a smaller number of individuals with clinically significant functional declines (reflected by a relatively large number of lapses). The number of vigilance lapses is a random variable with a (subject-specific) expected value proportional to test time. Moreover, by design, two lapses cannot occur simultaneously. These are characteristics shared by the Poisson distribution, a distribution in

which the standard deviation is proportion to its mean value. Therefore, it is not surprising that a variance stabilizing distribution for the Poisson distribution is appropriate to use in transforming raw lapses. When comparing among populations that vary according to the degree of functional impairment or vary according to the percentages of individuals with functional impairments, mean number of lapses, and so standard deviation values differ, violating the constant variance assumption used in parametric statistical models including analysis of variance. To address this problem a standard transformation has been developed. The precise form of this is given in the footnote below.[4] The descriptive statistics and analysis of variance results based on transformed total lapses for apnea severity are presented in Tables 7.25 and 7.26, respectively. The transformation was moderately successful in reducing variance heterogeneity. For example, in the higher risk group, prior to transformation, the SD values were 3.43, 1.40, 1.83, and 8.06, respectively, for AHI<5, 5 to <15, 15 to <30, and ≥30 events per hour. After transformation these values became 1.45, 0.97, 1.19, and 2.40 for these AHI categories, respectively. Thus, prior to transformation the ratio of the maximum to minimum SD (variance) was 5.75 (33.1). The transformation reduced this ratio by more than 100% (400%) to 2.47 (6.12).

Table 7.25.   PVT mean transformed total vigilance lapses per trial by severity of sleep apnea for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	2.88	1.45	1.00	9.57	137	3.29	1.78	1.00	10.22
5 - <15	69	2.41	0.97	1.00	5.27	16	2.46	0.75	1.28	3.70
15 - <30	29	2.87	1.19	1.00	4.84	2	4.62	0.25	4.44	4.80
>= 30	28	3.71	2.40	1.00	12.61					
Total N	245					155				

---

[4] The specific transformation on the number of lapses per PVT trial (Y) used is $\sqrt{Y} + \sqrt{(Y+1)}$. For example, 0, 1, 2, 4, and 8 lapses are transformed into 1, 2.41, 3.24, 4.23, and 5.82, respectively.

Table 7.26. Two-way ANOVA (apnea severity and risk group) for PVT mean transformed total vigilance lapses per trial. Controlling for risk group, statistically significant increases in total vigilance lapses were associated with severe sleep apnea.

Test	F value		df	P value	Pair-wise Contrasts
AHI (Overall)	5.97	3	395	0.001	≥30 vs. <5: p=0.012
Risk Group	4.74	1	395	0.030	≥30 vs. 5-<15: p<0.0001
AHI Linear Trend	8.83	1	395	0.003	≥30 vs.15-<30: p=0.065
AHI Quadratic Trend	8.26	1	395	0.004	15-<30 vs. 5-<15:p=060
					5-<15 vs. <5: p=0.007
Risk Group * AHI Category Interaction	1.03	2	393	0.357	

Comparisons between Tables 7.26 and 7.18 show that the statistical conclusions based on raw lapses remained intact after transformation. The least squares estimated mean values (accounting for risk group) for transformed lapses per trial were 3.09, 2.53, 3.15, and 3.90, respectively, for subjects with AHI values <5, 5 to <15, 15 to <30, and ≥30 episodes per hour (F=5.97, df=1,395, p=0.001). Both linear and non-linear trends remained statistically significant. For subjects with severe sleep apnea, there was still a significant increase in performance lapses after transformation.

The descriptive statistics and results of analysis of variance to assess the role of mean sleep duration are presented in Tables 7.27 and 7.28, respectively, based on mean main bout length of relative inactivity category and in Tables 7.29 and 7.30 based on mean cumulative duration of inactivity in the main sleep bout. Again, statistical conclusions after transformation were similar to those based on untransformed lapses. There was at most a weak association between PVT transformed lapses when duration was expressed by mean main bout length of relative inactivity category. In contrast, the association between PVT lapses and duration was statistically significant and strong when duration was expressed as mean cumulative duration of inactivity in the main sleep bout. High levels of performance lapses were found in those individuals with mean cumulative durations of inactivity in the main sleep bout of less than 5 hours.

Table 7.27. PVT mean transformed total vigilance lapses per trial by mean main bout length of relative of different durations category during the prior week as determined by actigraphy for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	3.03	1.84	1.00	9.57	13	4.01	2.57	1.28	10.17
6<=hr< 7	53	3.05	1.49	1.28	9.32	23	3.16	1.41	1.00	6.46
7<=hr<= 8	87	2.68	1.52	1.00	12.61	61	2.96	1.35	1.00	7.30
hr>8	47	2.78	1.39	1.00	7.47	32	3.29	1.84	1.00	9.05
Total N	207					129				

Table 7.28. Two-way ANOVA (mean main bout length of relative inactivity category and risk group) for PVT mean transformed total vigilance lapses per trial. Controlling for risk group, the linear trend in mean values did reach statistical significance (p=0.102).

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	1.81	3	331	0.145	
Risk Group	4.50	1	331	0.035	
Duration Linear Trend	2.68	1	331	0.102	
Duration Quadratic Trend	1.58	1	331	0.209	
Risk Group * Duration Category Interaction	0.63	3	328	0.598	

Table 7.29. PVT mean transformed total vigilance lapses per trial by mean cumulative duration of inactivity in the main sleep bout during the prior week as determined by actigraphy for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	44	3.53	2.22	1.00	12.61	10	3.37	1.03	1.28	4.34
5<=hr< 6	47	2.78	1.19	1.00	7.33	28	3.33	1.96	1.28	10.17
6<=hr< 7	67	2.64	1.05	1.00	5.11	35	3.32	1.53	1.00	7.30
7-8(ref)	44	2.50	1.44	1.00	7.47	44	3.10	1.75	1.00	9.05
hr> 8	5	2.55	1.05	1.28	3.95	12	2.56	1.26	1.00	5.28
Total N	207					129				

Table 7.30. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout category and risk group) for PVT mean transformed total vigilance lapses per trial. Controlling for risk group, a statistically significant linear trend reflecting increase in total vigilance lapses was associated with shorter mean durations of cumulative inactivity.

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	3.18	4	330	0.014	<5 vs. >8: p=0.007
Risk Group	7.87	1	330	0.005	<5 vs. 7-8: p=0.002
Duration Linear Trend	8.38	1	330	0.004	<5 vs., 6-<7: p=0.007
Duration Quadratic Trend	0.25	1	330	0.619	<5 vs. 5-<6: p=0.03
Risk Group * Duration Category Interaction	0.56	4	326	0.690	

7.5 Analyses of Other Variables Measured by PVT

A. Duration of Lapse Domain

As noted earlier, the PVT provides a number of other parameters besides the key variables of median response times and the number of response times that exceeded 500 ms (i.e., the number of lapses). These other PVT parameters reflect specific aspects of functional impairments reflecting differential effects of sleepiness and other challenges on functional performance. We describe the results in this section for completeness. The *duration of the lapse domain* refers to shifts in lapse duration calculated from the 10% slowest RTs, a metric that reflects vigilance response slowing. The PVT device provides a transformed parameter for duration of the lapse domain. Specifically, the parameter analyzed is defined as 1000 times the reciprocal of the RT. For example, an RT equal to 250 is transformed to 4.0. On this transformed scale, smaller values reflect worse performance. Tables 7.31 and 7.32 provide

descriptive statistics and the analysis of variance for the effect of the severity of sleep apnea on performance as measured in this domain. As before, decrements in performance were particularly found in those with severe sleep apnea.

Table 7.31. PVT mean of the 10% slowest response times by sleep apnea severity as determined by PSG for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
AHI	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	2.47	0.52	0.57	3.86	137	2.33	0.58	0.51	3.90
5 - <15	69	2.57	0.36	1.71	3.44	16	2.52	0.28	2.02	3.03
15 - <30	29	2.42	0.38	1.78	3.06	2	1.80	0.09	1.74	1.86
>= 30	28	2.18	0.67	0.14	3.28					
Total N	245					155				

Table 7.32. Two-way ANOVA (apnea severity and risk group) for PVT mean of the slowest 10% response times. Controlling for risk group, statistically significant decreases in the mean of the slowest 10% response times were associated with increased apnea severity.

Test	F value	df		P value	Pair-wise contrasts
AHI (Overall)	4.64	3	395	0.003	≥30 vs. <5: p=0.009
Risk Group	5.62	1	395	0.018	≥30 vs. 5-<15: p=0.0004
AHI Linear Trend	9.44	1	395	0.002	≥30 vs. 15-<30: p=0.125
AHI Quadratic Trend	4.86	1	395	0.028	15-<30 5-<15: p=0.065
					5-<15 vs. <5: p=0.067
Risk Group * AHI Category Interaction	1.00	2	393	0.370	

The least squares estimated mean values (accounting for risk group) were 2.40, 2.52, 2.32, and 2.11 (1000/msec) for subjects with AHI values <5, 5-<15, 15-<30, and ≥30 episodes per hour (F=4.64, df=1,395, p=0.003). Least squares mean values for the higher and lower risk groups (controlling for apnea severity) were 2.41 and 2.67 (1000/msec), respectively (F=5.62, df=1, 395, p=0.018). Both linear and non-linear trends were significant with a precipitous drop for subjects with severe sleep apnea.

We next analyzed the effect of average sleep duration, again employing the two measures we used to estimate sleep duration. The descriptive statistics and analysis of variance results for the effect of mean main bout length of relative inactivity category on the duration of the lapse domain are presented in Tables 7.33 and 7.34, respectively. The descriptive statistics and analysis of variance results for the effect of mean cumulative duration of inactivity in the main sleep bout category on the duration of the lapse domain are presented in Tables 7.35 and 7.36, respectively. Results differed substantially for these two metrics of sleep duration. No

statistically significant association was observed when duration was defined according to mean main bout length of relative inactivity category. In contrast, highly statistically significant results were observed for mean cumulative duration of inactivity in the main sleep bout category.

Table 7.33. PVT mean duration of the lapse domain by mean main bout length of relative inactivity category as determined by actigraphy for CDL holders in the higher and lower risk groups. The duration of the lapse domain is a transformation of the 10% slowest response times obtained by multiplying 1000 times the reciprocal response time.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	2.46	0.58	0.57	3.06	13	2.21	0.78	0.51	3.43
6<=hr< 7	53	2.39	0.46	0.57	3.03	23	2.36	0.62	1.25	3.90
7<=hr<= 8	87	2.51	0.49	0.14	3.81	61	2.38	0.42	1.21	3.27
hr>8	47	2.48	0.53	1.02	3.86	32	2.33	0.61	0.61	3.56
Total N	207					129				

Table 7.34. Two-way ANOVA (mean main bout length of relative inactivity category and risk group) for PVT mean of the duration of the lapse domain. There was no significant association between mean duration and PVT slowing in the lapse domain for this duration metric.

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	0.61	3	331	0.606	
Risk Group	4.44	1	331	0.036	
Duration Linear Trend	0.68	1	331	0.409	
Duration Quadratic Trend	0.12	1	331	0.729	
Risk Group * Duration Category Interaction	0.30	3	328	0.826	

Table 7.35. PVT mean duration of the lapse domain by mean cumulative duration of inactivity in the main sleep bout category. The duration of the lapse domain is a transformation of the 10% slowest response times obtained by multiplying 1000 times the reciprocal response time.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	44	2.28	0.63	0.14	3.16	10	2.34	0.54	1.85	3.43
5<=hr< 6	47	2.47	0.41	1.43	3.13	28	2.32	0.61	0.51	3.90
6<=hr< 7	67	2.47	0.38	1.59	3.42	35	2.26	0.50	1.21	3.27
7-8(ref)	44	2.62	0.56	1.02	3.86	44	2.37	0.55	0.61	3.29
hr> 8	5	2.64	0.39	2.13	3.06	12	2.59	0.51	1.85	3.56
Total N	207					129				

Table 7.36. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout category and risk group) for PVT mean of the duration of the lapse domain. Controlling for risk group, there were statistically significant decreases in the mean response values associated with decreases in mean sleep duration as determined by this duration metric.

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	2.92	4	330	0.021	<5 vs. >8: p=0.006
Risk Group	8.16	1	330	0.005	<5 vs. 7-8: p=0.005
Duration Linear Trend	8.84	1	330	0.003	<5 vs. 6-<7: p=0.117
Duration Quadratic Trend	0.14	1	330	0.713	<5 vs. 5-<6: p=0.084
					5-<6 vs. >8: p=0.081
Risk Group * Duration Category Interaction	0.64	4	326	0.634	6-<7 vs. >8: p=0.051

Least squares mean values for subjects with <5, 5 to <6, 6 to <7, 7 to 8, and >8 hours sleep on average were 2.3, 2.4, 2.4, 2.5, and 2.6 (1000/msec), respectively. As with other variables, highly significant differences were found for subjects with an average sleep duration of less than 5 hours of mean cumulative duration of inactivity in the main sleep bout.

B. Analysis of PVT Mean 10% Fastest Reaction Times (*Optimum Response Times*)

Another variable that can be extrapolated from the PVT test is the optimum response time. Optimum response times refer to the average of the 10% fastest RTs per trial, and reflect the very best performance an operator is capable of producing. Tables 7.37 and 7.38 provide descriptive statistics and the analysis of variance for the association between the mean 10% fastest RTs per trial and sleep apnea severity. Least squares mean values for subjects with AHI values >=30, 15-<30, 5-<15, and <5 events per hours (controlling for risk group) were 213.6, 211.3, 202.2, and 203.9 msec, respectively (F=3.60, df=1,395, p=0.014). Thus, decreases in the

optimum response times were already present for subjects in the moderate apnea severity category.

Table 7.37. PVT mean of the 10% fastest response times by sleep apnea severity as determined by PSG for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
AHI	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	203.47	20.58	159.54	267.48	137	204.47	19.23	140.45	264.22
5 - <15	69	200.70	16.42	161.70	245.25	16	204.75	15.22	179.05	230.78
15 - <30	29	208.23	17.80	177.29	247.73	2	240.36	17.14	228.23	252.48
>= 30	28	212.45	22.61	179.51	265.34					
Total N	245					155				

Table 7.38. Two-way ANOVA (apnea severity and risk group) for PVT mean of the fastest 10% response times. Controlling for risk group, statistically significant decreases in the mean of the fastest 10% response times were associated with increased apnea severity.

Test	F value	df	P value	Pair-wise contrasts
AHI (Overall)	3.60	3 395	0.014	≥30 vs. <5: p=0.017
Risk Group	1.07	1 395	0.301	≥30 vs. 5-<15: p=0.007
AHI Linear Trend	8.88	1 395	0.003	≥30 vs. 15-<30: p=0.649
AHI Quadratic Trend	0.52	1 395	0.470	15-<30 vs. 5-<15: p=0.025
				15-<15 vs. <5: p=0.067
Risk Group * AHI Category Interaction	2.45	2 393	0.087	

Tables 7.39 and 7.40 provide descriptive statistics and the analysis of variance for the association between the mean 10% fastest RTs per trial and mean main bout length of relative inactivity category. Similarly, Tables 7.41 and 7.42 provide descriptive statistics and the analysis of variance for the association between the mean 10% fastest RTs per trial and mean cumulative duration of inactivity in the main sleep bout category. Again there was no evidence of an association between the PVT optimum responses and mean main bout length of relative inactivity category. There was also no significant linear trend observed for this variable extracted from the PVT among different durations of mean cumulative duration of inactivity in the main sleep bout (p=0.117). The pattern of the mean values displayed minor changes in both risk groups for shorter durations of cumulative durations of inactivity. Thus, this PVT variable, i.e., optimum response times, was not as affected by sleep duration as some the other variables derived from the PVT test.

Table 7.39. PVT mean of the 10% fastest response times by mean main bout length of relative inactivity category for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	199.14	21.15	171.13	267.29	13	200.43	20.07	176.03	239.35
6<=hr< 7	53	204.74	21.27	174.30	267.48	23	202.98	22.33	165.10	261.59
7<=hr<= 8	87	205.22	19.25	161.70	265.17	61	206.36	16.76	164.56	264.22
hr>8	47	200.71	17.86	159.54	245.25	32	202.50	17.79	160.50	238.40
Total N	207					129				

Table 7.40. Two-way ANOVA (mean main bout length of relative inactivity category and risk group) for PVT mean of the fastest 10% response times.

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	1.41	3	331	0.238	
Risk Group	0.11	1	331	0.738	
Duration Linear Trend	0.31	1	331	0.580	
Duration Quadratic Trend	3.42	1	331	0.065	
Risk Group * Duration Category Interaction	0.12	3	328	0.950	

Table 7.41. PVT mean of the 10% fastest response times by mean cumulative duration of inactivity in the main sleep bout category for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	44	205.8	24.0	161.7	267.3	10	204.5	27.8	176.0	261.6
5<=hr< 6	47	203.2	18.6	174.3	267.5	27	204.6	18.6	165.1	239.4
6<=hr< 7	67	206.2	18.4	179.5	260.3	34	204.2	17.7	164.6	250.2
7-8(ref)	43	198.7	18.1	159.5	245.2	44	205.9	17.5	160.5	264.2
hr> 8	5	197.1	5.1	191.1	203.7	12	198.8	16.4	180.6	229.8
Total N	206					127				

Table 7.42. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout category and risk group) for PVT mean of the fastest 10% response times.

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	0.91	4	327	0.457	
Risk Group	0.64	1	327	0.424	
Duration Linear Trend	2.46	1	327	0.117	
Duration Quadratic Trend	0.69	1	327	0.406	
Risk Group * Duration Category Interaction	0.71	4	323	0.584	

C. Analysis of PVT Decrement Function (Fatigability Function)

The *fatigability function* refers to the vigilance decrement function or the extent to which subjects maintained performance across time on task. Each trial provides a value representing the linear regression slope of the reciprocal of the mean response over a one-minute interval as a function of minute (1 to 10). As a consequence of the transformation, more negative slopes reflect greater inability to sustain attention. Tables 7.43 and 7.44 provide descriptive statistics and the analysis of variance for the association between these slopes and sleep apnea severity. Least squares estimated expected values for subjects with AHI values >=30, 15-<30, 5-<15, and <5 events per hours (controlling for risk group) were –0.0172, -0.0033, -0.0075, and –0.0138. The quadratic trend in these mean values was statistically significant (F=5.62, df=1,392, p=0.018).   Thus, for subjects with AHI≥30 events per hour relative to subjects with AHI<5, ≥5 but <15, and ≥15 but less the 30 events per hour, the average rate in performance decrement over the 10 minute PVT testing bout was 25% larger, 229% larger, 521% larger, respectively. It is unclear as to why the largest difference was between subjects with severe sleep apnea compared to moderate sleep apnea.

Table 7.43. Mean PVT slopes by sleep apnea severity as determined by PSG for CDL holders in the higher and lower risk groups.

AHI	**Higher Risk**					**Lower Risk**				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	118	-0.016	0.029	-0.096	0.061	135	-0.011	0.030	-0.110	0.068
5 - <15	69	-0.010	0.027	-0.086	0.047	16	-0.006	0.019	-0.043	0.026
15 - <30	29	-0.006	0.021	-0.055	0.054	2	-0.001	0.045	-0.033	0.031
>= 30	28	-0.020	0.038	-0.149	0.038					
Total N	244					153				

Table 7.44. Two-way ANOVA (apnea severity and risk group) for mean PVT slopes.

Test	F value	df	P value	Pair-wise contrasts
AHI (Overall)	2.07	3 392	0.104	>=30 vs. 15-<30: p=0.069
Risk Group	2.24	1 392	0.136	>=30 vs. <5: p=0.069
AHI Linear Trend	0.09	1 392	0.764	
AHI Quadratic Trend	5.62	1 392	0.018	
Risk Group * AHI Category Interaction	0.00	2 390	0.996	

The descriptive statistics and analysis of variance results for the effect of mean main bout length of relative inactivity category on PVT slopes are presented in Tables 7.45 and 7.42, respectively. No significant association was observed.

Table 7.45. PVT mean slopes by mean main bout length of relative inactivity category for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
Duration	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	-0.016	0.037	-0.096	0.037	12	-0.018	0.036	-0.110	0.024
6<=hr< 7	53	-0.011	0.029	-0.094	0.054	22	-0.008	0.028	-0.079	0.040
7<=hr<= 8	86	-0.016	0.028	-0.149	0.039	61	-0.009	0.028	-0.073	0.068
hr>8	47	-0.007	0.026	-0.055	0.061	32	-0.018	0.031	-0.101	0.051
Total N	206					127				

Table 7.46. Two-way ANOVA (mean main bout length of relative inactivity category and risk group) for mean PVT slope.

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	0.46	3	328	0.713	
Risk Group	0.12	1	328	0.732	
Duration Linear Trend	0.47	1	328	0.491	
Duration Quadratic Trend	0.37	1	328	0.542	
Risk Group * Duration Category Interaction	1.83	3	325	0.141	

Tables 7.47 and 7.48 provide descriptive statistics and the analysis of variance for the association between the mean PVT slopes and mean sleep duration as estimated by duration of cumulative period of inactivity. In contrast, to the data presented for duration of main bout of relative inactivity in Tables 7.45 and 7.46, both the linear trend and overall differences, were statistically significant. The least squares estimated mean values for subjects with <5, 5 to <6, 6 to <7, 7 to 8, and >8 hours mean cumulative duration of inactivity in the main sleep bout category were -0.0247, -0.0056, -0.0136, -0.0112 and -0.0049, respectively. Thus, for subjects with a mean duration of <5 hours per night relative to subjects with mean durations of ≥5 hours but less than 6, ≥6 hours but less than 7, 7 to 8 hours inclusive, and >8 hours, the average rate in performance decrement over the 10 minute PVT testing bout was 441% larger, 182% larger, 220% larger, and 504% larger, respectively. As with apnea severity, there was a very large difference between the most severely pathological category compared to the next most severely pathological category.

Table 7.47. PVT mean slope by mean cumulative duration of inactivity in the main sleep bout category for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
Duration	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	44	-0.025	0.038	-0.149	0.054	10	-0.022	0.038	-0.110	0.021
5<=hr< 6	47	-0.003	0.025	-0.062	0.039	27	-0.010	0.029	-0.079	0.060
6<=hr< 7	67	-0.016	0.022	-0.086	0.030	34	-0.009	0.023	-0.068	0.040
7-8(ref)	43	-0.009	0.027	-0.067	0.061	44	-0.013	0.030	-0.081	0.068
hr> 8	5	0.011	0.027	-0.026	0.047	12	-0.012	0.038	-0.101	0.051
Total N	206					127				

Table 7.48. Two-way ANOVA (mean duration of cumulative inactivity during major sleep bout and risk group) for mean PVT slope.

Test	F value	df	P value	Pair-wise contrasts
Duration (Overall)	3.81	4 327	0.005	<5 vs. >8: p=0.015
Risk Group	0.17	1 327	0.678	<5 vs. 7-8: p=0.008
Duration Linear Trend	4.01	1 327	0.046	<5 vs. 6-<7: p=0.022
Duration Quadratic Trend	0.75	1 327	0.386	<5 vs. 5-<6: p=0.0002
				5-<6 vs. 6-<7: p=0.067
Risk Group * Duration Category Interaction	1.25	4 323	0.289	

7.6. <u>Divided Attention Driving Task (DADT)</u>

A. <u>Description of Test</u>

The Divided Attention Driving Task (DADT) was developed to simulate the cognitive load of the driving task [Moskowitz and Burns, 1977] and has been used in studies of the impairment of driving ability by alcohol [Moskowitz and Burns, 1977] and by medication [Smiley, 1987]. It also has recently been used for studies in sleep apnea patients and impairments of the same magnitude are described as those found in people whose blood alcohol concentration is over the legal limit [George et al, 1996]. The DADT uses a driving-like paradigm in which subjects must keep a cursor within the middle of a randomly moving target using a steering wheel device (lane tracking), while at the same time identifying and responding to numbers that appear at irregular intervals at the corners of the screen. Our version of the test took 30 minutes. Average deviation from the desired center point is computed for each 2-minute period. The major outcome measure is the mean two-minute sum of the absolute values of the tracking errors over the period of study. In addition, we computed a *tracking decrement function* using multiple linear regression. For this parameter, we analyzed the absolute value of the deviations of the cursor from the center for each of the 15 two-minute blocks of data across the 30-minute period. An extensive graphical analysis revealed that with a 30 minute testing bout, a logarithmic transformation produced a linear relationship between total tracking error per two-minutes and time on task. Following this transformation, a two-stage random effects analysis was used [Feldman, 1988].

B. <u>Sum of the Absolute Values of the Tracking Errors</u>

Tables 7.49 and 7.50 provide descriptive statistics and the analysis of variance for the association between sleep apnea severity and mean tracking error. Least squares estimated expected values for subjects with AHI values >=30, 15-<30, 5-<15, and <5 events per hours (controlling for risk group) were 380.0, 283.4, 227.8, and 251.6 cm. We see that mean absolute tracking error was significantly associated with severity of sleep apnea (F=5.13, df=1,372, p=0.002). Both linear and quadratic trends were significant. While subjects in the AHI category 15-<30 events per hour appeared to have elevated mean tracking errors compared to those with mild or no sleep apnea, these differences were not statistically significant. Thus, as for other variables, the major effect of sleep apnea is found in those subjects with severe sleep apnea. This is shown graphically in Figure 7.15.

Table 7.49. Mean DADT absolute lane deviation tracking errors (cm) by sleep apnea severity as determined by PSG for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	109	244.2	156.4	70.4	1324.0	140	258.4	177.3	67.4	1111.2
5 - <15	60	229.6	132.5	73.3	899.2	16	207.9	51.4	102.2	288.3
15 - <30	28	274.0	107.0	134.6	628.0	2	352.0	249.2	175.8	528.2
>= 30	22	375.1	263.2	112.0	1167.1					
Total N	219					158				

Table 7.50. Two-way ANOVA (apnea severity and risk group) for DADT sum of the absolute values of the tracking errors. Highly significant differences between tracking error in subjects with severe sleep apnea and the other categories of severity are found.

Test	F value	df		P value	Pair-wise contrasts
AHI (Overall)	5.13	3	372	0.002	≥30 vs. <5: p=0.0008
Risk Group	0.27	1	372	0.602	≥30 vs. 5-<15: p=0.0002
AHI Linear Trend	13.57	1	372	0.000	≥30 vs. 15-<30: p=0.036
AHI Quadratic Trend	5.58	1	372	0.019	
Risk Group * AHI Category Interaction	0.42	2	370	0.658	
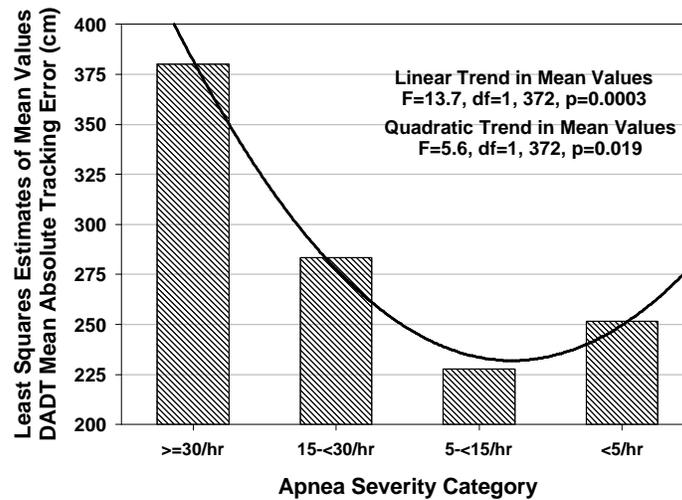
Figure 7.15. Divided Attention Driving Task (DADT) least squares estimated mean absolute tracking error (cm) controlling for risk group. After controlling for risk group, significant linear trend emerges (p=0.0003) and quadratic trends (p=0.019) emerge. The solid line is from a quadratic least squares fit of the mean values. The large increase in tracking error in subjects with an AHI≥30 episodes/hour is seen.

Tables 7.51 and 7.52 provide descriptive statistics and the analysis of variance for the association between the sum of the absolute tracking errors and mean main bout length of relative inactivity category. Tables 7.53 and 7.54 provide the analogous results using mean cumulative duration of inactivity in the main sleep bout category. No association between the sum of the absolute tracking errors and mean main bout length of relative inactivity category was observed. In contrast, reduced mean cumulative durations of inactivity in the main sleep were associated with decreases in lane tracking ability (Linear trend F=4.24, df=1,314, p=0.040).

Table 7.51. DADT mean tracking error by mean sleep duration as determined by mean main bout length of relative inactivity of different durations for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	17	289.7	179.1	129.9	913.4	13	231.7	208.9	81.7	905.8
6<=hr< 7	47	251.4	154.9	70.4	899.2	24	233.2	93.0	67.4	475.6
7<=hr<= 8	80	247.5	170.0	85.0	1324.0	63	246.9	153.0	76.3	841.6
hr>8	44	243.7	117.6	106.7	628.0	32	263.7	176.6	76.7	1111.2
Total N	188					132				

Table 7.52. Two-way ANOVA (mean main bout length of relative inactivity category) for mean sum of the absolute response errors.

Test	F value	df		P value	Pair-wise t-test p<0.05
Duration (Overall)	0.13	3	315	0.942	
Risk Group	0.07	1	315	0.788	
Duration Linear Trend	0.12	1	315	0.735	
Duration Quadratic Trend	0.37	1	315	0.546	
Risk Group * Duration Category Interaction	0.49	3	312	0.691	

Table 7.53. DADT mean tracking error by mean cumulative duration of inactivity in the main sleep bout in different categories as determined by actigraphy for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
Duration	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	37	318.6	194.2	96.9	913.4	10	269.5	131.4	128.2	519.8
5<=hr< 6	43	242.9	128.2	70.4	877.2	28	246.7	188.6	67.4	905.8
6<=hr< 7	65	244.8	173.5	85.0	1324.0	37	254.8	116.3	77.5	744.2
7-8(ref)	38	213.8	88.7	93.3	480.0	45	238.4	183.8	76.3	1111.2
hr> 8	5	198.6	60.9	113.5	260.9	12	236.5	72.0	139.8	368.1
Total N	188					132				

Table 7.54. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout category and risk group) for mean sum of the absolute response errors.

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	2.29	4	314	0.060	<5 vs. >8: p=0.053
Risk Group	0.18	1	314	0.672	<5 vs. 7-8: p=0.004
Duration Linear Trend	4.24	1	314	0.040	<5 vs. 6-<7: p=0.027
Duration Quadratic Trend	1.03	1	314	0.311	<5 vs. 5-<6: p=0.026
Risk Group * Duration Category Interaction	0.36	4	310	0.834	

Least squares estimates of the mean tracking errors were 310.3, 245.2, 249.5, 226.8, and 223.8 cm, respectively, for subjects with <5, 5-<6, 6-<7, 7-8, and >8 hours of mean cumulative duration of inactivity in the main sleep bout during the week prior to testing. Pairwise contrasts

revealed significant differences between subjects in the <5 hours category compared to subjects in each of the other categories. Differences among the other categories were not statistically significant.

C. Determinants of DADT Mean Absolute Lane Deviations

We used, as in previous analyses, multiple linear regression to determine the effect of a number of variables on absolute value of lane tracking error. The results of our analyses are shown in Table 7.55 using mean main bout length of relative inactivity category and Table 7.56 using mean cumulative duration of inactivity in the main sleep bout.

For models including duration of major bout of relative inactivity (Table 7.55), apnea severity remained an independent predictor of tracking errors (p=0.045) with both significant linear (p=0.037) and quadratic trends (p=0.033). Increases in age (p=0.022) and female gender (p=0.011) were also significantly associated with increased tracking errors. However, mean main bout length of relative inactivity category was not significant overall (p=0.992) or with regard to a linear trend (p=0.791).

Table 7.55. Multiple linear regression model for DADT mean absolute tracking error. Explanatory variables included sleep apnea severity, mean main bout length of relative inactivity category during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	243.5812	18.4642	0.0000	---
Age - 45.5 (yrs)	1.7055	0.7423	0.0222	---
BMI - 29.9 (kg/m^2)	0.8141	1.8636	0.6625	---
Female gender	85.8656	33.4534	0.0107	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.0450
Linear trend	---	---	---	0.0245
Quadratic trend	---	---	---	0.0328
AHI Model parameters	---	---	---	---
>30 vs. <5	101.2869	48.3853	0.0371	---
15-30 vs. <5	4.6031	37.9810	0.9036	---
5-<15 vs. <5	-34.7507	22.4789	0.1231	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.9917
Linear trend	---	---	---	0.7913
Quadratic trend	---	---	---	0.8242
Bout Duration model parameters				
<6 vs. >8 hrs.	7.9818	31.8406	0.8022	---
6-<7 vs. >8 hrs.	0.8471	25.2536	0.9733	---
7-8 vs. >8 hrs.	-1.3700	20.8810	0.9477	---
Health related QoL Score	17.5170	10.0009	0.0808	---
Model Summary	---	---	---	---
R-square	0.080	---	---	---
Root MSE	145.6144	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	26.5175	21.8384	0.2255	0.0488
Current Smoking	0.4346	9.8533	0.9648	0.1690
Alcohol >2 drinks / day	-17.6311	10.8065	0.1038	0.1944

Notes: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

When we incorporated the duration of cumulative inactivity (Table 7.56), the linear trend in mean DADT tracking errors among apnea severity categories was marginally statistically significant (p=0.078) and the quadratic trend was statistically significant (p=0.0495). The linear trend in tracking error also appeared marginally statistically significant for mean cumulative duration of inactivity in the main sleep bout category (p=0.054). Tracking error significantly increased as age increased (p=0.037) and was larger for female subjects (p=0.003). As in previous analyses, we found that the magnitude of the decrement in lane tracking ability

produced by severe sleep apnea, as compared to no apnea (81.29), was of similar magnitude to that produced by short mean cumulative duration of inactivity in the main sleep bout, i.e., less than 5 hours compared to 7-8 hours (51.79).

Table 7.56. Multiple linear regression model for DADT mean absolute tracking error. Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout during prior week as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	232.821814	16.369018	0.0000	---
Age - 45.5 years	1.554120	0.742419	0.0371	---
BMI - 29.9 kg/m^2	0.276109	1.874999	0.8830	---
Female gender	101.180174	34.135246	0.0033	---
<u>Apnea Hypopnea Index</u>	**---**	**---**	**---**	---
Overall difference in means	---	---	---	**0.1088**
Linear trend	---	---	---	**0.0782**
Quadratic trend	---	---	---	0.0495
<u>AHI Model parameters</u>	---	---	---	---
>30 vs. <5	81.293258	49.242958	0.0997	---
15-30 vs. <5	-4.453259	37.642050	0.9059	---
5-<15 vs. <5	-34.759646	22.565564	0.1245	---
<u>Sleep Duration (actigraphy)</u>	---	---	---	---
Overall difference in means	---	---	---	**0.3610**
Linear trend	---	---	---	**0.0542**
Quadratic trend	---	---	---	0.8512
<u>Sleep Duration model parameters</u>	---	---	---	---
<5 vs. 7-8 hrs.	51.791621	29.267919	0.0777	---
5-<6 vs. 7-8 hrs.	13.836899	24.006759	0.5648	---
6-<7 vs. 7-8 hrs.	10.137378	21.376407	0.6357	---
>8 vs. 7-8 hrs.	-26.418531	36.324785	0.4676	---
Health related QoL Score	16.372484	9.895193	0.0990	---
<u>Model Summary</u>	---	---	---	---
R-square	0.093220	---	---	---
Root MSE	144.836794	---	---	---
<u>Added variables</u>				**AHI p-value**
Frequent Snoring (>=3/wk)	21.605029	21.728033	0.3208	0.1250
Current Smoking	1.733908	9.855922	0.8605	0.2511
Alcohol >2 drinks / day	-18.353704	10.619167	0.0849	0.3333

D. **Risk of Impaired Function Defined by Divided Attention Driving Task (DADT) Mean Lane Deviation > 250**

We next employed our exploratory modeling to examine the relationship between impairment as measured by the Divided Attention Task and sleep apnea severity and our different metrics of sleep duration (see Figures 7.16 and 7.17). We defined impairment as a mean lane tracking error greater than 250 cm. In George et al [1996], mean (SD) tracking error was 228 (145) cm for 21 male patients with obstructive sleep apnea (AHI mean=73 SD=29; mean age=49.3), 71 (31) for age and gender matched controls, and 161 (115) for a group of controls after consumption of alcohol (BAL 103 mg/dl). Thus, values above the criterion of 250 cm reflect impairment at least as large as that among control subjects impaired by alcohol consumption and larger than the mean value in a sample of patients with generally severe sleep apnea. For sleep apnea, the predicted impairment probability increases at higher levels of respiratory disturbance during sleep. Visually it increases above an AHI of 25 episodes/hour. The marked interaction between mean cumulative duration of inactivity in the main sleep bout category and apnea severity is seen in Figure 7.17. Thus, particularly high probability of impairment occurs when there is both severe sleep apnea and a duration of cumulative inactivity <5 hours/night.
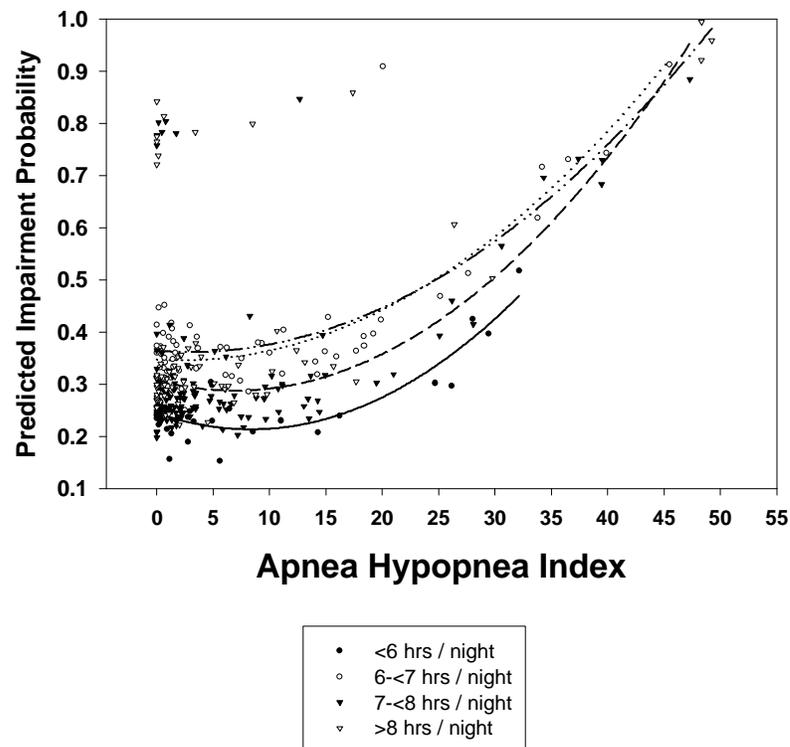
Figure 7.16. Predicted probability of having a DADT mean lane deviation > 250 as a function of AHI and in different mean main bout length of relative inactivity categories. Additional variables included in the model were age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 34.1%.
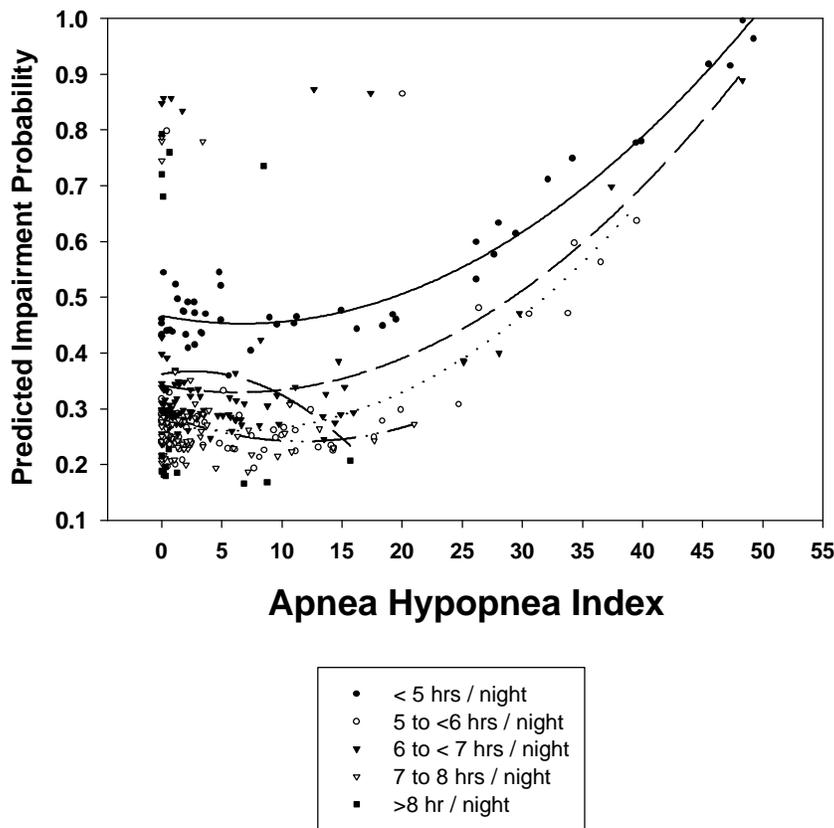
Figure 7.17. Predicted probability of having a DADT mean lane deviation > 250 as a function of AHI and in different mean cumulative duration of inactivity in the main sleep bout categories. Additional variables included in the model were age, BMI, gender, and self-report of health-related quality-of-life. The weighted impairment prevalence for the sample with complete data was 34.1%.

E. **DADT Tracking Decrement Function**

Tables 7.57 and 7.58 provide descriptive statistics and the analysis of variance for the association between the DADT tracking decrement function (i.e., slope) and sleep apnea severity. Least squares estimated expected values for subjects with AHI values >=30, 15-<30, 5-<15, and <5 events per hours (controlling for risk group) were 39.2, 14.1, 4.1, and 8.5. We see that individuals with severe sleep apnea developed increasing 'lane deviations' over time on the test more than twice as fast as those with moderate sleep apnea (p=0.059) and almost 10 times as fast as those with mild sleep apnea (p=0.002). Both linear and quadratic trends were significant. Thus, changes in this parameter were particularly sensitive to severity of sleep apnea.

Table 7.57. Mean DADT slope by sleep apnea severity as determined by PSG for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	109	8.3	41.7	-118.3	296.5	140	8.4	47.4	-89.8	271.0
5 - <15	60	5.9	38.6	-57.3	199.2	16	-0.7	25.3	-41.2	61.1
15 - <30	28	15.5	30.6	-49.0	92.7	2	2.5	22.6	-13.5	18.5
>= 30	22	39.9	99.2	-86.6	408.2					
Total N	219					158				

Table 7.58. Two-way ANOVA (apnea severity and risk group) for DADT slope.

Test	F value	df	P value	Pair-wise contrasts
AHI (Overall)	3.24	3 372	0.022	≥30 vs. <5: p=0.005
Risk Group	0.06	1 372	0.809	≥30 vs. 5-<15: p=0.002
AHI Linear Trend	8.74	1 372	0.003	≥30 vs. 15-<30: p=0.059
AHI Quadratic Trend	4.02	1 372	0.046	15-<30 vs. 5-<15: p=0.025
Risk Group * AHI Category Interaction	0.16	2 370	0.849	

Tables 7.59 and 7.60 provide descriptive statistics and the analysis of variance for the association between the DADT slopes and mean main bout length of relative inactivity category. Tables 7.61 and 7.62 provide descriptive statistics and the analysis of variance for the association between the DADT slope and mean cumulative duration of inactivity in the main sleep bout category.

In contrast to mean tracking errors and some of the other functional measures described above, both definitions of reduced sleep duration demonstrated significant accelerations in decreases in lane tracking ability over time with reduced durations of sleep. Differences were particularly pronounced at the shortest durations. For both metrics the linear trends were statistically significant (p<0.02). For mean main bout length of relative inactivity category, the least squares estimated mean values for drivers with <6 hours, 6 to <7 hours, 7 to 8 hours, and >8 hours were 30.2, 10.7, 4.7, and 4.7, respectively, with an estimated standard deviation of 47.5. Thus, even the category of <6 hours appeared to have greater functional impairments than those with 6 to <7 hours. Similar results were observed when sleep duration was defined using mean cumulative duration of inactivity in the main sleep bout category.

Table 7.59. Mean DADT slope by mean main bout length of relative inactivity category as determined by actigraphy for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	17	46.2	81.8	-50.7	296.5	13	10.2	53.7	-22.2	184.2
6<=hr< 7	47	19.2	64.2	-53.6	408.2	24	-3.0	35.7	-87.2	91.0
7<=hr<= 8	80	2.8	36.5	-118.3	132.5	63	8.0	41.0	-89.8	200.5
hr>8	44	5.9	26.9	-67.5	71.0	32	4.2	54.5	-72.6	271.0
Total N	188					132				

Table 7.60. Two-way ANOVA (mean main bout length of relative inactivity category and group) for mean DADT slope.

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	2.60	3	315	0.052	<6 vs. >8; p=0.013
Risk Group	1.26	1	315	0.263	<7 vs. 7-8; p=0.008
Duration Linear Trend	6.86	1	315	0.009	<6 vs. 6-<7; p=0.061
Duration Quadratic Trend	2.49	1	315	0.116	
Risk Group * Duration Category Interaction	2.34	3	312	0.074	

Table 7.61. Mean DADT slope by mean cumulative duration of inactivity in the main sleep bout category as determined by actigraphy for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	37	34.1	84.8	-60.7	408.2	10	-8.2	40.6	-89.8	76.9
5<=hr< 6	43	16.4	47.6	-85.7	199.2	28	7.2	48.4	-54.0	184.2
6<=hr< 7	65	5.4	25.8	-60.4	71.0	37	5.3	35.2	-87.2	91.0
7-8(ref)	38	-3.1	28.2	-118.3	46.4	45	10.6	53.6	-72.6	271.0
hr> 8	5	-5.1	40.6	-62.8	47.8	12	-7.7	27.5	-68.7	37.7
Total N	188					132				

Table 7.62. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout category and risk group) for mean DADT slope.

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	2.03	4	314	0.090	<5 vs. >8: p=0.027
Risk Group	0.29	1	314	0.590	<5 vs. 7-8: p=0.026
Duration Linear Trend	5.82	1	314	0.016	<5 vs. 6-<7: p=0.023
Duration Quadratic Trend	0.09	1	314	0.759	
Risk Group * Duration Category Interaction	2.11	4	310	0.080	

F. Determinants of DADT Vigilance Decrement Function

The prediction model was applied to the DADT vigilance decrement function to assess the effect of the different variables using multiple regression, as described previously. (For results for mean bout of relative inactivity, see Table 7.63 and for cumulative duration of relative inactivity, see Table 7.64.) The linear trends for apnea severity retained their significance in both models. In contrast to many of the functional parameters described above, mean main bout length of relative inactivity category appeared to influence function (p=0.090 for linear trend and p=0.077 for <6 hours vs. >8 hours) while the effect of mean cumulative duration of inactivity in the main sleep bout category dissipated when controlling for the other variables in the model. Overall, however, for this vigilance decrement function, the presence of severe sleep apnea had a much larger negative impact than short sleep durations did. Thus, this particular aspect of function is particularly negatively impacted by sleep apnea. Given how important maintaining lane tracking ability over time is for the commercial driver, this is an issue of concern. Finally, it is interesting to note that while overall mean tracking error was influenced by age and gender, the tracking error vigilance decrement function did not appear to be so affected.

Table 7.63. Multiple linear regression model for DADT tracking decrement slope. Explanatory variables included sleep apnea severity, mean main bout length of relative inactivity category during prior week as determined by actigraphy, age, gender, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	4.9457	5.9624	0.4074	---
Age - 45.5 (yrs)	-0.1203	0.2397	0.6160	---
BMI - 29.9 (kg/m^2)	0.3094	0.6018	0.6075	---
Female gender	-3.7461	10.8026	0.7290	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.0334
Linear trend	---	---	---	0.0041
Quadratic trend	---	---	---	0.0491
AHI Model parameters	---	---	---	---
>30 vs. <5	43.6188	15.6243	0.0056	---
15-30 vs. <5	7.6684	12.2647	0.5323	---
5-<15 vs. <5	-3.1512	7.2588	0.6645	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.2453
Linear trend	---	---	---	0.0901
Quadratic trend	---	---	---	0.1243
Bout Duration model parameters				
<6 vs. >8 hrs.	18.2661	10.2818	0.0766	---
6-<7 vs. >8 hrs.	-0.6683	8.1548	0.9347	---
7-8 vs. >8 hrs.	-0.1058	6.7428	0.9875	---
Health related QoL Score	-3.5155	3.2294	0.2772	---
Model Summary	---	---	---	---
R-square	0.048	---	---	---
Root MSE	47.0211	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	3.5628	7.2247	0.6222	0.0253
Current Smoking	-0.8431	2.9883	0.7780	0.4456
Alcohol >2 drinks / day	-5.9918	3.5157	0.0893	0.9200

Notes: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

Table 7.64. Multiple linear regression model for DADT tracking decrement slope. Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout category during prior week as determined by actigraphy, age, gender, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	7.173657	5.338019	0.1799	---
Age - 45.5	-0.147635	0.242106	0.5424	---
BMI - 29.9	0.184112	0.611446	0.7635	---
Female gender	-0.436024	11.131675	0.9688	---
<u>**Apnea Hypopnea Index**</u>	---	---	---	---
Overall difference in means	---	---	---	0.0690
Linear trend	---	---	---	0.0097
Quadratic trend	---	---	---	0.0863
<u>AHI Model parameters</u>	---	---	---	---
>30 vs. <5	39.865584	16.058375	0.0136	---
15-30 vs. <5	8.871692	12.275261	0.4704	---
5-<15 vs. <5	-3.291459	7.358743	0.6550	---
<u>**Sleep Duration (actigraphy)**</u>	---	---	---	---
Overall difference in means	---	---	---	0.6559
Linear trend	---	---	---	0.1508
Quadratic trend	---	---	---	0.6247
<u>Sleep Duration model parameters</u>	---	---	---	---
<5 vs. 7-8 hrs.	5.592791	9.544415	0.5583	---
5-<6 vs. 7-8 hrs.	0.893975	7.828725	0.9092	---
6-<7 vs. 7-8 hrs.	-2.432754	6.970953	0.7273	---
>8 vs. 7-8 hrs.	-14.650583	11.845695	0.2171	---
Health related QoL Score	-2.978173	3.226872	0.3567	---
<u>**Model Summary**</u>	---	---	---	---
R-square	0.042338	---	---	---
Root MSE	47.232004	---	---	---
<u>**Added variables**</u>				**AHI p-value**
Frequent Snoring (>=3/wk)	4.028921	7.216894	0.5771	0.0431
Current Smoking	-1.056497	3.024386	0.7271	0.5266
Alcohol >2 drinks / day	-5.415470	3.523674	0.1253	0.9283

7.7. <u>Digit Symbol Substitution Test (DSST)</u>

The Digit Symbol Substitution Test (DSST) assesses cognitive throughput (i.e., cognitive speed and accuracy trade-offs). It is derived from the Wechsler IQ test. The subject sees 9 symbols, and a number identifies each symbol. This test lasts 90 seconds. The subject sees one

symbol at a time and is asked to type the corresponding number. As soon as the subject responds, a new symbol appears. Performance on this test is affected by excessive sleepiness.

Tables 7.65 and 7.66 provide descriptive statistics and the analysis of variance for the association between the DSST number correct and sleep apnea severity.

Table 7.65. Mean DSST number of correct by sleep apnea severity as determined by PSG for CDL holders in the higher and lower risk groups.

	Higher Risk					Lower Risk				
AHI	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5	119	41.77	8.59	14.40	63.80	137	45.39	9.15	22.20	69.60
5 - <15	69	42.75	7.34	28.40	59.80	16	42.16	8.94	22.20	56.80
15 - <30	30	41.43	8.06	28.00	61.00	2	44.30	6.08	40.00	48.60
>= 30	28	37.90	7.06	21.50	53.00					
Total N	246					155				

Table 7.66. Two-way ANOVA (apnea severity and risk group) for DSST number correct.

Test	F value	df	P value	Pair-wise contrasts
AHI (Overall)	2.03	3	396	0.109	≥30 vs. <5: p=0.016
Risk Group	9.22	1	396	0.003	≥30 vs. 5-<15: p=0.024
AHI Linear Trend	5.73	1	396	0.017	
AHI Quadratic Trend	2.00	1	396	0.158	
Risk Group * AHI Category Interaction	1.33	2	394	0.265	

Least squares estimated mean values for subjects with AHI≥30, 15-<30, 5-<15, and <5 event per hours were 39.4, 42.9, 43.5, and 43.6 correct responses, respectively. A statistically significant linear trend reflecting decreases in the numbers of correct DSST responses for increasing apnea severity was observed (F=5.7, df=1,396, p=0.017).

Tables 7.67 and 7.68 provide descriptive statistics and the analysis of variance for the association between the DSST number correct and mean cumulative duration of inactivity in the main sleep bout category. Tables 7.69 and 7.70 provide the analogous results for mean cumulative duration of inactivity in the main sleep bout category. Significant differences were only observed for mean cumulative duration of inactivity in the main sleep bout category for this variable (DSST number correct).

Table 7.67. Mean DSST number correct by mean main bout length of relative inactivity category as determined by actigraphy for CDL holders in the higher and lower risk groups.

Duration	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<6 hr	20	40.69	9.01	14.40	54.60	13	43.87	12.79	22.20	64.33
6<=hr< 7	53	41.26	6.77	25.20	56.80	23	44.55	10.80	22.20	69.60
7<=hr<= 8	88	41.36	9.16	21.50	61.00	61	44.49	7.25	30.00	61.80
hr>8	47	40.34	6.48	28.80	53.40	32	47.26	9.45	25.00	67.80
Total N	208					129				

Table 7.68. Two-way ANOVA (mean main bout length of relative inactivity category and risk group) for mean DSST number correct.

Test	F value		df	P value	Pair-wise t-test p<0.05
Duration (Overall)	0.15	3	332	0.928	
Risk Group	18.33	1	332	0.000	
Duration Linear Trend	0.40	1	332	0.529	
Duration Quadratic Trend	0.01	1	332	0.939	
Risk Group * Duration Category Interaction	0.94	3	329	0.419	

Table 7.69. Mean DSST number correct by mean cumulative duration of inactivity in the main sleep bout category as determined by actigraphy for CDL holders in the higher and lower risk groups.

AHI	Higher Risk					Lower Risk				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
<5 hr	45	38.94	8.51	14.40	56.80	10	38.04	9.22	22.20	48.00
5<=hr< 6	47	40.91	7.15	23.80	53.20	28	45.27	9.67	22.20	69.60
6<=hr< 7	67	42.46	8.60	25.00	61.00	35	43.92	8.67	28.80	62.00
7-8(ref)	44	41.58	7.07	28.80	56.50	44	46.25	9.24	25.00	67.80
hr> 8	5	37.40	7.26	28.80	48.40	12	50.05	4.60	41.60	56.40
Total N	208					129				

Table 7.70. Two-way ANOVA (mean cumulative duration of inactivity in the main sleep bout category risk group) for mean DSST number correct.

Test	F value	df		P value	Pair-wise contrasts
Duration (Overall)	2.58	4	331	0.037	<5 vs. >8: p=0.015
Risk Group	12.08	1	331	0.001	<5 vs. 7-8: p=0.006
Duration Linear Trend	6.39	1	331	0.012	<5 vs. 6-<7: p=0.010
Duration Quadratic Trend	0.32	1	331	0.572	<5 vs. 6-<7: p=0.038
Risk Group * Duration Category Interaction	2.15	4	327	0.075	

Least squares estimated mean values controlling for risk group for subjects with mean cumulative duration of inactivity in the main sleep durations of <5, 5-<6, 6-<7, 7-8, and >8 hours were 39.8, 43.0, 43.5, 43.9, and 45.6 correct responses, respectively. The linear trend in the expected number of correct responses was statistically significant (F=6.4, df=1,331, p=0.012). Less than 5 hours of cumulative inactivity was associated with a precipitous drop in the number of DSST correct responses.

As for other measures of performance, we carried out multiple linear regression to assess the role of the different key variables. The results are shown in Tables 7.71 and 7.72. Results depended upon which metric was used to define sleep duration. In the first model using mean main bout length of relative inactivity category, the significant linear trend for sleep apnea severity was retained (p=0.019) while the duration itself was not significant. In contrast, when duration was defined on the basis of mean cumulative duration of inactivity in the main sleep bout category, and controlling for other variables the effect of increasing severity of apnea was marginal (p=0.0872) but that for sleep duration highly significant (p=0.0026). It should be pointed out that the presence of apnea itself affects this variable (see Chapter Four). As for some other measures of performance, the magnitude of the effect of severe apnea, as compared to no apnea, and that of short sleep (<5 hours) compared to normal sleep (7-8 hours), were of similar orders of magnitude (-3.35 and –4.04, respectively).

Table 7.71. Multiple linear regression model for DSST number correct. Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout category as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	45.0770	0.8943	0.0000	---
Age - 45.5 (yrs)	-0.4417	0.0359	0.0000	---
BMI - 29.9 (kg/m^2)	0.1050	0.0867	0.2265	---
Female gender	1.6583	1.6458	0.3144	---
Apnea Hypopnea Index	---	---	---	---
Overall difference in means	---	---	---	0.1302
Linear trend	---	---	---	0.0185
Quadratic trend	---	---	---	0.3551
AHI Model parameters	---	---	---	---
>30 vs. <5	-4.5592	2.0858	0.0295	---
15-30 vs. <5	-1.9937	1.8219	0.2746	---
5-<15 vs. <5	-0.0326	1.0840	0.9760	---
Sleep Bout Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.2885
Linear trend	---	---	---	0.2105
Quadratic trend	---	---	---	0.2887
Bout Duration model parameters				
<6 vs. >8 hrs.	-1.5966	1.5287	0.2971	---
6-<7 vs. >8 hrs.	-2.3611	1.2304	0.0559	---
7-8 vs. >8 hrs.	-1.1985	1.0196	0.2407	---
Health related QoL Score	-0.5882	0.4825	0.2237	---
Model Summary	---	---	---	---
R-square	0.363	---	---	---
Root MSE	7.1915	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	-0.0376	1.0574	0.9717	0.0660
Current Smoking	-0.7037	0.4889	0.1510	0.2224
Alcohol >2 drinks / day	0.4336	0.6214	0.4858	0.1138

Note: Age, BMI, and the self-report health-related QoL measure were centered by subtracting their weighted mean value in order to make the intercept interpretable as the expected value of a 45.5 year old male with a BMI of 29.9 kg/m2 without sleep apnea (AHI<5) who had an average QoL value of 2.44 and who had a mean main bout length of relative inactivity of more than 8 hours.

Table 7.72. Multiple linear regression model for DSST number correct. Explanatory variables included sleep apnea severity, mean cumulative duration of inactivity in the main sleep bout category as determined by actigraphy, age, body mass index (BMI), and a single-item health-related quality of life indicator.

Variable	Parameter Estimate	SE	Parameter p-value	Factor p-value
Intercept	45.106263	0.784977	0.0000	---
Age - 45.5	-0.427793	0.035808	0.0000	---
BMI - 29.9	0.149684	0.087023	0.0864	---
Female gender	1.131554	1.673294	0.4994	---
Apnea Hypopnea Index	**---**	**---**	**---**	---
Overall difference in means	---	---	---	0.3935
Linear trend	---	---	---	0.0872
Quadratic trend	---	---	---	0.4948
AHI Model parameters	---	---	---	---
>30 vs. <5	-3.354266	2.119468	0.1145	---
15-30 vs. <5	-1.492558	1.802488	0.4083	---
5-<15 vs. <5	0.001586	1.082401	0.9988	---
Sleep Duration (actigraphy)	---	---	---	---
Overall difference in means	---	---	---	0.0245
Linear trend	---	---	---	0.0026
Quadratic trend	---	---	---	0.8553
Sleep Duration model parameters	---	---	---	---
<5 vs. 7-8 hrs.	-4.045699	1.382818	0.0037	---
5-<6 vs. 7-8 hrs.	-2.009980	1.158603	0.0837	---
6-<7 vs. 7-8 hrs.	-1.229326	1.044227	0.2400	---
>8 vs. 7-8 hrs.	1.419968	1.782536	0.4263	---
Health related QoL Score	-0.473462	0.475551	0.3202	---
Model Summary	---	---	---	---
R-square	0.377967	---	---	---
Root MSE	7.119325	---	---	---
Added variables				**AHI p-value**
Frequent Snoring (>=3/wk)	0.156854	1.048773	0.8812	0.2578
Current Smoking	-0.839810	0.484787	0.0842	0.5209
Alcohol >2 drinks / day	0.389442	0.612844	0.5256	0.2624

7.8. Relationships Among Objective Measures

The results from the analyses of these various tests of performance in general lead to similar conclusions. An analysis was, therefore, performed to determine the extent to which the MSLT, PVT, DADT, and DSST response parameters were measuring similar functional domains. As with the analysis of subjective assessments, we present a Pearson partial correlation

matrix controlling for risk group in order to simplify the presentations. Table 7.73 contains the pair-wise Pearson partial correlations among these seven objective sleepiness measures.

Table 7.73. Pearson partial correlations among seven objective assessments of sleepiness controlling for risk group. P-values are listed in parentheses. For PVT median response times, PVT vigilance lapses, and DADT sum of tracking errors larger values reflect poorer performance.   For DSST number correct, smaller values indicate poorer performance. For the PVT vigilance slope and the DADT tracking decrement slope, the more negative the value, the greater the increase in response times over time.

	PVT MEDIAN	PVT LAPSE	PVT SLOPE	DADT TRACKING	DADT SLOPE	DSST CORRECT
Mean Sleep Latency Test (MSLT)	-0.071 (p=0.171)	-0.103 (p=0.047)	+0.030 (p=0.568)	-0.141 (p=0.006)	-0.120 (p=0.021)	+0.035 (p=0.505)
PVT Median Response Time (PVT MEDIAN)	1	+0.742 (p<0.0001)	-0.250 (p<0.0001)	0.495 (p<0.0001)	+0.306 (p<0.0001)	-0.334 (p<0.0001)
PVT Number of Lapses (PVT LAPSE)		1	-0.280 (p<0.0001)	0.505 (p<0.0001)	+0.471 (p<0.0001)	-0.302 (p<0.0001)
PVT Vigilance Slope (PVT Slope)			1	-0.176 (p=0.0006)	-0.185 (p=0.0003)	0.081 (p=0.120)
Divided Attention Driving Task Mean Tracking Error (DADT TRACKING)				1	0.539 (p<0.0001)	-0.336 (p<0.0001)
Divided Attention Driving Tracking Slope (DADT Slope)					1	-0.124 (p=0.017)

As would be expected, the correlation between PVT median response times and number of vigilance lapses was large (r=0.742). There was also a statistically significant correlation between the PVT vigilance slopes and the DADT tracking decrement slope r=-0.185 (p=0.0003), providing convergent validity for both of these independently measured functional decrement over time assessments. Similarly, the correlation between PVT median response time and DADT mean tracking error was fairly large (r=0.495). There were also reasonably large (negative) correlations between the DSST number correct relative to PVT median response time (r=-0.334), PVT vigilance lapses (r=-0.304), and DADT tracking error (r=-0.336). Statistically significant but small correlations were present between MSLT and PVT lapses (r=-0.103, p=0.047) and between MSLT and DADT tracking slope (p=-0.120, p=0.021) and between DADT mean tracking (r=-0.141).   These correlations are supportive of our strategy to do multiple assessments of performance so that our study can maintain concurrent validity checks of critical functional assessment measurements.

7.9. Summary and Discussion

This lengthy chapter of our final report contains extensive data about the performance of our sample of CDL holders on various tests that are known to be affected by sleep deprivation and/or alcohol administration. We had deliberately set out to do multiple assessments of

performance. Particularly critical measures were the following: multiple sleep latency (a measure of the physiological pressure for sleep), reaction time from psychomotor vigilance reaction time (PVT) test, number of performance lapses on PVT task, lane deviations in the divided attention task (a task designed to simulate the "cognitive load" of driving), and decrements in lane tracking while performing the divided attention task.

Although these are separate measures involving three completely different testing paradigms, the results of our analyses of the data for each of these variables lead to similar conclusions.

First, we found that a substantial fraction of individuals that we tested were excessively sleepy by objective testing and had performance decrements in tasks known to be affected by excessive sleepiness. When we defined cut-points of impairment derived from either the clinical literature or literature studying the effects of alcohol intoxication, we found remarkably similar percentages of individuals that we tested who were "impaired". Specifically, we found the following: 25.6% of subjects had pathological excessive sleepiness (i.e., multiple sleep latency test on average less than 5.0 minutes); 30.6% had medium response time on PVT greater than 275 msec. This is found when individuals are twice over the legal limit for blood alcohol; 29.2% had more than 3 performance lapses in a ten-minute trial. This again is found in individuals who are over twice the blood alcohol limit; 34.1% had, on average, a lane deviation error of >250 cm in the Divided Attention Driving Task. (This is greater than the average deviation of 161 cm for controls rendered intoxicated (BAC 103 mg/dl).) Thus, approximately 1/4 to 1/3 of the sample crossed thresholds previously identified as being impaired. This is a very large minority and an obvious cause for concern.

These performance decrements are related to whether subjects have sleep apnea or not. For all tests we found significant decrements in performance and/or marked excessive sleepiness for individuals with severe sleep apnea, i.e., an apnea/hypopnea index greater than 30 episodes/hour. For some tests, but not all, decrements are also found on average for individuals with moderate sleep apnea, i.e., AHI between 15 and 30 episodes/hour. Thus, we can state unequivocally that on average CDL holders with AHI≥30 episodes/hour are excessively sleepy and have resulting decrements in performance on tasks related to driving ability. Since such individuals can be treated, there is a need to develop strategies to identify such individuals in the workplace.

Our data for the impact of sleep duration are not as clear cut. We recognized challenges involving at-home sleep duration assessments using actigraphy in a population containing individuals with severe sleep apnea (see Chapter Four). Since cumulative duration of inactivity in the main sleep bout may be confounded by movements associated with apneic events, we defined a metric unaffected by such movements, namely, mean main bout length of relative inactivity category. However, this metric may overestimate sleep duration since it includes wakefulness after sleep onset within the main sleep bout. Thus, to meet these challenges, we chose to present results for both metrics of sleep duration. For several functional parameters, mean main bout length of relative inactivity category was not significant while mean cumulative duration of inactivity in the main sleep bout category was significant after controlling for sleep apnea severity. This finding is consistent with the hypothesis that cumulative duration of

inactivity in the main sleep bout is confounded by apnea-related activity during sleep. Hence, we cannot interpret this result as implying an affect of sleep duration per se. However, for two of our primary measures of function, MSLT and PVT lapses, we found simultaneously significant contributions of sleep apnea and short duration of the major bout of relative inactivity on reduced function. These findings were replicated by a vigilance decrement function defined on the basis of a divided attention driving task. Since this variable to assess sleep duration overestimates sleep duration, as we discussed in Chapter Four, we can reasonably assume that sleep duration at home is having an effect on these measures. Thus, we believe that we have established that cumulative partial sleep deprivation occurs in commercial drivers and has an impact on some key performance measures related to the driving task. In these tasks, when we did demonstrate an association between performance on the task or degree of excessive sleepiness, and the duration of the major bout of relative inactivity, the magnitude of effect of severe sleep apnea (AHI>30 episodes/hour) and of "sleep duration" of <6 hours was approximately equivalent.

Interventions designed at reducing excessive daytime sleepiness in commercial drivers should, we propose, include both sleep apnea and sleep duration components.

The prevalence of severe sleep apnea in our sample was 4.6% while that for the mean main bout length of relative inactivity <6 hour/night was 9.8%. (The prevalence of mean cumulative duration of inactivity in the main sleep duration <5 hours/night, at which level we saw impact on awake functioning, was 13.5%). Thus, severe sleep apnea and short duration (measured either way) not only have a similar magnitude of effect on performance, albeit on some of the tests we utilized, but both are prevalent in this population of commercial drivers. The prevalence of short sleep duration is somewhat greater, however, than for severe sleep apnea. Both of these issues need to be a source of concern to the commercial trucking industry.

Both of these variables, i.e., severity of sleep apnea and sleep duration, affect performance. This results in a complex interaction in the final determination of performance or degree of excessive sleepiness. The individuals most at risk are those with both severe sleep apnea and short sleep duration. As discussed in Chapter Five, short sleep duration is itself a risk factor for sleep apnea, thereby further complicating the picture.

Nevertheless, our results clearly indicate that from a public policy perspective, the drivers with documented risk for excessive sleepiness and decrements in performance are those with severe sleep apnea ($\geq$30 episodes/hour) and/or average nightly sleep durations of less than 6 hours.

These clear relationships, that are found for almost all key objective tests, are in contrast to the lack of such relationships with self-report measures of sleepiness. This is not because self-reported sleepiness in CDL holders does not occur; indeed it occurs relatively frequently. But its occurrence is not related to the presence of risk factors such as sleep apnea. Why this should be so is unclear. What is clear is that one cannot rely solely on self-report measure of sleepiness to identify individuals with correctable risk factors. Such individuals can only be identified by probing for the risk factors themselves, or alternatively carrying out routine performance tests on drivers. Tests such as the psychomotor vigilance reaction time task or divided attention task are

simple to perform, have limited learning effects and normative data are available. Drivers showing the type of performance decrements reported here might subsequently be assessed for the identified and correctable problems—severe sleep apnea and chronic inadequate sleep.

CHAPTER EIGHT

Findings and Recommendations

8.1 Findings

This study on the prevalence of effects of sleep apnea in commercial drivers led to the following findings:

1. Sleep apnea is common in holders of commercial drivers licenses. While common, its prevalence is considerably less than reported in an earlier study. Rather, the prevalence is similar to studies in other more general populations.

2. Mild sleep apnea occurs in 17.6% of holders of commercial drivers licenses; moderate sleep apnea in 5.8%; and severe sleep apnea in 4.7%.

3. The major determinants of prevalence of sleep apnea are age, with increasing prevalence of apnea with increasing age, and degree of obesity as measured by body mass index (BMI). (Increasing obesity leads to increasing prevalence of sleep apnea.) The effects of age and obesity are multiplicative. Thus, age has a more marked effect on prevalence of apnea in individuals who are obese, while obesity has a more marked effect on prevalence in older individuals. Equations are developed to allow calculation of prevalence of sleep apnea based on the distribution of age and BMI in any target population of CDL holders.

4. Sleep apnea prevalence is also dependent on average nightly sleep duration at home, with shorter durations of sleep being associated with higher prevalence of apnea.

5. Sleep duration at home was difficult to measure in this study due to movements occurring in some individuals during sleep as a result of the presence of sleep apnea. Even, however, using cautious definitions of durations of sleep, we found that 9.8% of commercial drivers slept less than 6 hours/night on average. This is a duration that we found to be associated with performance impairments in tests in our laboratory.

6. Short sleep durations are associated with subjects terminating sleep early in the morning. 35.6% of CDL holders terminate sleep before 6:00 am and 12.3% before 5:00 am.

7. A large percentage of CDL holders (32.6%) have excessive levels of self-reported sleepiness. However, there is no association between measures of self-reported sleepiness and the presence of sleep apnea. The basis for this lack of association is unclear, but self-report measures of sleepiness cannot be used to identify drivers likely to have sleep apnea.

8. When we applied clinical definitions of abnormality, or definitions arising from studies employing alcohol use in volunteers, we found a remarkably similar percentage of subjects were in the abnormal range for different tests of awake performance. This percentage ranged from 25.6% to 34.1% for different tasks. This high percentage is a source of concern.

9. All objective tests of performance such as assessment of reaction times, performance lapses, lane tracking ability and objectively measured sleepiness show relationships with severity of sleep apnea. Changes take place particularly in those with severe sleep apnea, i.e., those subjects with more than 30 episodes/hour of breathing abnormalities/hour of sleep.

10. We did not find a clear relationship between sleep duration measured at home and all tests of performance that we performed. We did find clear relationships for important variables, specifically measurement of objective degree of sleepiness, performance lapses and for decline in lane tracking ability over time while performing continuously.

11. Decrements in performance and excessive, objectively measured, sleepiness are found in individuals with severe sleep apnea (apnea/hypopnea index ≥ 30 episodes/hour) and those sleeping less than 6 hours/night. In those tests affected by both factors, the effect on performance measures of both of these abnormalities are similar in magnitude.

8.2 Recommendations

A. Educational Needs

Our results lead us to believe that there is a need for an extensive national program of education for all components of the commercial driving industry about the problems with inadequate sleep and the reasons for it, as well as the symptoms and risk factors for obstructive sleep apnea. This education needs to be directed at drivers, their spouses, safety officers in companies, and owners as well as physicians carrying out routine medical examinations of drivers. Programs on these topics have been developed for the general population by the National Center for Sleep Disorders Research at the National Institutes of Health and such programs could be the basis for this educational effort [see Dinges et al, 1999].

B. Altering the Routine History and Physical Examination for Commercial Drivers to Assess for Presence of Sleep Apnea

Results from routine medical examinations for commercial drivers can be used to develop tools that could help the examining physician to determine the likelihood of apnea at the time of examination. These tools could assist the physician to decide which drivers will need sleep studies to rule out (or in) the presence of sleep apnea.

Since the routine medical examination is such a key component of this strategy, consideration should be given to developing techniques to assure high quality in such examinations. This could be done, for example, by having a small random percentage of such drivers assessed with a repeat evaluation by an expert in sleep medicine, so that the quality of the initial evaluation could be assessed.

The routine medical examination will also be a place where a detailed history about sleep patterns can be obtained so that subjects with cumulative partial sleep deprivation could be identified.

C. Need for Weight Management Programs

Our results reveal that holders of CDLs and active commercial drivers have, as a group, a high prevalence of obesity. This is true even for individuals in their twenties. This high level of obesity may be a consequence of life-style. The high prevalence of obesity is a major risk factor for the presence of sleep apnea. Moreover, we know that obesity is a risk factor for a number of other medical problems—diabetes, hypertension and other cardiovascular diseases. Consideration should be given to developing at least pilot weight management programs for commercial drivers to assess the efficacy of this strategy.

D. Research Needs

While this study has produced considerable informative data, we believe that there are several critical gaps in our current knowledge that could be addressed by appropriate research.

1. Role of Sleep Apnea on Commercial Drivers in Crash Causation

A critical area where there are no definitive data is whether sleep apnea in commercial drivers is a risk factor for crashes and, if so, how big a risk factor. There have been considerable studies of this for passenger car drivers but not for commercial drivers. The most logical design, in our view, given the likely multifactorial risk factors for crashes, is a case-control study. This would be similar in concept to the study of Teran-Santos et al [1999] for passenger cars (see Chapter Two) but avoiding the problem related to the way controls were selected in that study.

This design would require study of drivers who had a major crash. (This could be further qualified to likely fall-asleep crashes based on criteria such as those used by the National Transportation Safety Board in their study of fatigue-related crashes [1995].) Controls, who were potentially matched for gender or age and who had not had a crash in a defined period, e.g., three years, would also be studied. Drivers in both groups (cases and controls) would get overnight sleep studies to detect sleep apnea. This would permit determination of the increased risk of such crashes as a result of sleep apnea of different degrees of severity. The prevalence estimates of sleep apnea reported here would be invaluable to allow calculation of an appropriate sample size for this study.

2. Efficacy of Screening for Sleep Apnea and Effects of Treatment

Our study leads to the logical conclusion that we need to begin to develop sensitive and specific methodologies for identification of drivers who have severe sleep apnea. This methodology needs to be cost-effective and a number of different approaches are currently being evaluated in general population studies. We need to determine how well such methodologies work in commercial drivers. To evaluate this, a study would need to be done to compare the accuracy of such case identification strategies in comparison to the "gold standard", i.e., an overnight sleep study in a laboratory.

But, case identification will only be of value if drivers so identified can be effectively treated with devices such as nasal continuous positive airway pressure (CPAP) and show clear benefit from the therapy. Moreover, it would need to be demonstrated that drivers would continue to use the therapy, i.e., be adherent to it. This would argue that a research study on case identification should also include a treatment evaluation component, where outcomes of treatment and adherence to therapy can be assessed. For commercial drivers, an outcome of importance would be their performance on, for example, a driving simulator. Such a study done as a pilot in certain commercial vehicle companies would develop guidelines for the rest of the industry.

3. Economic Analysis of Detection of Sleep Apnea

Case identification of apnea and its treatment is associated with costs. There are also unknown costs of allowing drivers with apnea to continue driving since we believe they will be at an increased risk for commercial vehicle crashes. There is a need for a careful analysis of this so that the risk/benefit can be assessed in economic terms.

4. Identification of Commercial Drivers with Impairment as a Result of Excessive Sleepiness

An alternative strategy to be considered, based on our results, is that techniques might be used to identify drivers with performance decrements based on tests such as the psychomotor vigilance reaction time (PVT) task and/or divided attention task. If drivers are identified as having decrements in performance on such tasks, the next step would be to develop strategies to determine whether these performance decrements are the result of severe sleep apnea and/or chronic inadequate sleep. Studies need to be done to assess the feasibility of this strategy and whether individuals with correctable risk factors for excessive sleepiness are identified. Moreover, whether correction of the risk factor is feasible also needs to be evaluated. The relative efficacy of this approach, compared to focusing solely on case identification of sleep apnea, needs to be assessed.

E. Consensus Panel to Make Recommendations Regarding Public Policy

The results of our study also raise a large number of policy questions. Rather than give specific recommendations, we propose that a consensus panel be established to discuss these questions. We propose that this panel would be given all of the results from the study described here and could, if desired, ask for additional analyses based on the data we collected. The meeting of the panel would commence with a formal presentation by investigators from this study on our results. We envisage that the panel would contain experts in sleep medicine, commercial vehicle operation, transportation policy, traffic safety, crash causation, medical policy making, etc.

We believe that our study raises the following public policy questions. These questions are relevant to many modes of transportation.

1. If a driver has a likely fall-asleep crash, what steps should be taken to evaluate this driver for the presence of sleep apnea and/or a life-style that leads to cumulative partial sleep deprivation (chronic inadequate sleep)?
2. What definition(s) should be used for a likely fall-asleep crash?
3. Should drivers being considered for employment have tests of performance such as the psychomotor reaction time test and/or divided attention task so that drivers with impairment secondary to excessive sleepiness can be identified, evaluated and appropriately treated?
4. Should steps be taken to evaluate all commercial drivers at high risk for severe sleep apnea as a condition of employment by formal overnight sleep studies?
5. If yes to 4, given the data presented, and the dependence of apnea prevalence on age and degree of obesity, how should high risk for severe sleep apnea be defined?
6. If a commercial driver has been found to have sleep apnea, what steps should be taken to ensure that he/she is adherent to therapy? How frequently should this driver be assessed medically, and what form should this assessment take?
7. What steps should be taken to identify the commercial driver who is chronically excessively sleepy because of chronic inadequate amounts of sleep? How should this be corrected and what steps should be taken to assess the effectiveness of its correction?

It seems likely that policy can be developed for some of these issues but the others will necessitate more research. There will be, in the future, a need to evaluate the effect of any policy change on the safety and work practices in the industry.

BIBLIOGRAPHY

Akerstedt T, Gillberg M. The Karolinska Sleepiness Scale. Int J Neurosci 52:29-32, 1990.

Akerstedt T. Personal communication to A. Pack – April 19, 2000.

Ambulatory Monitoring, Inc. Action-W Version 1.06. Windows data analysis for Motionlogger actigraphs, 1996.

American Academy of Sleep Medicine Task Force. Sleep-disordered breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. Sleep 22:667-689, 1999.

Ancoli-Israel S, Kripke DF, Klauber MR, Mason WJ, Fell R, Kaplan O. Sleep-disordered breathing in community-dwelling elderly. Sleep 14:486-495, 1991.

Anonymous. Executive summary of the clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults. Arch Intern Med 158:1855-1867, 1998.

ATA Foundation. Empty seats and musical chairs. Critical success factors in truck driver retention. Report prepared by the Gallup Organization, 1997.

ATA Foundation. Compensation study. Printed by ATA, 1997.

Ayas NT, Epstein LJ. Oral appliances in the treatment of obstructive sleep apnea and snoring. Curr Opin Pulm Med 4:355-360, 1998.

Bamford CR, Quan SF. Bacterial meningitis—a possible complication of nasal continuous positive airway pressure therapy in a patient with obstructive sleep apnea syndrome and a mucocele. Sleep 16:31-32, 1993.

Bassetti C, Aldrich MS. Sleep apnea in acute cerebrovascular diseases: final report on 128 patients. Sleep 22:217-223, 1999.

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 151:1459-1465, 1995.

Bettega G, Pepin J-L, Veale D, Deschaux C, Raphael B, Levy P. Obstructive sleep apnea syndrome: fifty-one consecutive patients treated by maxillofacial surgery. Am J Respir Crit Care Med 162:641-649, 2000.

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 157:144-148, 1998.

Bliwise D, Carskadon M, Carey E, Dement W. Longitudinal development of sleep-related respiratory disturbance in adult humans. J Gerontol 39:290-293, 1984.

Blood ML, Sack RL, Percy DC, Pen JC. A comparison of sleep detection by wrist actigraphy, behavioral response, and polysomnography. Sleep 20:388-395, 1997.

Bloom JW, Kaltenborn WT, Quan SF. Risk factors in a general population for snoring. Importance of cigarette smoking and obesity. Chest 93:678-683, 1988.

Brochu M, Poehlman ET, Ades PA. Obesity, body fat distribution, and coronary artery disease. J Cardiopulm Rehab 20:96-108, 2000.

Carmines EG, Zeller RA. *Reliability and Validity Assessment*. Quality Applications in the Social Sciences No. 17, Beverly Hill, Sage Publications, 1979.

Carskadon MA, Dement WC. The multiple sleep latency test: What does it measure? Sleep 5:S67-S72, 1982.

Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook RR, Keenan S. Guidelines for the multiple sleep latency test (MSLT): A standard measure of sleepiness. Sleep 9:519-524, 1986.

Cassel W, Ploch T, Becker C, Dugnus D, Peter JH, von Wichert P. Risk of traffic accidents in patients with sleep-disordered breathing: reduction with nasal CPAP. Eur Respir J 9:2606-2611, 1996.

Cirignotta F, D'Alessandro R, partinen M, Zucconi M, Cristina E, Gerardi R, Cacciatore FM, Lugaresi E. Prevalence of every night snoring and obstructive sleep apnoeas among 30-60-year-old men in Bologna, Italy. Act Neurol Scand 79:366-372, 1989.

Clinical Guidelines on the Identification, Evaluation and Treatment of Overweight and Obesity in Adults. NIH Publication No. 98-4083, 1998.

Cochran WG. Sampling Techniques, New York: Wiley, 1977.

Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, second edition. Hillsdale: Lawrence Erlbaum Associates pp 8-14, 1998.

Colhoun HM, Hemingway H, Poulter NR. Socio-economic status and blood pressure: an overview analysis. J Hum Hypertension 12:91-110, 1998.

Connor J, Whitlock G, Norton R, Jackson R. The role of driver sleepiness in car crashes: a systematic review of epidemiological studies. Accident Analysis and Prevention 33:31-41, 2000.

Conradt R, Hochban W, Brandenburg U, Heitmann J, Peter JH. Long-term follow-up after surgical treatment of obstructive sleep apnea by maxillomandibular advancement. Eur Respir J 10:123-128, 1997.

Davies RJO, Stradling JR. The epidemiology of sleep apnoea. Am J Respir Crit Care Med 152:711-716, 1995.

Davies RJ, Stradling JR. The efficacy of nasal continuous positive airway pressure in the treatment of obstructive sleep apnea syndrome is proven. Am J Respir Crit Care Med 161:1775-1776, 2000.

Dinges DF, Powell JW. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. Behav Res Methods, Instrumentation and Computers 17:6520655, 1985.

Dinges DF, Orne MT, Whitehouse WG, Orne EC. Temporal placement of a nap for alertness: contributions of circadian phase and prior wakefulness. Sleep 10:313-329, 1987.

Dinges DF, Kribbs NB. Performing while sleepy: effects of experimentally-induced sleepiness. In Monk TH (ed), *Sleep, Sleepiness and Performance*. John Wiley and Sons, Ltd., Chichester, United Kingdom, 1991.

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, pp 177-188, 1992.

Dinges DF, Douglas SD, Zaugg L, Campbell DE, McMann JM, Whithouse WG, Orne EC, Kapoor SC, Icaza E, Orne MT. Leukocytosis and natural killer cell function parallel neurobehavioral fatigue induced by 64 h of sleep deprivation. J Clin Invest 93:1930-1939, 1994.

Dinges DF, Pack F, William 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 20:267-277, 1997.

Douglas NJ. Upper airway resistance syndrome is not a distinct syndrome. Pro/Con editorial. Am J Respir Crit Care Med 161:1413-1414, 2000.

Dunn G. Design and analysis of reliability studies. Statistical Methods in Medical Research 1:123-157, 1992.

Dyken ME, Somers VK, Yamada T, Ren ZY, Zimmerman MB. Investigating the relationship between stroke and obstructive sleep apnea. Stroke 27:401-407, 1996.

Engleman EM, Martin JE, Dreary IJ, Douglas NJ. Effect of CPAP therapy on daytime function on patients with mild sleep apnoea-hypopnea syndrome. Thorax 52:114-119, 1997.

Engleman HM, Kingshott RN, Wraith PK, Mackay TW, Deary IJ, Douglas NJ. Randomized placebo-controlled crossover trial of continuous positive airway pressure for mild sleep apnea/hypopnea syndrome. Am J Respir Crit Care Med 159:461-467, 1999.

Feldman HA. Families of lines: random effects in linear regression analysis. J Appl Physiol 64:1721-1732, 1988.

Findley L, Fabrizio M, Thommi G, Suratt P. Severity of sleep apnea and automobile crashes. N Engl J Med 320:868-869, 1989.

Findley L, Unverzagt M, Guchu R, Fabrizio M, Buckner J, Suratt P. Vigilance and automobile accidents in patients with sleep apnea or narcolepsy. Chest 108:619-624, 1995.

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

Findley L, Smith C, Hooper J, Dineen M, Suratt PM. Treatment with nasal CPAP decreases automobile accidents in patients with sleep apnea. Am J Respir Crit Care Med 161:857-859, 2000.

Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL. Overweight and obesity in the United State: prevalence and trends, 1960-1994. Intl J Obesity 22:39-47, 1998.

Fleiss JK. *The Design and Analysis of Clinical Experiments.* New York, John Wiley and Sons, Chapter 1, 1986.

Flemons WW, Whitelaw WA, Brant R. Likelihood ratios for sleep apnea clinical prediction rule. Am J Respir Crit Care Med 150:1279-1285, 1994.

Fujimoto WY. The importance of insulin resistance in the pathogenesis of type 2 diabetes mellitus. Am J Med 108:9S-14S, 2000.

Geer RT, Jobes DR, Gilfor J, Traber KB, Dinges D. Reduced psychomotor vigilance in anesthesia residents after 24-hr on-call. Anesthesiology 83:A1008, 1995.

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

George CF, Boudreau AC, Smiley A. Effects of nasal CPAP on simulated driving performance in patients with obstructive sleep apnoea. Thorax 52:648-653, 1997.

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

George CF, Smiley A. Sleep apnea and automobile crashes. Sleep 22:790-795, 1999.

Gillum RF. Risk factors for stroke in blacks: a critical review. Am J Epidemiology 150:1266-1274, 1999.

Gislason T, Almqvist M, Eriksson G, Taube A, Boman G. Prevalence of sleep apnea syndrome among Swedish men - an epidemiological study. J Clin Epidemiol 41:571-576, 1988.

Gislason T, Benediktsdottir B, Bjornsson JK, Kjartansson G, Kjeld M, Kristbjarnarson H. Snoring, hypertension, and the sleep apnea syndrome. An epidemiologic survey of middle-aged women. Chest 103:1147-1151, 1993.

Gonzalez-Rothi R, Foresman G, Block A. Do patients with sleep apnoea die in their sleep? Chest 94:531-538, 1988.

Gould GA, Whyte KF, Rhind GB, Airlie MA, Catterall JR, Shapiro CM, Douglas NJ. The sleep hypopnea syndrome. Am Rev Respir Dis. 137:895-898, 1988.

Grunstein R, Wilcox I, Yang TS, Gould Y, Hedner J. Snoring and sleep apnoea in men: association with central obesity and hypertension. Int J Obes Relat Metab Disord 17:533-540, 1993.

Guilleminault C, Stoohs R, Clerk A, Cetel M, Maistros P. A cause of excessive daytime sleepiness. The upper airway resistance syndrome. Chest 104:781-787, 1993.

Hack M, Davies RJ, Mullinis R, Choi SJ, Ramdassingh-Dow S, Jenkinson C, Stradling JR. Randomised prospective parallel trial of therapeutic versus subtherapeutic nasal continuous positive airway pressure on simulated steering performance in patients with obstructive sleep apnoea. Thorax 55:224-231, 2000.

Haraldsson PO, Carenfelt C, Diderichsen F, Nygren A, Tingvall C. Clinical symptoms of sleep apnea syndrome and automobile accidents. ORL; Journal of Oto-Rhino-Laryngology & Its Related Specialties 52:57-62, 1990.

Haraldsson PO, Carenfelt C, Lysdahl M, Tingvall C. Does uvulopalatopharyngoplasty inhibit automobile accidents? Laryngoscope 105:657-661, 1995a.

Haraldsson PO, Carenfelt C, Lysdahl M, Tornros J. Long-term effect of uvulopalatopharyngoplasty on driving performance. Arch Otol Head Neck Surg 121:90-94, 1995b.

Harrell FE, Lee KI, Mark DB. Tutorial in biostatistics: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat in Med 15:361-387, 1996.

Hermansen K. Diet, blood pressure and hypertension. Br J Nutr 83:S113-S119, 2000.

Hochban W, Conradt R, Brandenburg U, Heitmann J, Peter JH. Surgical maxillofacial treatment of obstructive sleep apnea. Plast Reconstr Surg 99:619-628, 1997.

Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. Quantification of sleepiness: a new approach. Psychophysiology 10:431-436, 1973.

Howard M, Worsnop C, Campbell D, Swann P, Pierce R. Obstructive sleep apnoea and excessive sleepiness in Australian road transport drivers. Presentation and proceedings paper at the 4th International Conference on Fatigue in Transportation: Coping with the 24 Hours Society, Fatigue Management Alternatives to Prescriptive Hours of Service. Fremantle, Western Australia, March 2000.

Hoy CJ, Vennelle M, Kingshott RN, Engleman HM, Douglas NJ. Can intensive support improve continuous positive airway pressure use in patients with the sleep apnea/hypopnea syndrome? Am J Respir Crit Care Med 159:1096-1100, 1999.

Hung J, Whitford EG, Parsons RW, Hillman DR. Association of sleep apnoea with myocardial infarction in men. Lancet 336:261-264, 1990.

Issa FG, Sullivan CE. Alcohol, snoring and sleep apnoea. J Neurol Neurosurg Psychiatry 45:353-359, 1982.

Jarjour NN, Wilson P. Pneumocephalus associated with nasal continuous positive airway pressure in a patient with sleep apnea syndrome. Chest 96:1425-1426, 1989.

Jean-Louis G, von Gizycki H, Zizi F, Spielman A, Hauri P, Taub H. The actigraph data analysis software: II. A novel approach to scoring and interpreting sleep-wake activity. Perceptual and Motor Skills 85:219-226, 1997.

Jenkinson C, Davies RJ, Mullins R, Stradling JR. Comparison of therapeutic and subtherapeutic nasal continuous positive airway pressure for obstructive sleep apnoea: a randomised prospective parallel trial. Lancet 353:2100-2105, 1999.

Jennum P, Sjøl A. Snoring, sleep apnoea and cardiovascular risk factors: the MONICA II study. Int J Epidemiol 22:439-444, 1993.

Johns MW. A new method for measuring daytime sleepiness: The Epworth Sleepiness Scale. Sleep 14:540-545, 1991.

Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 15:376-381, 1992.

Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea - The Epworth Sleepiness Scale. Chest 103:30-36, 1993.

Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. Sleep 17:703-710, 1994.

Kannel WB. Fifty years of Framingham Study contributions to understanding hypertension. J Hum Hypertension 14:83-90 2000.

Kopelman PG. Obesity as a medical problem. Nature 404:635-643, 2000.

Koskenvuo M, Partinen M, Kaprio J, Vuorinen H, Telakivi T, Kajaste S, Salmi T, Heikkila K. Snoring and cardiovascular risk factors. Ann Med 26:371-376, 1994.

Kribbs NB, Pack AI, Kline LR, Smith PL, Schwartz AR, Schubert NM, Redline S, Henry JN, Getsy JE, Dinges DF. Objective measurement of patterns of nasal CPAP use by patients with obstructive sleep apnea. Am Rev Respir Dis 147:887-895, 1993a.

Kribbs NB, Pack AI, Kline LR, Getsy JE, Schuett JS, Henry JN, Maislin G, Dinges DF. Effects of one night without nasal CPAP treatment on sleep and sleepiness in patients with obstructive sleep apnea. Am Rev Respir Dis 147:1162-1168, 1993b.

Kribbs NB, Dinges DF. Vigilance decrement and sleepiness. In: Harsh JR and Ogilvie RD (eds), *Sleep Onset Mechanisms*. American Psychological Association, Washington DC, pp 113-125, 1994.

Krieger J, Kurtz D: Objective measurement of compliance with nasal CPAP treatment for obstructive sleep apnoea syndrome. Eur Respir J 1:436-438, 1988.

Kripke DF, Ancoli-Israel S, Klauber MR, Wingard DL, Mason WJ, Mullaney DJ. Prevalence of sleep-disordered breathing in ages 40-64 years: A population-based survey. Sleep 20:65-76, 1997.

Landis JR, Coch GG. The measurement of observer agreement for categorical data. Biometrics 33:159-174, 1997.

Lavie P. Incidence of sleep apnea in a presumably healthy working population: a significant relationship with excessive daytime sleepiness. Sleep 6:312-318, 1983.

Lindberg E, Janson C, Gislason T, Björnsson E, Hetta J, Boman G. Sleep disturbances in a young adult population: can gender differences be explained by differences in psychological status? Sleep 20:381-387, 1997.

Lindberg E, Elmasry A, Gislason T, Janson C, Bengtsson H, Hetta J, Nettelbladt M, Boman G. Evolution of sleep apnea syndrome in sleepy snorers - a population-based prospective study. Am J Respir Crit Care Med 159:2024-2027, 1999.

Lowe AA, Santamaria JD, Fleetham JA, Price C. Facial morphology and obstructive sleep apnea. Am J Orthod Dentofac Orthop 90:484-491, 1986.

Lowe AA, Fleetham JA, Adachi S, Ryan CF. Cephalometric and computed tomographic predictors of obstructive sleep apnea severity. Am J Orthod Dentofac Orthoped 107:589-595, 1995.

Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, Schwab RJ, Dinges DF. A survey screen for prediction of apnea. Sleep 18:158-166, 1995.

Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22:719-748, 1959.

Marklund M, Franklin KA, Sahlin C, Lundgren R. The effect of a mandibular advancement device on apneas and sleep in patients with obstructive sleep apnea. Chest 113:707-713, 1998.

McNicholas WT. Sleep apnoea and driving risk. Eur Respir J 13:1225-1227, 1999.

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 50:284-289, 1995.

Mitler MM, Miller JC, Lipsitz JJ, Walsh JK, Wylie CD. The sleep of long-haul truck drivers. N Engl J Med 337:755-761, 1997.

Moskowitz H, Burns M. The effects of alcohol and valium, singly and in combination, upon driving-related skills performance. Proceedigns of the 21st Conference of the Association for Automotive Medicine. Vancouver, BC, September 15-17, pp 226-240, 1997.

National Task Force on the Prevention and Treatment of Obesity. Overweight, obesity and health risk. Arch Intern Med 160:898-904, 2000.

National Transportation Safety Board. Factors that affect fatigue in heavy truck accidents. Volume 1: Analysis safety study. NTSB/SS-95-01. National Transportation Safety Board: Washington DC, 1995.

Nelson S, Hans M. Contribution of craniofacial risk factors in increasing apneic activity among obese and nonobese habitual snorers. Chest 111:154-162, 1997.

Neven AK, Middelkoop HA, Kemp B, Kamphuisen HA, Springer MP. The prevalence of clinically significant sleep apnoea syndrome in The Netherlands. Thorax 53:638-642, 1998.

Nieto FJ, Young TB, Lind BK, Shahar E, Samet JM, Redline S, D'Agostino RB, Newman AB, Lebowitz MD, Pickering TG, for the Sleep Heart Health Study. Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study. JAMA 283:1829-1836, 2000.

Nino-Murcia G, McCann CC, Bliwise DL, Guilleminault C, Dement WC. Compliance and side effects in sleep apnea patients treated with nasal continuous positive airway pressure. West J Med 150:165-169, 1989.

Pack AI. Obstructive sleep apnea. Adv Intern Med 39:517-567, 1994.

Partinen M, Guilleminault C, Quera-Salva M, Jamieson A. Obstructive sleep apnea and cephalometric roentgenograms. The role of anatomic upper airway abnormalities in the definition of abnormal breathing during sleep. Chest 93:1199-1205, 1988.

Pendlebury ST, Pépin J-L, Veale D, Lévy P. Natural evolution of moderate sleep apnoea syndrome: significant progression over a mean of 17 months. Thorax 52:872-878, 1997.

Pennsylvania Department of Health. 1998 Behavioral Health Risks of Pennsylvania Adults.

Peppard PE, Young T, Palta M, Skatrud J. Prospective study of the association between sleep-disordered breathing and hypertension. N Engl J Med 342:1378-1384, 2000.

Phoha RL, Dickel MJ, Mosko SS. Preliminary longitudinal assessment of sleep in the elderly. Sleep 13:425-429, 1990.

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. The Laryngoscope 109:1648-1654, 1999.

Rather RJ, Warriner S, Johanson D. From beltline to steering wheel: the vanishing space. South Dakota Med J 24:21-24, 1981.

Rechstaffen A, Kales A. A Manual of Standardized Terminology: Techniques of Scoring System for Sleep Stages of Human Subjects. Brain Information Services/Brain Research Institute, University of California, Los Angeles, 1968.

Redline S, Kump K, Tishler PV, Browner I, Ferrette V. Gender differences in sleep disordered breathing in a community-based sample. Am J Respir Crit Care Med 149:722-726, 1994.

Redline S, Sanders M. Hypopnea, a floating metric: implications for prevalence, morbidity estimates, and case finding. Sleep 20:1209-1217, 1997a.

Redline S, Tischler PV, Hans MG, Tosteson TD, Strohl KP, Spry K. Racial differences in sleep-disordered breathing in African-Americans and Caucasians. Am J Respir Crit Care Med 155:186-192, 1997b.

Redline S, Kapur VK, Sanders MH, Quan SF, Gottlieb DJ, Rapoport DM, Bonekat WH, Smith PL, Kiley, JP, Iber C. Effects of varying approaches for identifying respiratory disturbances on sleep apnea assessment. Am J Respir Crit Care Med 161:369-374, 2000.

Redmond DP, Hegge FW. Observations on the design and specification of a wrist-worn human activity monitoring system. Behav Res Meth Instru Computers 17:659-669, 1985.

Reeves-Hoche MK, Hudgel DW, Meck R, Witteman R, Ross A, Zwillich CW. Continuous versus bilevel positive airway pressure for obstructive sleep apnea. Am J Respir Crit Care Med 151:443-449, 1995.

Riley RW, Powell NB, Guilleminault C. Obstructive sleep apnea syndrome: a review of 306 consecutively treated surgical patients. Otolaryngol Head Neck Surg 108:117-125, 1993.

Rolfe I, Olson LG, Saunders NA. Long-term acceptance of continuous positive airway pressure in obstructive sleep apnea. Am Rev Respir Dis 144:1130-1133, 1991.

Rosekind MR, Graeber RC, Dinges DF, et al. Crew factors in flight operations. IX. Effects of cockpit rest on crew performance and alertness in long-haul operations. NASA Technical Memorandum Rep. No. 103884, 252 pp, 1994.

Rosenbaum PR. Discussing hidden bias in observational studies. Ann Intern Med 115:901-905, 1991.

Rosenbaum PR. Observational studies. Springer Series in Statitistics. New York: Springer-Verlag, pp 87-135, 1995.

Roth T, Roehrs T, Zwyghuizen-Doorenbos A, Stepanski E, Wittig R. Sleep and memory. Psychopharmacology Series 6:140-145, 1988.

Rothman KJ, Greenland S, Walker A. Concepts of interaction. Am J Epidemiology 112:467-470, 1980.

Sakurai Y. Duration of obesity and risk of non-insulin-dependent diabetes mellitus. Biomed Pharmacotherapy 54:80-84, 2000.

Samuel S, Pack FM, Pack AI, Maislin G, Winokur A, Dinges DF. Sleep-wake behaviors of elderly living in residential communities: results from a case-control study of excessive daytime sleepiness. Sleep Res 25:131, 1996.

Schmidt-Nowara WW, Coultas DB, Wiggins C, Skipper BE, Samet JM. Snoring in a Hispanic-American population. Risk factors and association with hypertension and other morbidity. Arch Intern Med 150:597-601, 1990a.

Schmidt-Nowara JR, Jennum P. Epidemiology of sleep apnea. In: Guilleminault C, Partinen M, ed. *Obstructive Sleep Apnea Syndrome—Clinical Research and Treatment*. New York: Raven Press, pp 1-8, 1990.

Schmidt-Nowara W, Lowe A, Wiegand L, Cartwright R, Perez-Guerra F, Menn S. Oral appliances for the treatment of snoring and obstructive sleep apnea: a review. Sleep 18:501-510, 1995.

Schwab RJ, Gupta KB, Gefter WB, Hoffman EA, Pack AI. Upper airway soft tissue anatomy in normals and patients with sleep disordered breathing. Significance of the lateral pharyngeal walls. Am J Respir Crit Care Med 152:1673-1689, 1995.

Seidel JC. Obesity, insulin resistance and diabetes—a worldwide epidemic. Brit J Nutr 83:S5-S8, 2000.

Sforza E, Addati G, Cirignotta F, Lugaresi E. Natural evolution of sleep apnoea syndrome: a five year longitudinal study. Eur Respir J 7:1765-1770, 1994.

Sheehan MT, Jensen MD. Metabolic complications of obesity. Pathophysiologic considerations. Med Clin N A 84:363-385, 2000.

Sher AE, Schectman KB, Piccirillo JF. The efficacy of surgical modifications of the upper airway in adults with obstructive sleep apnea syndrome. Sleep 19:156-177, 1996.

Smiley A. Effects of minor tranquilizers and antidepressants on psychomotor performance. J Clin Psychiatry 48 (Suppl):22-28, 1987.

Stoohs RA, Dement WC. Snoring and sleep-related breathing abnormality during partial sleep deprivation. N Engl J Med 328:1279, 1993.

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 17:619-623, 1994.

Stoohs RA, Bingham LA, Itoi A, Guilleminault C, Dement WC. Sleep and sleep-disordered breathing in commercial long-haul truck drivers. Chest 107:1275-1282, 1995.

Stradling JR, Crosby JH. Predictors and prevalence of obstructive sleep apnoea and snoring in 1001 middle aged men. Thorax 46:85-90, 1991.

Stutts JC, Wilkins JW, Vaughn BV. Why Do People Have Drowsy Driving Crashes? Input From Drivers Who Just Did. AAA Foundation for Traffic Safety, 1999.

Sullivan CE, Issa FG, Berthon-Jones M, Eves, L. Reversal of obstructive sleep apnoea by continuous positive airway pressure applied through the nares. Lancet 1:862-865, 1981.

Svanborg E, Larsson H. Development of nocturnal respiratory disturbance in untreated patients with obstructive sleep apnea syndrome. Chest 104:340-343, 1993.

Terán-Santos J, Jiménez-Gòmez A, Cordero-Guevara J. The association between sleep apnea and the risk of traffic accidents. N Engl J Med 340:847-851, 1999.

Van den Hoed J, Hraemer H, Guilleminault C, Zarcone VP, Miles LE, Dement WC, Mitler MM. Disorders of excessive daytime somnolence: polygraphic and clinical data for 100 patients. Sleep 4:23-37, 1981.

Waite PD, Wooten V, Lachner J, Guyette RF. Maxillomandibular advancement surgery in 23 patients with obstructive sleep apnea syndrome. J Oral Maxillofac Surg 47:1256-1261, 1989.

Waldhorn RE, Herrick TW, Nguyen MC, O'Donnell AE, Sodero J, Potolicchio SJ. Long-term compliance with nasal continuous positive airway pressure therapy of obstructive sleep apnea. Chest 97:33-38, 1990.

Weaver TE, Laizner AM, Evans LK, Maislin G, Chugh KD, Lyon K, Smith PL, Schwartz AR, Redline S, Pack AI, Dinges DF. An instrument to measure functional status outcomes for disorders of excessive sleepiness. Sleep 20:835-843, 1997.

Wessendorf TE, Teschler H, Wang YM, Konietzko N, Thilmann AF. Sleep-disordered breathing among patients with first-ever stroke. J Neurol 247:41-47, 2000.

Worsnop CJ, Naughton MT, Barter CE, Morgan TO, Anderson AI, Pierce RJ. The prevalence of obstructive sleep apnea in hypertensives. Am J Respir Crit Care Med 157:111-115, 1998.

Wu H, Yan-Go F. Self-reported automobile accidents involving patients with obstructive sleep apnea. Neurology 46:1254-1257, 1996.

Wyatt JK, Dijk D-J, Ronda JM, Jewett ME, Powell JW, Dinges DF, Czeisler CA. Interaction of circadian- and sleep/wake homeostatic-processes modulate psychomotor vigilance test (PVT) performance. Sleep Research 26:759, 1997.

Yamamoto H, Akashiba T, Kosaka N, Ito D, Horie T. Long-term effects of nasal continuous positive airway pressure on daytime sleepiness, mood and traffic accidents in patients with obstructive sleep apnoea. Respir Med 94:87-90, 2000.

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 328:1230-1235, 1993.

Young T, Peppard P, Palta M, Hla KM, Finn L, Morgan B, Skatrud J. Population-based study of sleep-disordered breathing as a risk factor for hypertension. Arch Intern Med 157:1746-1752, 1997a.

Young T, Blustein J, Finn L, Palta M. Sleep-disordered breathing and motor vehicle accidents in a population-based sample of employed adults. Sleep 20:607-613, 1997b.

Young T, University of Wisconsin-Madison Medical School. Written report to Mr. Albert Alvarez, Office of Motor Carrier Research and Standards, March 29, 1999 included in Review of the Federal Highway Administration's Study of Sleep Apnea Among Operators of Commercial Motor Vehicles, 1999.

Appendix A:

List Of Abbreviations Used In Report

Abbreviation	Meaning	Note
AHI	Apnea-hypopnea index	Number of apneas plus hypopneas per hour of sleep. This is measure of severity of sleep-disordered breathing
ATA	American Trucking Associations	National association of owners of trucking companies. The Association has active research programs in this area
AUC	Area under the curve	Measured used in analyses using Receiver Operator Curve as a measure of goodness of fit. 1.0 represents optimal fit
BMI	Body mass index	A measure of degree of obesity when weight is corrected for height
CDL	Commercial driver license	
CI	Confidence interval	Used in statistics to give measure of accuracy of an estimate of a particular parameter
DADT	Divided attention driving task	A task that has been developed to simulate and assess the cognitive performance related to the driving task
DSST	Digital symbol substitution test	A test used to assess what is known as cognitive throughput
EEG	Electroencephalogram	Technique used to measure small electrical signals generated by the brain. It is used to help identify the different stages of sleep
EMG	Electromyogram	Technique used to measure electrical activity in muscle. Used with EEG to help identify different stages of sleep

EOG	Electro-oculogram	Technique used to measure eye movements by recording electrical activity of muscles moving the eyes. This is particularly valuable for identifying rapid-eye-movement sleep
ESS	Epworth sleepiness scale/score	A questionnaire that is used to assess degree of self-reported excessive daytime sleepiness
FOSQ	Functional Outcomes of Sleepiness Questionnaire	Questionnaire used to assess impact of excessive sleepiness on activities of daily living (quality of life)
ICC	Intra-class correlation coefficient	A statistical measure used to assess, for example, reliability of a particular measurement
KSS	Karolinska Sleepiness Scale	A questionnaire used to measure how sleepy an individual is at a given moment in time
MAP	Multivariable apnea prediction	A formula used to assess likelihood of an individual having obstructive sleep apnea. This scale is from zero to one
MSLT	Multiple sleep latency test	A test used to assess the level of physiological sleepiness, i.e., pressure for sleep, in an individual
N	Number	Typically used for the number of subjects in a study, or the number of measurements used in a particular statistical analysis
NAB	Neurobehavioral assessment battery	This is a battery of tests used to assess when the individual's performance is affected by sleepiness and/or sleep deprivation
OR	Odds ratio	A statistical term indicating increased relative risk
PSG	Polysomnogram	This is an overnight sleep study during which a large number of variables, including sleep state (EEG, EOG, EMG), respiration and oxygen level are measured

PVT	Psychomotor vigilance reaction time task	A reaction time task that is used to assess performance impairments that relate to excessive sleepiness
QoL	Quality of life	A generic term referring to instruments that are used to assess quality of life
RDI	Respiratory disturbance index	This is the same as AHI, i.e., number of apneas plus hypopneas per hour of sleep
RERA	Respiratory effect related arousal	This describes a specific event that can occur during sleep. It is a when an arousal (sleep interruption) occurs as a result of an individual partially obstructing their airway and having to make an effort to breathe
ROC	Receiver operator curve	A specific methodology used to assess, for example, the accuracy of a diagnostic test
SaO$_2$	Oxygen saturation	The level of oxygen in the arterial blood
SDB	Sleep-disordered breathing	A generic term used to define the presence of breathing abnormalities during sleep
SSS	Stanford Sleepiness Scale	A questionnaire used to define how sleepy an individual is at a moment in time

Appendix B:

Summary Of Previous Studies On

Sleep Apnea

Reference Study Population	Sample Size (n) Sex and criteria	Number is AI or AHI >5	Estimated OSAS Prevalence (%)
Lavie P [1983] Industrial workers (n=1262) in Israel	n=78 males, selected on the basis of questionnaires	17 M	0.7
Franceschi et al [1982] All patients (n=2518) admitted during 1 year to a general hospital	n=87, selected on the basis of questionnaires and clinical data	7 M, 18 F	1.0
Gislason et al [1988] Random sample of 3201 males aged 30-69 years in Uppsala	n=61 males, selected because of habitual snoring and daytime sleepiness	15 M	2.7
Cirignotta et al [1989] Males 30-69 years old in Bologna	n=40, snored every night	21 M	2.7
Schmidt-Nowara et al [1990] 1195 males and females who had answered a questionnaire	n=32, with ≥20 apneas/h in a screening investigation		M: 2.3 F: 1.1
Stradling et al [1991] Random sample of 1001 males aged 35-65 years in Oxford	n=21 males with >5 SaO$_2$ dips/h in a screening investigation	3	0.3
Gislason et al [1993] Random sample of 1505 females aged 40-59 years in Reykjavik	n=35 females selected because of habitual snoring and daytime sleepiness	15 F	2.5
Young et al [1993] Random sample of 3513 state employees 30-60 years of age in Wisconsin	n=602 (250 F and 353 M) habitual snorers (n=335) and a random sample of not habitual snorers (n=247)	97 M, 31 F	M: 4.0 F: 2.0

Neven et al [1998] General population: 1572 males ≥35 years; 894 females ≥50 years	home thermistry: 169 M and 25 F polysomnography: 24 M and 1 F	14 M, 1 F	0.9
Bixler et al [1998] n=4364 males selected by telephone survey	M 20-44 years = 236 M 45-64 years = 430 M 65-100 years = 75		1.2 4.7 1.7

Appendix C:

Sleep Apnea Risk Of Car Crashes

SLEEP APNEA AND RISK OF CAR CRASHES: SUMMARY OF EPIDEMIOLOGICAL EVIDENCE
(Modified from Conner et al [2001])

Study	Participants	Exposure	Confounders considered	Crash outcomes	Results	Comments
Aldrich [1989], US, cross-sectional	424 Consecutive adult sleep clinic patients, 279 men, 145 women, non-drivers excluded, 70 "healthy" adult controls	Type of sleep disorder, by polysomnography in clinic patients only. Questionnaire in al participants	Gender (analyzed separately), approximate age matching	Self-reported number of automobile accidents ever	No patient group had a higher rate of crashes (from any cause) than controls	No polysomnography in controls, uncontrolled confounding by age, mileage, drug and alcohol use, outcome measured over variable period, no effect estimates reported
Alpert et al [1992], US, cross-sectional	200 Men >65 years of age recruited for a previous sleep study from inpatients of a VA hospital between 1985-1990	Respiratory disturbance index (RDI), from sleep recording, obtained from clinic records	Gender (restricted), age (restricted)	Number of accidents as driver in past 5 years, from official records of motor vehicle moving violations	23% of patients with RDI ≥10 had at least one accident compared with 11% with RDI <10 (P=0.02), RR[a] for crash (95% CI) for RDI >10 compared with RDI <10 = 2.1 (1.1-4.1)	Selection criteria not given, restricted to patients with driving records. Possible biased underreporting of crashes. Uncontrolled confounding by mileage, alcohol and drug use. No effect estimates reported.

Study	Participants	Exposure	Confounders considered	Crash outcomes	Results	Comments
Barbe et al [1998], Spain, cross-sectional	60 Sleep clinic patients, 60 "healthy" controls (hospital workers). Exclusion for: no drivers license, non-residents, drug abuse, shift-work, psychiatric disorders, epilepsy, narcolepsy and periodic limb movements (1/61 cases declined to participate	Sleep apnea status; by polysomnography in patients and by clinical history in controls (confirmatory polysomnography in two). Epworth Score for usual daytime sleepiness	Gender (matched), age (matched), mileage (adjusted), alcohol (stratified), drug use	Number of automobile accidents in past 3 years, from self report and insurance records	OR for at least one crash (95% CI) for apneics compared with controls = 2.3 (0.97-5.33). Adjusted for mileage OR = 2.6 (1.06-6.43). OR for more than one crash = 5.2 (1.07-25.29). Other potential confounders reported not to have affected the estimates. No association between Epworth Score and risk of crash in apnea patients	Response rate high (one refusal). Control group may not be representative of driving population. All major confounders considered. Association between Epworth Score and risk of crash not examined in controls
Bearpark [1990], Australia, cross-sectional	135 Male sleep clinic patients, 289 male volunteer controls from an Easter Show	Apnea Index (AI), by polysomnography in patients. BMI and 7-item mini sleep questionnaire (MSQ) in all groups	Gender (restricted), age (matched), mileage: no difference between groups	Self reported accidents in the last 2 years (as driver)	Number of crashes in last 2 years, controls 0.4, snorers 0.3, apneics 0.3. OR* for crash (95% CI) compared with controls for snorers = 0.62 (0.27-1.42), for apneics = 0.63 (0.38-1.05)	No response rates reported, no polysomnography in controls, uncontrolled confounding by drug and alcohol use (controls drank less than patients, snorers used more sleeping pills than controls). No effect estimates reported

Study	Participants	Exposure	Confounders considered	Crash outcomes	Results	Comments
Findley et al [1988], US, cross-sectional	64 Selected from 77 sleep clinic patients. Also compared with all drivers in Virginia (records) n = 3.7 million	Sleep apnea status, by polysomnography in patients. No exposure measurement in general population	None	Crash per driver in 5 year period from state driving records	% Subjects having a crash, apneic patients: 31%, non-apneics: 6% (P=0.01). Mean number of crashes: apneic patients 0.41, non-apneics: 0.06 (P<0.01), Virginia drivers: 0.16 (P<0.02). RR[a] for crash (95% CI) for apneics compared with non-apneics = 7.2 (1.8-30). RR[a] for crash (95% CI) for apneics compared to Virginia drivers = 2.6	Response rate not reported, selection criteria not stated, restricted to patients with driving records (90% of apneics, 76% of non-apneics). Possible biased underreporting of crashes. Uncontrolled confounding by age, gender, mileage, alcohol and drug use. Very small sample. No effect estimates reported. Non-apneic group not representative controls (much lower crash rate than all Virginia drivers)
Findley et al [1989], US, cross-sectional	46 Patients diagnosed with OSA at a sleep clinic. Controls: all licensed drivers in Virginia	Sleep apnea status, by polysomnography in patients. No exposure measurement in comparison group	None	Crashes per driver in 5 year period from state driving records	Mean number of crashes; mild apnea: 0.13, moderate apnea: 0.24; severe apnea: 0.46, Virginia drivers: 0.16. RR[a] for crash (95% CI) for severe apneics compared with Virginia drivers = 3.4 (1.9-6)	Relationship to sample in previous study (Findley et al [ref]) not clear. Same comments apply.

Study	Sample	Exposure measurement	Confounders controlled	Outcome measurement	Results	Comments
Findley et al [1995], US, cross-sectional	62 Sleep apnea patients. 12 age- and sex-matched sleep clinic patients. 10 age- and sex-matched volunteers

10 Narcolepsy patients. 10 age- and sex-matched patients, 10 age- and sex-matched volunteers | A severity of sleep apnea by polysomnography

Presence of narcolepsy by multiple sleep latency test. No sleep testing for volunteer controls | Age (matching), gender (matching) | "Steer clear" computerized driving vigilance test. Crashes per driver in 5 year period from state driving records | Simulator performance negatively associated with presence and severity of sleep apnea (P<0.05). Simulator performance negatively associated with presence (but not severity) of narcolepsy (P<0.05). Crash rates in performance groups: normal performance, 0.05; poor performance, 0.20; very poor performance, 0.38 | No analysis of crash rates by diagnostic group. Uncontrolled confounding by mileage and drug and alcohol use. Possible biased underreporting of crashes |
| George et al [1987], Canada, cross-sectional | 27 Male sleep clinic patients, 270 male controls (10 age-matched controls for each patient) from driving records | Presence of obstructive sleep apnea in patients by polysomnography (20), or clinical diagnosis (7). No exposure measurement in controls | Age (matching), gender (restriction) | Number of car accidents as driver, ever, from state records | % Subjects ever had a crash: apneics, 97%; non-apneics, 54% (P<0.01); mean number of crashes, apneics, 2.63; non-apneics, 1.28. OR[a] (95% CI) for at least one crash in apneics compared with non-apneics = 10.9 (2.4-68) | Very small exposed sample, selection criteria not stated, possible biased underreporting of crashes. Uncontrolled confounding by mileage and drug and alcohol use, outcome measured over variable period, no effect estimates reported |

Study	Participants	Exposure	Confounders considered	Crash outcomes	Results	Comments
Gonzalez-Rothi et al [1988], US, cross-sectional	126 of 140 consecutive sleep clinic patients	Sleep apnea status, by polysomnography from record review	None	Self-reported vehicular mishaps, including "near miss", actual crashes, or subjects who no longer operated a motor vehicle for fear of falling asleep	Treated apnea group: 34% (19 events), untreated apnea group: 27% (6 events), control group: 7% (2 events)	Selection criteria unclear. Uncontrolled confounding by age, gender, mileage and alcohol and drug use. Mean age and gender distribution varied between groups. Objective and subjective outcome measures grouped together, and not analyzed separately. No effect estimates reported or able to be calculated from data
Haraldsson et al [1990], Sweden, cross-sectional	282 Male ENT-department patients who were regular car drivers, 140 with sleep apnea symptoms (response rate 97%), 142 with nasal obstruction (response rate 89%)	Clinical diagnosis of sleep apnea (n=73), habitual snoring (n=67) or no sleep disorder (n=142), by self report	Age, gender (restricted), mileage	Self-reported single-car and combined-car accidents in the past 5 years	Odds Ratio for involvement in crashes compared with controls, adjusted for mileage. All crashes: snorers OR=1.4 (P>0.05); apneics OR=1.5 (P<0.05). Single vehicle crashes: snorers OR=1.2 (P>0.05); apneics (OR=6.8 (P<0.05).	Uncontrolled confounding by alcohol, age and drug use.

Study	Participants	Exposure	Confounders considered	Crash outcomes	Results	Comments
Terán-Santos et al [1999], Spain, case-control	102 Drivers treated at Emergency Departments (response rate 71%), 152 age- and sex-matched controls from primary healthcare centers (response rate 89%)	Sleep apnea status by respiratory polysomnography. Epworth Sleepiness Scale, symptoms of sleep apnea by questionnaire	Age, gender (matched), illicit drug use (restricted), alcohol, mileage	Crash resulting in treatment of driver at Emergency Department	Adjusted odds ratio for crash with sleep apnea (apnea-hypopnea index >10) 7.2 (2.4-21.8)	Sensitivity analysis for non-responders supported positive association. Control group may not be representative of driving population. Exclusion from cases of most severely injured and drivers of cars where only passengers were injured.
Wu and Yan-Go [1996], US, cross-sectional	253 Sleep clinic patients, with a driver's license. Response rate 86%	Sleep apnea status on polysomnography, falling asleep at inappropriate times by self report	Age, gender, alcohol use	Self reported car crashes	Adjusted OR (95% CI) for association of crashes with: sleep apnea 2.58 (1.06-6.31); falling asleep at inappropriate times 5.72 (2.39-9.21)	Uncontrolled confounding by mileage, outcome measured over variable period, outcome measure not clearly defined. References to "driving accidents or near accidents due to sleepiness", "having an accident," "self reported MVAs"

Study	Participants	Exposure	Confounders considered	Crash outcomes	Results	Comments
Young et al [1997], US, cross-sectional	913 Employed adults enrolled in an ongoing study of the natural history of sleep disordered breathing. Licensed driving >1000 miles per year	Sleep disordered breathing (SDB) status measured by polysomnography and self-reported snoring frequency	Age, gender, mileage, alcohol	Crash history over 5 year study period, by record matching with motor vehicle accident data from Wisconsin State records	Adjusted OR (95% CI) for at least one crash in men: no SDB 1.0 (referent); snorers 3.4 (1.8-6.9); mild SDB 4.2 (1.6-11.3); severe SDB 3.4 (1.4-8.0). In women: snorers 0.9 (0.5-1.6); mild SDB 0.8 (0.3-2.0); severe SDB 0.6 (0.2-2.5)	Response rates not stated. From a previous description [Young et al, 1993]. 82% responded to initial questionnaire and 43% responded to recruitment for polysomnography. Outcome measure may be affected by biased underreporting. Controlled for all major confounders. Small number of events in subgroups, resulting in poor precision

[a]Estimate calculated from data present. RR, relative risk; OR, odds ratio; CI, confidence interval; AI, apnea index, SDB, sleep disordered breathing; RDI, respiratory disturbance index.

Appendix D:

Paper Describing Multivariable Apnea

Prediction

A Survey Screen for Prediction of Apnea

*Greg Maislin, *Allan I. Pack, *Nancy B. Kribbs, †Philip L. Smith,
†Alan R. Schwartz, ‡Lewis R. Kline, *Richard J. Schwab
and *David F. Dinges

*Center for Sleep and Respiratory Neurobiology, University of Pennsylvania School of Medicine,
Philadelphia, Pennsylvania, U.S.A.;
†Johns Hopkins Center for Sleep Disorders, Baltimore, Maryland, U.S.A.; and
‡West Penn Center for Sleep Disorders, Pittsburgh, Pennsylvania, U.S.A.

Summary: Questionnaire data from patients presenting at three sleep disorders centers were used to develop and assess a screening tool for sleep apnea based on the reporting of the frequency of various symptoms of sleep apnea and other sleep disorders plus age, body mass index (BMI) and gender. Patients were not specifically referred for suspicion of sleep apnea. Separate factor analyses of survey responses from 658, 193 and 77 respondents from the first, second and third sites, respectively, each yielded four orthogonal factors, one of which accounted for all the questions concerned with the frequency of disordered breathing during sleep. The survey was shown to be reliable in a subset of patients from one of the sites (test–retest correlation = 0.92). Survey data were then compared to a clinical measure of sleep apnea (respiratory disturbance index) obtained from polysomnography. A multivariable apnea risk index including survey responses, age, gender and BMI was estimated using multiple logistic regression in a total sample of 427 respondents from two of the sites. Predictive ability was assessed using receiver operating characteristic (ROC) curves. The area under the ROC curve was 0.79 (p < 0.0001). For BMI alone, it was 0.73, and for an index measuring the self-report of the frequency of apnea symptoms, it was 0.70. The multivariable apnea risk index has potential utility in clinical settings. **Key Words:** Sleep apnea syndrome—Prediction—Survey screen.

A low-cost reliable method for discriminating between patients likely and not likely to have sleep apnea syndrome among those presenting at sleep disorders centers and in other settings would be useful for efficiently targeting specific persons for more in-depth clinical evaluation. Studies (1–4) directed at predicting the diagnosis of sleep apnea syndrome utilizing the self-reporting of apnea symptoms (e.g. snoring, breathing cessation, choking, gasping) and demographic information [e.g. body mass index (BMI), age and gender] have assessed the predictive utilities of derived decision rules. This study extends these findings using a large multisite sample of patients presenting at sleep disorders centers, who were not specifically referred for suspicion of sleep apnea. First, we developed and administered a survey. A confirmatory factor analysis was used to examine the validity of an index derived from respondents' self-reports of the frequency of loud snoring, breathing cessation, and snorting and gasping. Reliability was assessed by having a subset of patients fill out another questionnaire approximately 2 weeks later and then comparing the results using test–retest correlations. Internal consistency of the items in the questionnaire was assessed using Cronbach alpha statistics. Then, a multivariable apnea risk index was estimated using logistic regression models that included the value of the index, age, gender and BMI. A comparison among the predictive powers of the multivariable apnea risk index, the apnea symptom frequency index alone and BMI alone was completed using receiver operating characteristic (ROC) curves. Procedures for using the multivariable risk index in clinical settings are discussed.

METHODS

Subjects

Survey responses were obtained from 1,071 patients. This included 770 consecutive patients seen at the Penn

Center for Sleep Disorders in Philadelphia, a sample of 225 patients seen at the Johns Hopkins Center for Sleep Disorders in Baltimore and a sample of 85 patients seen at the West Penn Sleep Center in Pittsburgh. Other information was collected on subsets of these patients.

Subject enrollment form

All patients filled out a subject enrollment form that included demographic information such as date of birth, gender, race, marital status, presence of regular bedpartner/roommate, chief complaint, education level, work status and history of shift work. Of primary interest were 13 self-report symptom frequency questions as follows: "During the last month, have you had, or have been told about the following symptom (Show the frequency): (0) Never; (1) Rarely, Less Than Once a Week; (2) 1–2 Times Per Week; (3) 3–4 Times Per Week; (4) 5–7 Times Per Week; (.) Don't Know". The questions were chosen from four sets of symptoms thought to be associated with sleep disorders. These were sleep apnea-related symptoms, difficulty sleeping symptoms, excessive daytime sleepiness symptoms and narcolepsylike symptoms. In addition, other questions were asked about medications, weight changes, smoking and drinking habits, menstruation history and four ordinal questions regarding perceived quality of life. The subject enrollment form was sent out to patients prior to their first visit at the Philadelphia site. Patients filled out the forms at home and brought in the completed forms. At the Baltimore and Pittsburgh sites, subjects filled out the questionnaire at the time of their first visit to the clinical facility.

Physician enrollment form

This form was completed by the examining physician. It included data from a physical assessment (blood pressure, weight, height, presence of naso-oropharyngeal abnormalities, cardiovascular abnormalities, chest sounds) and medical history of various previous illnesses. BMIs were computed from these data as weight divided by height squared (kg/m²). Completed physician enrollment forms were available for 657 patients (519 patients from the Philadelphia site and 138 from the Baltimore site).

Sleep study form

This form included the results from an overnight polysomnographic study. The respiratory disturbance index (RDI) was calculated as the number of apneas plus hypopneas per hour of sleep. Other variables collected included O₂ nadirs in rapid eye movement (REM)

and nonrapid eye movement (NREM) sleep; sleep latencies and proportions of total sleep in stages 1, 2, 3/4 and REM sleep; arousals in each stage; and frequency of periodic leg movements. Scoring was performed using the method of Rechtschaffen and Kales (5). Completed sleep study forms were available for 977 patients (686 patients from Philadelphia, 208 from Baltimore and 83 from Pittsburgh).

Statistical methods

Confirmatory factor analysis, test–retest correlation coefficients (6) and Cronbach alpha statistics (7) were used to examine the validity, reliability and internal consistency, respectively, of the instrument to assess the frequency of symptoms expected to be associated with sleep-disordered breathing. A multivariable apnea risk index incorporating the mean apnea symptom frequency along with age, BMI and gender was developed using multiple logistic regression (8). ROC curves (9–12) were used to further assess the predictive utility. Assessment of validity, reliability and internal consistency was performed by site. After examining site-to-site differences, the data from Philadelphia and Baltimore were pooled prior to estimating the parameters for the final multivariable apnea risk index. All statistical computations were performed using SAS version 6.08 (The SAS Institute, Cary, NC, U.S.A.).

Available samples and examination of possible sample biases

In preparation for the confirmatory factor analysis, missing item values were imputed by setting them equal to the mean of the nonmissing responses. This was done for each of the four domains separately. The number of respondents missing zero, one, two or three apnea symptom items were 715 (66.8%), 156 (14.6%), 97 (9.1%) and 103 (9.6%), respectively. Thus, for 103 respondents, imputed values could not be derived. These were excluded from both the factor analysis samples and the sample used in the subsequent predictive analyses. Similarly, 32 (3.0%) respondents were missing all five items from the difficulty sleeping domain, 45 (4.2%) were missing all three items from the excessive daytime sleepiness domain and 84 (7.8%) were missing both items in the narcolepsylike symptoms domain. A subject was included in the sample used for predictive analyses if he or she had a value for the apnea symptom frequency index plus values for age, gender, and height and weight. Physician enrollment data were not available for 33%, 36% and 100% of the respondents from Philadelphia, Baltimore and Pittsburgh, respectively. Given the pattern of missing values, four types of samples were defined as follows:

TABLE 1. *Available samples at the sleep disorders centers*

Sample	Philadel-phia	Baltimore	Pittsburgh
Factor analysis/ predictive model	310	104	0
Factor analysis/ no predictive model	348	89	77
No factor analysis/ predictive model	11	2	0
No factor analysis/ no predictive model	101	21	8
Total	770	216	85

Sample 1: Subject had complete data for factor analysis and predictive analysis; Sample 2: complete data for factor analysis but not for predictive analysis; Sample 3: incomplete data for factor analysis and complete data for predictive analysis; and Sample 4: incomplete data for both factor analysis or predictive analysis. Table 1 summarizes the available data for each site. Tables 2 and 3 examine the possible biases among the available subsamples and provide summaries of demographic characteristics. Of the 770 subjects from the Philadelphia site, 310 and 348 were from Samples 1 and 2, respectively. These samples did not differ significantly on the proportions with RDI ≥ 10, age, marital status and race. Sample 1 did have a significantly smaller proportion of males, however. Additionally, Sample 1 and Sample 2 did not significantly differ in the mean values for any of the four symptom frequency indices described below. For these reasons, it appeared valid to pool Samples 1 and 2 for the purpose of factor analysis.

RESULTS

Confirmatory factor analysis

The 13 questions about symptom frequency were selected from four domains. The specific symptoms are listed in Table 4. Table 4 also presents the factor loadings for our four-factor model after Varimax rotation. These results were obtained using data from the Philadelphia site for the 658 patients from Samples 1 and 2. The loadings confirmed a clear factor structure; each of the 13 symptom frequency questions has a large loading on only one factor. Specifically, the three component questions associated with sleep-disordered breathing, that is, snorting or gasping, loud snoring and breathing stops/choke/struggle for breath, all factor very highly on Factor 1, but not on Factors 2, 3 or 4. Thus, factor analysis was used to examine the construct validity of defining an apnea symptom frequency index as the average of the three a priori defined apnea symptom frequency variables. The remaining three factors were interpreted as follows: Factor 2, difficulty sleeping factor; Factor 3, excessive daytime sleepiness factor; and Factor 4, narcolepsylike factor. The proportion of total variance explained by Factors 1–4 were 32%, 26%, 25% and 17%, respectively, computed as the sum of the squared factor loadings divided by the sum of the unweighted communalities.

The four-factor model was estimated for Samples 1 and 2 from the Baltimore site (n = 193). The factor structure identified at Philadelphia was reproduced at the Baltimore site. The factor loadings for Factor 1 were as follows: snorting and gasping (0.92), loud snoring (0.82) and breathing stops (0.77). The factor load-

TABLE 2. *Demographic characteristics of available subsamples: interval variables*

Sample	Philadelphia			Baltimore			Pittsburgh		
	n^a	Mean	SD	n	Mean	SD	n	Mean	SD
1. Factor analysis/predictive model									
Age	310	47.2[b,d]	12.8	104	46.7[c]	14.4	0	—	—
BMI	310	32.7[b,d]	8.5	104	32.8[c]	9.4	0	—	—
2. Factor analysis/no predictive model									
Age	224	47.4	13.4	61	50.5	13.8	77	47.6	13.4
BMI	110	33.5	9.0	5	38.8	11.3	0	—	—
3. No factor analysis/predictive model									
Age	11	56.4	11.9	2	53.4	25.6	0	—	—
BMI	11	39.3	12.0	2	27.4	1.7	0	—	—
4. No factor analysis/no predictive model									
Age	71	57.0	14.9	16	50.2	16.2	8	64.1	13.1
BMI	68	31.6	7.6	8	29.3	7.2	0	—	—

[a] Number of patients in cell with nonmissing data.

[b] No significant difference at Philadelphia site between Sample 1 and Sample 2 for age ($t = -0.15$, df = 532, p = 0.88) and BMI ($t = -0.8$, df = 418, p = 0.42).

[c] No significant difference in Baltimore site between Sample 1 and Sample 2 for age ($t = -1.68$, df = 163, p = 0.09) and BMI ($t = -1.40$, df = 107, p = 0.17).

[d] No significant site differences in predictive analysis samples (Sample 1 and Sample 3 combined) between Philadelphia and Baltimore for age ($t = 0.47$, df = 425, p = 0.64) and BMI ($t = 0.79$, df = 425, p = 0.79).

TABLE 3. *Demographic characteristics of available sub-samples: nominal variables*

Sample	Philadelphia n[a]	%	Baltimore n	%	Pittsburgh n	%
1. Factor analysis/predictive model						
RDI ≥ 10	310	61.0[b,d]	104	54.8[e]	0	
Male	310	69.7[c,d]	104	67.3	0	
Female	310	30.3	104	32.7	0	
White	308	69.5[b,d]	104	76.9	0	
Black	308	27.3	104	19.2	0	
Married	306	62.4[b,d]	103	62.1	0	
2. Factor analysis/no predictive model						
RDI ≥ 10	267	56.2	79	63.3	67	38.8
Male	347	60.0	89	71.9	77	64.9
Female	347	40.0	89	28.1	77	35.1
White	343	62.1	82	43.2	77	83.1
Black	343	32.9	82	46.6	77	16.9
Married	339	58.1	88	71.6	76	65.8
3. No factor analysis/predictive model						
RDI ≥ 10	11	81.8	2	100	0	
Male	11	36.4	2	50.0	0	
Female	11	63.6	2	50.0	0	
White	11	54.5	2	100	0	
Black	11	45.5	2	0	0	
Married	11	45.4	2	50.0	0	
4. No factor analysis/no predictive model						
RDI ≥ 10	83	49.0	19	57.9	6	66.7
Male	98	56.1	21	57.1	8	62.5
Female	98	43.9	21	42.9	8	37.5
White	94	54.3	21	47.6	8	100
Black	94	40.4	21	42.9	8	0
Married	89	44.9	21	52.4	8	50.0

[a] Number of patients in cell with nonmissing data.

[b] No significant difference in Philadelphia between Sample 1 and Sample 2 for RDI ≥ 10 ($\chi^2 = 1.4$, df = 1, p = 0.24), race ($\chi^2 = 4.8$, df = 2, p = 0.09), and marital status ($\chi^2 = 1.2$, df = 1, p = 0.27).

[c] Significant difference in Philadelphia between Sample 1 and Sample 2 for gender ($\chi^2 = 7.2$, df = 1, p = 0.007).

[d] No significant site differences in predictive analysis samples (Sample 1 and Sample 3 combined) between Philadelphia and Baltimore for RDI ≥ 10 (p = 0.27), race (p = 0.17) and marital status (p = 0.99).

[e] No significant difference in Baltimore between Sample 1 and Sample 2 for gender (p = 0.49), RDI ≥ 10 (p = 0.25) and marital status (p = 0.17). The difference in race distribution was significant ($\chi^2 = 22.9$, df = 2, p = 0.001).

ings for Factor 2 were awakenings (0.69), tossing and thrashing (0.63), difficulty falling asleep (0.61), legs feel jumpy (0.56) and morning headaches (0.41). The factor loadings for Factor 3 were falling asleep at work or school (0.81), falling asleep while driving (0.73) and excessive daytime sleepiness (0.64). The factor loadings for Factor 4 were sleep paralysis (0.96) and dreamlike state (0.57). A similar factor structure was identified for the subjects (n = 77) from Sample 2 at Pittsburgh.

Analysis of symptom frequency indices

The results from the factor analyses supported computation of symptom frequency indices as the mean

TABLE 4. *Rotated factor pattern matrix*

Self-reported symptom	Factor 1	Factor 2	Factor 3	Factor 4
Snorting or gasping	0.933	0.085	0.079	0.043
Loud snoring	0.836	−0.023	0.050	0.010
Breathing stops, choke or struggle for breath	0.674	0.183	0.101	0.117
Frequent awakenings	0.074	0.688	0.140	0.028
Tossing, turning or thrashing	0.284	0.646	0.052	0.105
Difficulty falling asleep	−0.089	0.564	0.008	0.134
Legs feel jumpy or jerky	0.136	0.449	0.140	0.239
Morning headaches	0.014	0.381	0.100	0.270
Falling asleep when at work or school	0.091	0.069	0.861	0.113
Falling asleep when driving	0.072	0.032	0.720	0.128
Excessive sleepiness during the day	0.051	0.236	0.522	0.134
Awaken feeling paralyzed, unable to move for short periods	0.060	0.164	0.142	0.783
Find yourself in a vivid dreamlike state when falling asleep or awakening even though you know you're awake	0.084	0.272	0.214	0.541

All questions scaled as 0 = never, 1 = rarely, 2 = 1–2/week, 3 = 3–4/week, 4 = 5–7/week and DK = don't know. DK and missing were set to the mean of nonmissing values within a priori defined domains for the purpose of factor analysis.

of the items within a domain. Index 1 represented a symptom frequency index for apnea. It was computed by averaging the nonmissing values for the frequency of loud snoring, breathing cessation, and snorting and gasping. We noted that missing values primarily arose from subjects indicating that they did not know the frequency of symptom occurrence rather than just skipping the question. Analogous indices were computed using the items from the other three factors identified in the factor analysis. Table 5 presents a comparison of the means and standard deviations for the four indices for each sample definition across the three sites.

At the Philadelphia site there were no significant differences between members of Sample 1 and Sample 2 for any of the symptom frequency indices. Because these samples were similar demographically, and on the basis of the predictor and outcome variables, it was appropriate to pool them for the purpose of assessing the value of Index 1 in predicting sleep apnea.

Test-retest reliability and internal consistency

The reliability characteristics of the indices were examined by performing a test–retest analysis. At the Philadelphia site a sample of patients (n = 30) filled out another subject enrollment form at the time of their first visit. This was typically 2 weeks after filling out the original form at home. The test–retest correlations

TABLE 5. *Mean (SD) index values[a] by sample and site*

	Philadelphia			Baltimore			Pittsburgh		
	n	Mean	(SD)	n	Mean	(SD)	n	Mean	(SD)
1. Factor analysis/predictive model									
I1	310	2.67	(1.30)[b]	104	1.89	(1.36)[c]	0	—	—
I2	310	1.88	(0.96)[b]	104	1.64	(0.97)[d]	0	—	—
I3	310	1.63	(1.17)[b]	104	1.75	(1.23)[d]	0	—	—
I4	310	0.74	(1.03)[b]	104	0.68	(0.94)[d]	0	—	—
2. Factor analysis/no predictive model									
I1	348	2.52	(1.37)	89	2.51	(1.30)	77	2.05	(1.57)
I2	348	1.78	(1.01)	89	1.83	(1.05)	77	1.53	(0.99)
I3	348	1.49	(1.14)	89	1.69	(1.20)	77	1.21	(1.06)
I4	348	0.67	(0.98)	89	0.71	(1.02)	77	0.60	(1.00)

[a] Index definitions: I1, apnea symptom index; I2, difficulty sleeping symptom index; I3, excessive daytime sleepiness symptom index; I4, narcolepsy symptom index.

[b] No significant difference in Philadelphia between Sample 1 and Sample 2 for I1 ($t = 1.45$, df = 656, p = 0.15), I2 ($t = 1.27$, df = 656, p = 0.20), I3 ($t = 1.60$, df = 656, p = 0.11) or I4 ($t = 0.89$, df = 656, p = 0.38).

[c] Significant difference in Baltimore between Sample 1 and Sample 2 for I1 ($t = -3.22$, df = 191, p = 0.002).

[d] No significant site difference in Baltimore between Sample 1 and Sample 2 for I2 ($t = -1.33$, df = 191, p = 0.19), I3 ($t = 0.34$, df = 191, p = 0.73) or I4 ($t = -0.18$, df = 191, p = 0.86).

for Indices 1, 2, 3 and 4 were 0.92 (n = 29), 0.85 (n = 30), 0.79 (n = 29) and 0.86 (n = 29), respectively. All test–retest reliability coefficients were significant, with p < 0.0001. We compared the demographic characteristics of the test–retest subsample to the remaining patients in the Philadelphia sample and found no significant differences in age, BMI, marital status and gender. Furthermore, the mean values of the four indices did not differ significantly. However, the test–retest subsample had a proportion with RDI ≥ 10 that was higher than the rest of the Philadelphia sample (80.0% vs. 57.1%, p = 0.02).

We examined the internal consistency of the four indices using Cronbach alpha statistics computed for the factor analysis samples (Samples 1 and 2) at each site separately. The Cronbach alpha values for Index 1 were estimated to be 0.85 at the Philadelphia site, 0.88 at the Baltimore site and 0.93 at the Pittsburgh site. In comparison, the Cronbach alpha values for Indices 2, 3 and 4 were 0.71, 0.76 and 0.66, respectively, at Philadelphia. Values similar to these were also obtained at the Baltimore and Pittsburgh sites.

Initial assessment of the predictive ability of Index 1

The initial assessment of the predictive ability of Index 1 was carried out in the predictive model sample (n = 321) at the Philadelphia site. This was done by categorizing Index 1. The results are shown in Table 6. An RDI cutoff of 10 was used to define the presence of apnea in order to compare our results with those in the literature (e.g. references 1–4). The table shows the estimated prevalence of sleep apnea for patients with Index 1 values in the indicated categories. The cate-

gories were chosen so that a value of 4 is obtained for patients indicating the maximum frequency for all three items. The category of 3.33–3.67 represents a patient indicating maximum frequency on at least one item but not all three. The remaining categories represent a variety of item patterns. Thus, we find that the prevalence of apnea ranges from 20% in those patients with Index 1 values <1 to 74% in those patients with Index 1 values equal to 4.

Development of the multivariable apnea risk index (MAP index)

Multiple logistic regression models were used to incorporate BMI, age and gender into a multivariable apnea risk index (MAP index). The model was developed as follows: First, main effects logistic regression models were estimated separately for the Philadelphia and Baltimore sites, pooling their Samples 1 and 3, respectively. Data from the Pittsburgh site could not be used because BMI was missing. Then, the significance of site differences in the relationship between each variable and the probability of apnea was assessed by pooling the data and adding an indicator variable

TABLE 6. *Predictive value of Index 1 alone at Philadelphia site*

Index 1 category	No. with RDI ≥ 10	Total no.	% RDI ≥ 10
0–<1	8	40	20
1–<2	17	37	46
2–<3	45	69	62
3	29	41	71
3.33–3.67	29	40	73
4	70	94	74

TABLE 7. *Site-specific main effects logistic regression models*

| Variable[a] | Philadelphia | | Baltimore | | |
	Beta[b]	(SE)[c]	Beta	(SE)	p-value[d]
Intercept	−5.718	(0.966)	−7.075	(1.610)	
I1	0.445	(0.114)	0.376	(0.203)	0.50
BMI	0.093	(0.019)	0.088	(8.227)	0.63
Age	0.023	(0.011)	0.058	(0.019)	0.08
Male	1.331	(0.315)	1.427	(0.545)	0.94

[a] All variables in both models significant at $p < 0.05$.

[b] Beta coefficient derived from maximum likelihood estimation of a multivariable logistic regression.

[c] Standard error of the estimated beta coefficient.

[d] p-value from chi-square test for site differences (interaction test) in pooled model.

TABLE 8. *Final model with interaction*

Variable	Beta[a]	(SE)[b]	p-value[c]
Intercept	−8.160	(1.334)	
I1	1.299	(0.403)	0.001
BMI	0.163	(0.038)	<0.0001
I1 by BMI	−0.028	(0.013)	0.026
Age	0.032	(0.0093)	0.0006
Male	1.278	(0.273)	<0.0001

[a] Beta coefficient derived from maximum likelihood estimation of a multivariable logistic regression using data pooled from Philadelphia and Baltimore.

[b] Standard error of the estimated beta coefficient.

[c] p-value from chi-square.

for site + site × predictor variable interactions. The interactions were tested one at a time. Table 7 summarizes the results. There were no significant differences between the sites for any of the variables. All variables at both sites were significant. The fact that Index 1 was significant at both sites in multivariable models including BMI is further evidence of construct validity. It demonstrates that Index 1 is a valid measure of apnea risk, not just of obesity.

In addition, there was no difference in outcome prevalence holding constant the variables in the main effects model (site effect $\chi^2 = 0.13$, df = 1, p = 0.72), nor was there any significant difference between sites in the demographic variables summarized in Tables 2 and 3. Therefore, the samples were pooled to increase the power for identifying interactions among the predictor variables. A significant interaction between Index 1 and BMI emerged ($\chi^2 = 4.97$, df = 1, p = 0.026). Quadratic terms were then added to this model one at a time for Index 1, age and BMI. This was to assess whether nonlinear relationships existed. None of these terms were significant in the model that included the Index 1 × BMI interaction. The results from this final logistic regression model are summarized in Table 8. The Hosmer-Lemeshow Goodness of Fit chi-square statistic (8) was used to test the hypothesis that the estimated model does not adequately fit the data. Because $\chi^2 = 7.36$, df = 8, p = 0.50, there was no indication of a poor fit of the model to the data.

As last steps, we added systolic and diastolic blood pressures to the model because they might be attainable at relatively low cost if found of additional predictive value. When they were simultaneously added to the model as either linear or linear + quadratic terms, no significant increase in the model's ability to predict was found. We then simultaneously added Indices 2, 3 and 4 to the model. These indices added no additional explanatory power to the model ($\chi^2 = 3.18$, df = 3, p = 0.36) and so were not included in the MAP index.

Figure 1 illustrates the nature of the statistically significant Index 1 × BMI interaction. This figure provides the model-estimated probability of apnea for a 49-year-old male with varying values of BMI computed for Index 1 values of 0–4. Other ages or female gender would only represent parallel shifting of the curves up or down. We find that Index 1 is useful in discriminating patients with and without sleep apnea only if the patient is not extremely obese (i.e. <40 BMI).

The estimated probabilities that a patient will have an RDI ≥10 may be computed using the coefficients listed in Table 8. First, the linear predictor (x) is computed by incorporating patient-specific data. Then, the estimated probability is computed from x using the following formula:

$$\text{Probability} = e^x/(1 + e^x)$$

where

$$x = -8.160 + 1.299 \cdot \text{Index } 1 + 0.163 \cdot \text{BMI}$$
$$- 0.028 \cdot \text{Index } 1 \cdot \text{BMI} + 0.032 \cdot \text{Age}$$
$$+ 1.278 \cdot \text{Male},$$

and Male = 1 if male and 0 if female.

Assessment of clinical utility of the multivariable apnea risk index

Figure 2 presents ROC curves for the predicted probabilities derived from the MAP index. Also plotted are the ROC curves for Index 1 alone and for BMI alone. These curves were derived by plotting sensitivity (true positive rate) versus the complement of specificity (false positive rate) using various cutpoints. For the MAP index and for BMI alone, cutpoints were defined at each decile of predicted probability derived from logistic regression models. For Index 1, this was done for the cutpoints defined in Table 6. The areas and SE (in parentheses) under the ROC curves for the MAP index, BMI alone and Index 1 alone, were 0.786 (0.023),
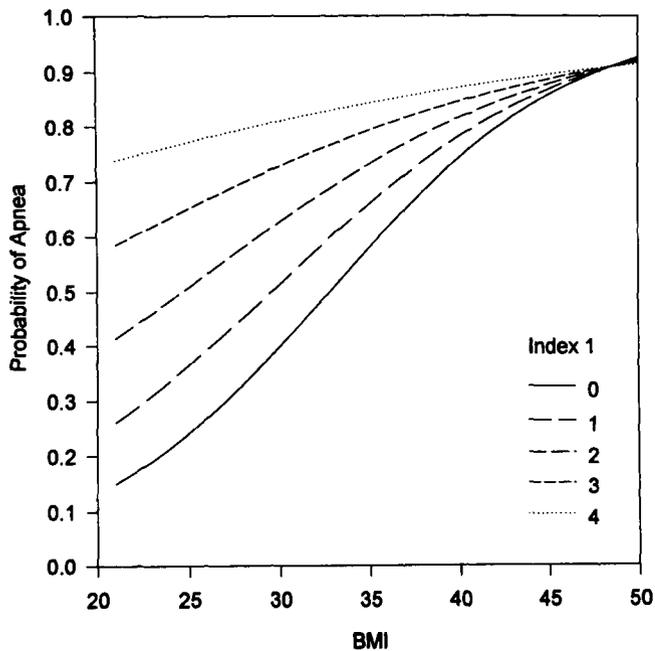
FIG. 1. Estimated probabilities that RDI ≥ 10 for a 49-year-old man. The probabilities are given as a function of BMI and computed separately for Index 1 values of 0–4.

FIG. 2. Receiver operating characteristic (ROC) curves for the predicted probabilities derived from the multivariable logistic regression model (MAP index), BMI alone and Index 1 (self-report index of frequency of apnea symptoms). The dotted line represents the line tangent to the MAP index ROC assuming equal costs of false positives and false negatives.

0.734 (0.025) and 0.695 (0.028), respectively. The areas under the ROC curves (11,12) represent the probability that a randomly selected patient with apnea has a predicted value larger than a randomly selected patient without apnea. The difference between 0.786 and 0.734 represents a measure of the increase in predictive power for the MAP index as compared to making prediction using BMI alone. This difference is statistically significant, as demonstrated by the simultaneous significance of the additional parameters in the logistic regression model (partial $\chi^2 = 72.5$, df $= 4$, p < 0.0001).

DISCUSSION

Several authors (1–4) have explored the utility of the self-reporting of symptoms usually associated with sleep-disordered breathing. Kapuniai et al. (1) computed the sensitivity, specificity, and positive and negative predictive value of a decision rule that required both loud snoring and breathing cessation to be present sometimes, often or always. For predicting an apnea + hypopnea index >10/hour, they reported a sensitivity and specificity of 78% and 67%, respectively, and positive and negative predictive values of 64% and 80%, respectively. This exploratory study was based on a relatively small number of patients, however. Similarly, Haraldsson et al. (3) compared a clinical diagnosis of sleep apnea based on a self-report of symptom frequency to a diagnosis made on the basis of a nocturnal polysomnography also in a relatively small

number of patients. Viner et al. (2) extended this work by investigating the predictive utility of a screen based on age, BMI, male sex and the presence of snoring in a large sample (n = 410). However, patients in this study were specifically referred for sleep apnea syndrome. Also, examinations of possible statistical interactions among these variables were not reported. The presence of statistical interaction would suggest that the predictive utility of a specific variable depended upon other factors. Hoffstein and Szalai (4) based their analyses on a continuous measure of RDI in a population of patients also specifically referred to a sleep disorders clinic because of a suspicion of sleep apnea. Again, the issue of prediction in general sleep disorders center populations was not addressed.

Our study differs from these investigations in several ways. First, it is based on general populations of patients presenting at sleep disorders centers. Secondly, our study was predicated on first examining the validity and reliability of using an index computed from the self-report of the frequency of symptoms usually thought to be associated with sleep-disordered breathing. This documentation is an essential first step when proposing that an instrument has practical utility. To this end, we embedded three questions thought to be associated with sleep-disordered breathing in a questionnaire that also contained 10 other questions about a variety of symptoms. These questions were chosen based on the judgment of sleep physicians and sleep

experts with the aim of asking about symptoms thought to reflect other sleep disorders. The factor analysis confirmed that the chosen questions could discriminate among patients with symptoms from several domains, namely, sleep-disordered breathing, difficulty sleeping, excessive daytime sleepiness and narcolepsy. Additionally, we estimated reliability using test–retest correlations. The test–retest correlations were high for all four indices, especially for the apnea symptom index (0.92). Although the test–retest subsample was similar demographically to the total sample, there was significantly more apnea. This could have biased the reliability result for the apnea index but should have had little effect on the other indices. It also appears that the three apnea symptom questions have substantial internal consistency as measured by a Cronbach alpha statistic. The values for the other symptom indices were also reasonably large when one considers the small number of items per index. Thirdly, we looked for statistical interactions among the variables in the multivariable model. We postulated that the presence of interactions between Index 1 and other variables might identify subgroups of patients in which the frequency of apnea symptoms is especially discriminating. In our study the statistical interaction between Index 1 and BMI suggested that the ability of the screening tool to differentiate between those at risk and not at risk for sleep apnea may be larger in the subpopulation not already at risk due to a large BMI.

It may be tempting to apply our model in other clinical settings. However, the predictive characteristics of diagnostic tests vary from one clinical population to another for at least two reasons. The distribution of other clinical characteristics, such as comorbid conditions, age, etc., may differ, and these differences may affect the sensitivity and specificity of diagnostic tests. In addition, there may be varying prevalences of the disease under study caused by differing referral patterns, etc. Thus, the results of this study must be validated prior to use in other settings. To do so, the intercept may be removed from the logistic regression equation presented above. The resulting modified MAP index becomes a unitless measure of relative apnea risk. To determine its utility in other clinical settings, the modified MAP index may be computed for a large number of patients. Then, the MAP values may be categorized into deciles or quintiles. The MAP index can be considered to be effective in the new clinical setting if the proportions with RDI ≥ 10 increase steeply as a function of category number.

Quantitative indices such as the MAP index can have a number of uses in clinical settings. For example, they can be used in primary care physician offices. A patient with a high probability could be referred immediately for a diagnostic sleep study. A patient with low probability could be first referred to a sleep specialist for clinical evaluation. Alternatively, indices such as the MAP index could be used in the context of managed care. A primary care physician could justify the cost of an overnight study for patients with high predicted probabilities. In experimental trials used to assess the clinical utility of in-home overnight polysomnography, analyses could be done stratified by initial risk as defined by the MAP index. In treatment efficacy trials, patient allocation could be made within strata defined by the MAP index. A more statistically powerful comparison among treatments would arise if there were significant relationships between the index values and treatment efficacy.

Many such uses require the determination of an optimal cutpoint on which to base the prediction of apnea. The determination of a specific cutpoint is based on outcome prevalence and the ratio of costs associated with false positive predictions to costs associated with false negative predictions (10). These costs are derived for the specific context in which the prediction is being made. Thus, they may include costs from a wide range of factors, including costs from the patient's perspective (e.g. ratio of expected healthy life years) and costs from society's perspective (e.g. ratio of total expected healthcare costs). The cutpoint that minimizes total errors may be used when costs are equal or when there is no information regarding costs (10). In that case, the optimal cutpoint is derived from the ROC curve using the ratio of nonapnea prevalence to apnea prevalence. The optimal cutpoint is located where the slope of the line tangent to the curve is equal to this ratio. For our sample this is 0.4/0.6, or 0.67. This tangent line is indicated as the dotted line in Fig. 2 and corresponds to an MAP index cutpoint of approximately 0.50. Thus, for example, a primary care physician could refer a patient directly for a sleep study if the MAP value was larger than 0.50 and could refer the patient for a clinical evaluation if the MAP value was less than 0.50. The estimated positive and negative predictive values (95% confidence interval) of this cutpoint are 0.75 (0.70–0.80) and 0.74 (0.66–0.82), respectively. The estimated sensitivity and specificity of this cutpoint are 0.88 (0.84–0.92) and 0.55 (0.48–0.62), respectively. These appear to be in a range consistent with potential clinical utility.

In conclusion, we have shown that the self-report of symptoms of sleep apnea can be obtained in a valid and reliable way. Age, gender and BMI added additional information. A multivariable apnea risk index was derived and a method for modifying the prediction rule for use in other settings was given. Finally, the derivation of optimal cutpoints was discussed. Thus, the MAP index is potentially useful in the allocation of medical resources, in complementing the diagnostic process and in aiding in efficient research designs.

REFERENCES

1. Kapuniai LE, Andrew DJ, Crowell DH, Pearce JW. Identifying sleep apnea from self-reports. *Sleep* 1988;11:430–6.
2. Viner S, Szalai JP, Hoffstein V. Are history and physical examination a good screening test for sleep apnea? *Ann Int Med* 1991;115:356–9.
3. Haraldsson PO, Carenfelt C, Knutsson E, Persson HE, Rinder J. Preliminary report: validity of symptom analysis and daytime polysomnography in diagnosis of sleep apnea. *Sleep* 1992;15: 261–3.
4. Hoffstein VH, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. *Sleep* 1992;16:118–22.
5. Rechtschaffen A, Kales A. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Los Angeles: Brain Information Service, 1968.
6. Carmines EG, Zeller RA. *Reliability and validity assessment,* No. 17 Series: Quantitative Applications in the Social Sciences. Newbury Park: Sage Publications, 1990.
7. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1971;16:297–334.
8. Hosmer DW, Lemeshow S. *Applied logistic regression.* New York: Wiley, 1989.
9. Beck JB, Shultz EK. The use of relative operating characteristic (ROC) curves in test performance evaluation. *Arch Pathol Lab Med* 1986;110:13–20.
10. McNeil BJ, Keeler E, Adelstein SJ. Primer on certain elements of medical decision making. *N Engl J Med* 1975;293:211–5.
11. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic curve (ROC). *Radiology* 1982; 143:29–36.
12. ROCFIT computer program. Department of Radiology, Biological Sciences Division, University of Chicago Hospitals and Clinics. Available ftp at random.bsd.uchicago.edu.

Appendix E:

Details Of Sample Size Analyses

Synopsis

This appendix (Appendix E) provides a technical appendix containing sample size analyses related to estimating apnea prevalence using a risk-group stratified, non-proportional sampling design. It presents our analyses indicating that our design criterion for a total sample size of 410 CDL holders (250 Higher Risk and 160 Lower Risk) was an appropriate and reasonable sample size for estimating the population prevalence of at least moderate sleep apnea as well as to provide adequate within strata estimates. For this discussion, the statistical criterion of "margin-of-error" (i.e., the half-width of a 95% confidence interval) is used as the criteria in which to judge sample size adequacy.

Prior to sampling from the Lower Risk population we determined that a sample of 160 Lower risk CDL holders was sufficient to produce reasonably precise estimates of at least moderate apnea prevalence(AHI=15 event/hr) taking into account the two-stage sampling design. We found that the rate of at least moderate apnea would be estimated with a margin-of-error no greater than approximately between ±3% and ±4%. We decided this was a sufficiently precise estimate based on comparisons with margins-of-error from typical national surveys. Furthermore, within strata, we estimated that for both the Higher and Lower risk groups, our design would result in estimated prevalence values for least moderate apnea with margins-of-error approximately equal to ±4-5%. Again, these appeared to represent sufficiently precise within strata estimates.

Statistical Measure of Precision

A widely used statistical criterion for determining whether a prevalence estimate is sufficiently precise is ½ the width of a 95% confidence interval for the true proportion (or percentage). This is often referred to as the "margin-of-error". For example, in national presidential preference polls, the margin-of-error is set to approximately ±3%. The value of ± 3% is deemed sufficiently small so as to allow the statistical estimate to be useful for decision-making. For true percentages near 50%, a sample size of 1068 is required to produce a margin-of-error equal to ±3%. This is why national presidential polls reported on television typically seek between 1000 and 1200 respondents.

Operationally, a margin-of-error of ±3% means that with "95% confidence" we can "rule out" that the true percentage is 3% larger or smaller than the percentage obtained in the random sample. For example, suppose that 54% of a simple random sample of 1100 registered voters stated that they preferred the incumbent in a national presidential election. Since the margin-of-error is ±3% we have determined with acceptable levels of statistical certainty that more than a majority preferred the incumbent (i.e., the true rate was no smaller than 51%) in the population from which the sample was drawn. A more precise definition is that a confidence interval of width 6% will cover the true percentage with probability equal to 0.95.

The relationship between the desired margin-of-error and the required sample size is highly dependent upon the true underlying percentage and is given the formula below:

$$\text{Margin-of-error} \cong 1.96 * \text{square root of } [\pi * (1-\pi)/n]$$

Therefore, the required sample size can be expressed as a function of the margin-of-error (E) as follows:

$$n @ (1.96)^2 * [p * (1-p)] / E^2$$

This presents a problem because the value of π, the true population prevalence is unknown. However, the required sample size is at a maximum when the true proportion is set to 0.50. Table E1 illustrates the nature of the relationships among the required sample size, the desired margin-of-error, and the true underlying proportion. These computations assume a simple random sample.

For prevalence values less than about 22% a sample size of about 412 is required to obtain a margin-of-error $\cong \pm 4\%$. Below, we describe how our stratified, non-proportional sampling was expected to provide increases in statistical precision (i.e., smaller margins-of-error) making the values in Table E1 conservative estimates of required sample sizes.

Sample Size Required For Simple Random Samples

We now discuss the implications of the design specification for margins-of-error equal to the values described above in terms of estimating the prevalence of sleep apnea in a population of holders of CDLs. Suppose that the true prevalence of at least moderate apnea (AHI $\geq$15) is 10%. In Young et al [1993], the prevalence of AHI $\geq$5 for men 30-60 was estimated to be 9.1% (95% CI=6.4-11.0) and so 10% is a very reasonable value to assume for the purpose of sample size analysis. Table E1 indicates that a sample of at least 384 is necessary to obtain a margin-of-error no greater than $\pm 3\%$. If the prevalence is higher, say 15%, the margin-of-error will be larger than 3% but smaller than 4%. At a prevalence of 20%, a sample size of 384 results in a margin-of-error of exactly 4%. Thus, we concluded that for a simple random sample with a prevalence at least moderate apnea of no more than 20%, a total sample size of 384 would provide reasonably precise estimates if we had a simple random sample.

Sample Size Required For Our Stratified Random Sample

We now discuss the effect of the sampling design on the required sample size. The calculations above are strictly valid only when a simple random sample has been obtained. In many cases, stratified random sampling results in increased precision [Cochran, 1977, pages 107-111]. The degree of precision gained from stratified random sampling depends on how different the strata are in terms of the percentage of individuals with the study characteristic. Since the multivariable apnea prediction (MAP) [Maislin et al, 1995] is designed to discriminate between individuals with and without sleep apnea, we expected gains in precision resulting from the stratified sampling. However, the gain is modified depending on the relative strata sample sizes. Although proportional sampling (when the sample sizes of the strata are proportional to the sizes of the strata in the population) has the distinct advantage that it allows for unweighted analyses, over-

sampling strata in proportion to their relative variance provides further gains in statistical precision. Such estimators have the property that the variance of the estimator is minimized. Since the variance of sample proportions is smallest for values close to 0 or 1 and largest for values close to 0.50, the Higher risk group provides less efficient prevalence estimators. Therefore, further statistical precision is gained by over-sampling from this group relative to the Lower risk group.

We note that our estimate of the prevalence of at least moderate sleep apnea is 9.6% with a margin-of-error equal 2.4% (95% confidence interval equal to 7.4% to 11.9%). A simple random sample would have resulted in a margin-of-error equal to 2.9%. Thus, we observe the increased efficiency in terms of a reduction in the margin-of-error from 2.9% to 2.4%. This reflects a study design effect (deff) equal to 0.68 [Kish, 1965]. Furthermore, we note that our design expectation that the population rate would be approximately equal to 10% was confirmed by the result of our study.

Application

In summary, our design purposely does not result in a pooled representative sample. Non-proportionality was chosen by design in order to enroll enough subjects in the subsequent study of the relationship between sleep apnea and function. However, we see from the discussion above, it has the added benefit of increasing the precision of prevalence estimates. Specifically, we defined two strata, Higher vs. Lower apnea risk according to respondents' Multivariable Apnea (MAP) [Maislin et al, 1995] risk score obtained from the population survey. Our design was to enroll 250 respondents from among the 500 (50%) with the highest MAP values. In fact, we enrolled 247 from the 551 (44.8%) survey respondents with the highest MAP values, fairly close to the desired target. We then enrolled 159 from the remaining 771 (20.6%) survey respondents with non-missing MAP values in random order. Thus, in the population, Higher and Lower risk strata were defined based on whether individual MAP values were above or below this *post hoc* determined threshold. The threshold MAP value that provided the boundary between Higher and Lower risk groups was 0.435. Among survey respondents, 41.5% had MAP values above 0.435 and 58.5% had MAP values below 0.435. Since 247 of 406 (60.8%) in-laboratory subjects were from the Higher risk strata, we over-sampled the Higher risk group relative to the theoretical proportional sizes of the population strata. Thus, we realized gains in efficiency as described above in addition to the required over-sampling of subjects with various levels of apnea for the purpose of subsequent analyses concerning apnea's impact on functional performance.

Table E1
Sample Sizes Necessary to Achieve Desired Margins-of-Error
For Various Population Values of Apnea Prevalence
Assuming Simple Random Sampling

Margin-of-Error

Prevalence	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
0.01	380	95	42	24	15	11	8	6	5	4
0.02	753	188	84	47	30	21	15	12	9	8
0.03	1118	279	124	70	45	31	23	17	14	11
0.04	1475	369	164	92	59	41	30	23	18	15
0.05	1825	456	203	114	73	51	37	29	23	18
0.06	2167	542	241	135	87	60	44	34	27	22
0.07	2501	625	278	156	100	69	51	39	31	25
0.08	2827	707	314	177	113	79	58	44	35	28
0.09	3146	787	350	197	126	87	64	49	39	31
0.10	3457	864	*384*	216	138	96	71	54	43	35
0.11	3761	940	*418*	235	150	104	77	59	46	38
0.12	4057	1014	451	254	162	113	83	63	50	41
0.13	4345	1086	483	272	174	121	89	68	54	43
0.14	4625	1156	514	289	185	128	94	72	57	46
0.15	4898	1225	544	306	196	136	100	77	60	49
0.16	5163	1291	574	323	207	143	105	81	64	52
0.17	5420	1355	602	339	217	151	111	85	67	54
0.18	5670	1418	630	354	227	158	116	89	70	57
0.19	5912	1478	657	370	236	164	121	92	73	59
0.20	6147	1537	683	384	246	171	125	96	76	61
0.22	6592	1648	732	*412*	264	183	135	103	81	66
0.24	7007	1752	779	438	280	195	143	109	87	70
0.26	7391	1848	821	462	296	205	151	115	91	74
0.28	7745	1936	861	484	310	215	158	121	96	77
0.30	8067	2017	896	504	323	224	165	126	100	81
0.35	8740	2185	971	546	350	243	178	137	108	87
0.40	9220	2305	1024	576	369	256	188	144	114	92
0.45	9508	2377	1056	594	380	264	194	149	117	95
0.50	9604	2401	1067	600	384	267	196	150	119	96

Notes: The criterion used to specify the desired statistical precision was that the margin-of-error (i.e., ½ width of a 95% confidence interval) for estimating the proportion with AHI≥15 was to be no larger than ±0.03 to ±0.04. For the target sample size of approximately 410, the bold italicized values indicate that this level of precision would be achieved as long as the true prevalence was between approximately 0.10 and 0.22 and would be exceeded if the true prevalence was less than 0.10. In Young et al [1993], the prevalence of AHI≥15 for men 30-60 was estimated to be 0.091 (95% CI=0.06-0.11). Furthermore, additional gains in precision were likely as a consequence of the stratified sampling plan blocked on the likelihood of having sleep apnea. Thus, it was concluded that the desired level of statistical precision would likely be achieved.

Appendix F:

Questionnaire Used In Initial Survey

Of Sample

SLEEP SURVEY FORM

I agree to allow my answers in this questionnaire to be used for research purposes. My understanding is that, if the answers are so used, my identity will be kept confidential.

_____
Signed

Please write down your phone number here: _____

For questions 1, 2 ,3, and 4, please completely fill in one circle per question that best describes you.

1) GENDER   2) ARE YOU CURRENTLY EMPLOYED
   Male   O   AS A TRUCK DRIVER?
   Female   O   Yes, full time   O
                                     Yes, part time   O
                                     No, (Go to #5)   O

                                  4) DOES YOUR CURRENT DRIVING
3) IS YOUR DRIVING?   SCHEDULE INCLUDE?
   Over the road   O   Only days   O
   Local   O   Only nights   O
   Both   O   Both days and nights O

For questions 5 through 8, please write down the correct number in the space provided and then fill in the correct circles. See the EXAMPLE below.

       EXAMPLE   5) DATE OF BIRTH
       M M/D D/YY   M M/D D/YY
Write #_ _/_ _/_ _   _ _/_ _/_ _

 0 - O O O O O O   0 - O O O O O O
 1 - O O O O O O   1 - O O O O O O
 2 - O O O O O O   2 - O O O O O O
 3 - O O O O O O   3 - O O O O O O

```
4  -  O O O O O O                    4  -  O O O O O O
5  -  O O O O O O                    5  -  O O O O O O
6  -  O O O O O O                    6  -  O O O O O O
7  -  O O O O O O                    7  -  O O O O O O
8  -  O O O O O O                    8  -  O O O O O O
9  -  O O O O O O                    9  -  O O O O O O
```

Page 2

6) How many years have you been driving a truck?

7) How many miles a year do you usually drive a truck?

```
      Write # - -                       Write # - - - - - -
        0  -  O O                         0  -  O O O O O O
        1  -  O O                         1  -  O O O O O O
        2  -  O O                         2  -  O O O O O O
        3  -  O O                         3  -  O O O O O O
        4  -  O O                         4  -  O O O O O O
        5  -  O O                         5  -  O O O O O O
        6  -  O O                         6  -  O O O O O O
        7  -  O O                         7  -  O O O O O O
        8  -  O O                         8  -  O O O O O O
        9  -  O O                         9  -  O O O O O O
```

```
   8) WEIGHT          9) COLLAR SIZE            10) HEIGHT
     (POUNDS       (e.g., 16.5 INCHES)   (FEET)  and (INCHES)

        _ _ _              _ _._            _            _ _
0  -  O O O        0  -  O O  O      4  -  O      0  -  O
1  -  O O O        1  -  O O  O      5  -  O      1  -  O
2  -  O O O        2  -  O O  O      6  -  O      2  -  O
3  -  O O O        3  -  O O  O      7  -  O      3  -  O
4  -  O O O        4  -  O O  O                   4  -  O
5  -  O O O        5  -  O O  O                   5  -  O
6  -  O O O        6  -  O O  O                   6  -  O
7  -  O O O        7  -  O O  O                   7  -  O
8  -  O O O        8  -  O O  O                   8  -  O
9  -  O O O        9  -  O O  O                   9  -  O
                                                 10 -  O
                                                 11 -  O
```

11) The following questions refer to your behavior while sleeping or trying to sleep. Please fill in one circle for each question. During the last month have you had, or have been told about the following?

	Never	Rarely, less than once a week	1-2 times per week	3-4 times per week	5-7 times per week	Don't Know
Loud snoring	O	O	O	O	O	O

	Never	Rarely, less than once a week	1-2 times per week	3-4 times per week	5-7 times per week	Don't Know
Your legs feel jumpy or jerk	O	O	O	O	O	O
Difficulty falling asleep	O	O	O	O	O	O
Frequent awakenings	O	O	O	O	O	O
Snorting or gasping	O	O	O	O	O	O

	Never	Rarely, less than once a week	1-2 times per week	3-4 times per week	5-7 times per week	Don't Know
Frequent tossing, turning, or thrashing	O	O	O	O	O	O
Your breathing, stops or you struggle for breath	O	O	O	O	O	O
Any snoring	O	O	O	O	O	O
Excessive sleepiness during waking hours	O	O	O	O	O	O
Morning Headache	O	O	O	O	O	O
Falling asleep while driving on or off the job	O	O	O	O	O	O
Awaken feeling paralyzed, unable to move for short periods	O	O	O	O	O	O
Find yourself in a vivid dreamlike state when falling asleep or awakening even though you know you're awake	O	O	O	O	O	O

Appendix G:

Pages From Sleep Apnea Diary

14. What time is it now?_____ hr_____min a.m./p.m.

15. Did you sleep or nap during the morning, afternoon, or evening today? YES NO

    16. If YES, how many times?_____ from:_____to:_____ from:_____to:_____

Rate how you felt overall today? *(mark an X somewhere along the line below)*

17. alert	|_____|	sleepy
18. stressed	|_____|	calm
19. happy	|_____|	unhappy
20. sick	|_____|	healthy
21. physically exhausted	|_____|	energetic
22. mentally exhausted	|_____|	sharp

23. Indicate amounts of caffeine today:_____cups of coffee _____colas _____tea

24. List any illness, infection, pain, discomfort, worry, or problem that you had today.

    _____

25. List all medications you took today:_____

26. Did you remove your actigraph today? YES NO

    If YES, when?_____hr_____min a.m./p.m. How long?_____hr_____min

1. What time is it now?_____ hr_____min a.m./p.m.

2. Did you sleep last night? YES NO
3. What time did you awaken this morning?_____hr_____min a.m./p.m.
4. What time did you fall asleep last night? _____hr_____min a.m./p.m.

5. How well did you sleep last night? *(mark an X somewhere along the line below)*
   very poorly |_____| excellent

6. Last night, did you have trouble falling asleep? YES NO
7. Last night, did you have trouble with waking up during the night? YES NO
8. Last night, did you have trouble with waking up too early and not being able to fall asleep again? YES NO

9. How many times did you awaken last night?_____ Duration of awakening(s)?_____

10. List all medications you took last night:_____

11. How do you feel now? *(mark an X somewhere along the line below)*
    extremely tired |_____| very refreshed

12. List any illness, infection, pain, discomfort, worry, or problem you had last night. _____

    _____

13. Did you remove your actigraph at any time last night or this morning? YES NO
    If YES, when?_____hr_____min a.m./p.m. How long?_____hr_____min

Since awakening this morning, please indicate whether you have experienced any of the following today for more than 5 minutes, and indicate the peak intensity of the experience (1=very low intensity; 2=low intensity; 3=moderate intensity; 4=high intensity; 5=very high intensity).

Experience	Please circle "Yes" or "No"		If "Yes," rate intensity (1-5)
Upset stomach/bowel	Yes	No	
Headache	Yes	No	
Difficulty concentrating	Yes	No	
Sadness	Yes	No	
Back aches/pains	Yes	No	
Giddines	Yes	No	
Difficulty remembering	Yes	No	
Muscular aches/pains	Yes	No	
Joint aches/pains	Yes	No	
Irritability	Yes	No	
Itchy skin	Yes	No	
Feeling too hot	Yes	No	
Feeling too cold	Yes	No	
Frightened	Yes	No	
Worried	Yes	No	
Quietness (more than usual)	Yes	No	
Excitement (more than usual)	Yes	No	
Tiredness (more than usual)	Yes	No	
Feeling confused	Yes	No	
Feeling anxious	Yes	No	

TUESDAY NIGHT

DATE:_____

14. What time is it now?_____ hr_____min a.m./p.m.

15. Did you sleep or nap during the morning, afternoon, or evening today? YES   NO
    16. If YES, how many times?_____   from:_____to:_____   from:_____to:_____

Rate how you felt overall today? *(mark an X somewhere along the line below)*
    17. alert |_____| sleepy
    18. stressed |_____| calm
    19. happy |_____| unhappy
    20. sick |_____| healthy
21. physically exhausted |_____◄_____| energetic
22. mentally exhausted |_____| sharp

23. Indicate amounts of caffeine today:_____cups of coffee _____colas _____tea

24. List any illness, infection, pain, discomfort, worry, or problem that you had today.

_____

25. List all medications you took today:_____

26. Did you remove your actigraph today? YES   NO
    If YES, when?_____hr_____min a.m./p.m.   How long?_____hr_____min

Appendix H:

Prevalence Estimates Of Sleep Apnea

AHI[&]	Higher Risk N=247					Lower Risk N=158					Weighted			
	N	Rate	SE	(95%C.I.)		N	Rate	SE	(95%C.I.)		Rate	SE	(95%C.I.)	
<5	149	0.603	0.040	0.525	0.682	146	0.924	0.022	0.881	0.967	0.791	0.021	0.750	0.832
5 - <15	53	0.215	0.056	0.104	0.325	10	0.063	0.077	0.000	0.214	0.126	0.051	0.027	0.226
15 - <30	26	0.105	0.060	-0.013	0.223	2	0.013	0.079	0.000	0.168	0.051	0.053	0.000	0.154
>= 30	19	0.077	0.061	-0.043	0.197		0.000		0.000	0.000	0.032	0.025	0.000	0.082

Notes: [&] Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.

Table H.1. Prevalence of different severities of apnea. AHI computed as total apneas plus hypopneas per hours of sleep. Hypopneas were defined as >=4% desaturation + air flow arousal hypopneas.

AHI[&]	Higher Risk N=246					Lower Risk N=158					Weighted			
	N	Rate	SE	(95%C.I.)		N	Rate	SE	(95%C.I.)		Rate	SE	(95%C.I.)	
<5	119	0.484	0.046	0.394	0.574	140	0.886	0.027	0.833	0.939	0.719	0.025	0.671	0.767
5 - <15	69	0.280	0.054	0.174	0.386	16	0.101	0.075	0.000	0.249	0.176	0.050	0.079	0.273
15 - <30	30	0.122	0.060	0.005	0.239	2	0.013	0.079	0.000	0.168	0.058	0.052	0.000	0.161
>= 30	28	0.114	0.060	-0.004	0.231		0.000		0.000	0.000	0.047	0.025	0.000	0.096

Notes: [&] Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.

Table H.2. Prevalence of different severities of apnea. AHI computed as total apneas plus hypopneas per hours of sleep. Hypopneas were defined as >=3% desaturation + air flow arousal hypopneas.

AHI[&]	Higher Risk N=246					Lower Risk N=158					Weighted			
	N	Rate	SE	(95%C.I.)		N	Rate	SE	(95%C.I.)		Rate	SE	(95%C.I.)	
<5	82	0.333	0.052	0.231	0.435	126	0.797	0.036	0.727	0.868	0.605	0.030	0.546	0.664
5 - <15	77	0.313	0.053	0.209	0.417	25	0.158	0.073	0.000	0.301	0.222	0.048	0.128	0.317
15 - <30	46	0.187	0.057	0.074	0.300	5	0.032	0.078	0.000	0.185	0.096	0.052	0.000	0.197
>= 30	41	0.167	0.058	0.053	0.281	2	0.013	0.079	-0.142	0.168	0.077	0.052	0.000	0.179

Notes: [&] Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.

Table H.3. Prevalence of different severities of apnea. AHI computed as total apneas plus hypopneas per hours of sleep. Hypopneas were defined as >=2% desaturation + air flow arousal hypopneas.

AHI[&]	Higher Risk N=247					Lower Risk N=158					Weighted			
	N	Rate	SE	(95%C.I.)		N	Rate	SE	(95%C.I.)		Rate	SE	(95%C.I.)	
<5	151	0.611	0.040	0.534	0.689	151	0.956	0.017	0.923	0.989	0.813	0.019	0.775	0.850
5 - <15	55	0.223	0.056	0.113	0.333	6	0.038	0.078	0.000	0.191	0.115	0.051	0.014	0.215
15 - <30	22	0.089	0.061	-0.030	0.208	1	0.006	0.079	0.000	0.162	0.041	0.053	0.000	0.144
>= 30	19	0.077	0.061	-0.043	0.197		0.000		0.000	0.000	0.032	0.025	0.000	0.082
Notes: [&] Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.														

Table H.4. Prevalence of different severities of apnea. AHI computed as total apneas plus hypopneas per hours of sleep. Hypopneas were defined based on only >=4% desaturation.

AHI[&]	Higher Risk N=246					Lower Risk N=158					Weighted			
	N	Rate	SE	(95%C.I.)		N	Rate	SE	(95%C.I.)		Rate	SE	(95%C.I.)	
<5	124	0.504	0.045	0.416	0.592	146	0.924	0.022	0.881	0.967	0.750	0.023	0.705	0.794
5 - <15	67	0.272	0.054	0.166	0.379	11	0.070	0.077	0.000	0.220	0.154	0.050	0.055	0.252
15 - <30	28	0.114	0.060	-0.004	0.231	1	0.006	0.079	0.000	0.162	0.051	0.053	0.000	0.154
>= 30	27	0.110	0.060	-0.008	0.228		0.000		0.000	0.000	0.046	0.025	0.000	0.094
Notes: [&] Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.														

Table H.5. Prevalence of different severities of apnea. AHI computed as total apneas plus hypopneas per hours of sleep. Hypopneas were defined based on only >=3% desaturation.

AHI[&]	Higher Risk N=246					Lower Risk N=158					Weighted			
	N	Rate	SE	(95%C.I.)		N	Rate	SE	(95%C.I.)		Rate	SE	(95%C.I.)	
<5	84	0.341	0.052	0.240	0.443	135	0.854	0.030	0.795	0.914	0.642	0.028	0.587	0.696
5 - <15	77	0.313	0.053	0.209	0.417	17	0.108	0.075	0.000	0.255	0.193	0.049	0.097	0.289
15 - <30	48	0.195	0.057	0.083	0.307	6	0.038	0.078	0.000	0.191	0.103	0.051	0.000	0.204
>= 30	37	0.150	0.059	0.035	0.266		0.000		0.000	0.000	0.062	0.024	0.000	0.110
Notes: [&] Apnea hypopnea index category (events/hours): mild 5-<15, moderate 15-<30, severe >=30.														

Table H.6. Prevalence of different severities of apnea. AHI computed as total apneas plus hypopneas per hours of sleep. Hypopneas were defined based on only >=2% desaturation.

Appendix I:

Analysis Of Variables

Determining Prevalence

	AHI³5 (AUC=0.726)	AHI³15 (AUC=0.727)	*AHI³30 (AUC=0.889)*
	b	**b**	**b**
Intercept	-16.9618	-15.6945	-23.8087
Age	0.2873	0.2071	0.3879
Age*Age	-0.00237	-0.00150	-0.00259
BMI	0.3187	0.2495	0.1479
BMI*BMI	-0.00302	-0.00153	0.00176

<u>Table I.1</u>. Quadratic effects apnea syndrome prevalence models for males. AUC is area under the receiver operating characteristic curve and is a summary measure of predictive value. The logistic regression models were estimated using weighted data to account for sampling design. Weights were adjusted for Epworth non-response (missing values). Weights used were 0.706 for higher risk and 1.457 for lower risk

	b	**OR**	
Intercept	-3.4272	-----	
Age 30-39†	-0.1080	0.90	
Age 40-49†	0.1504	1.16	
Age 50-59†	1.6564	5.24	
Age >=60†	0.4990	1.65	
BMI 25-<30‡	0.4656	1.59	
BMI 30-<35‡	0.8640	2.37	
BMI >=35‡	1.8860	6.59	
Notes:			
† Indicator variable for comparison with Age 20-29 years			
‡ Indicator variable for comparison with BMI < 25 kg/m^2			

<u>Table I.2</u>. Non-parametric logistic regression model for AHI≥5 episodes per hour plus Epworth ≥ 10.

Appendix J:

Questionnaire Instruments

Used In-Laboratory

Demographic Information

Information in this area will be kept Confidential

1. **Name:** _____

2. **Address:** _____

   _____

   _____

3. **Telephone:**_____

4. **Social Security #** _____-_____-_____

CONSENT FORM:

I agree to allow the answers on the following pages to be used for research purposes to help in the understanding of the causes of and treatments for sleep disorders. My understanding is that, if my answers are so used, my identity will be kept confidential.

I agree. ☐   **I do not agree.** ☐ **Please ✓ one box.**

Signed_____ **5. Today's Date :**_____/_____/_____

NOTE TO OUR PATIENTS:

Even if you do not agree to the above, please complete the questionnaire as it is an important part of the information required by your physician to properly care for you.

6a. Scheduled visit date (mm/dd/yy) ____ / ____ / ____

6b. D.O.B. (mm/dd/yy) ____ / ____ / ____

7. Sex: 1) Male ☐   2) Female ☐

8. Race: 1) White, Not of Hispanic origin ☐   4) Hispanic ☐
2) Black, not of Hispanic origin ☐   5) Other ☐
3) Asian or Pacific Islander ☐   6) American Indian
/ Alaskan Native ☐

9. Marital Status: 1) Married (or Common Law) ☐   3) Separated/Divorced ☐
2) Single ☐   4) Widow(er) ☐

10. Height (ins)_____   **11. Weight (lbs)**_____   **12. Collar Size**_____

13. Problems you are seeking help for from the Sleep Clinic (check all that apply):

1) Snoring_____   5) Restless sleep_____   9) Talk or walk in sleep_____

2) My breathing stops_____   6) Other heart disease_____   10) Other (specify)____

3) Sleepiness during the day__   7) Daytime breathing _____

problems_____

4) High Blood Pressure____   8) Can't fall asleep_____   11) Research subject___

14. Who referred you here? (check one)

1) My family doctor_____   4) A neurologist_____   7) Myself_____

2) A lung specialist_____   5) A psychiatrist_____   8) Other (specify)___

3) A heart specialist_____   6) A family member or friend__ 9) Research subject_

15. Highest level of schooling completed: (check one)

1) Grade School _____   4) High School_____   7) Graduate School__

2) Middle School_____   5) 2 Years College_____

3) Jr. High School_____   6) 4 Years College_____

16. Are you? (Check One)

1)Working Full Time ☐
5) Unemployed, not looking for work ☐

2)Working Part Time ☐
6) A Student ☐

3)Home Keeper ☐
7) Retired ☐

4)Unemployed, looking for work ☐
8) Unable to work ☐
*Reason*_____

17. Most recent occupation:_____

18. Do you work rotating night shift work?   1) YES ☐   2) NO ☐

19. Do you work steady night shift work?   1) YES ☐   2) NO ☐

J2
Pittsburgh sleep quality index

The following questions relate to your usual habits during the past *month only*. Your answers should indicate the most accurate reply for the majority of days and nights in the past month. Please answer all questions.

1. During the past month, when have you usually gone to bed at night?

   USUAL BED TIME ☐☐:☐☐ P.M. ☐ A.M. ☐

2. During the past month, how long has it usually taken you to fall asleep each night?

   NUMBER OF MINUTES ☐☐☐☐

3. During the past month, when have you usually gotten up in the morning?

   USUAL GETTING UP TIME ☐☐:☐☐ A.M. ☐ P.M. ☐

4. During the past month, how many hours of actual sleep did you get at night?

   HOURS OF SLEEP PER NIGHT ☐☐.☐☐

5. During the past month, how often have you had trouble sleeping because you.....

	1) Not during the past month	2) Less than once a week	3) Once or twice a week	4) Three or more times a week
(a) Cannot get to sleep within 30 minutes	☐	☐	☐	☐
(b) Wake up in the middle of the night or early morning	☐	☐	☐	☐
(c) Have to go to the bathroom	☐	☐	☐	☐
(d) Cannot breathe comfortably	☐	☐	☐	☐
(e) Cough or snore loudly	☐	☐	☐	☐
(f) Feel too cold	☐	☐	☐	☐
(g) Feel too hot	☐	☐	☐	☐
(h) Had bad dreams	☐	☐	☐	☐
(i) Have pain	☐	☐	☐	☐

(j) Other reason(s), please describe_____

6. During the past month, how would you rate your sleep quality overall? (Check one)

Very Good ☐   Fairly Good ☐   Fairly Bad ☐   Very Bad ☐

7. During the past month, how often have you had trouble staying awake while…

	1) Not during the past month	2) Less than once a week	3) Once or twice a week	4) Three or more times a week
Driving	☐	☐	☐	☐
Eating meals	☐	☐	☐	☐
Engaging in social activity	☐	☐	☐	☐

8. During the past month, how often have you taken medicine (prescribed or "over the counter") to help you sleep? (please check one)

1) Not during the past month	2) Less than once a week	3) Once or twice a week	4) Three or more times a week
☐	☐	☐	☐

9. During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

No problem at all____ Only a slight problem____ Somewhat of a problem____ A very big problem_____

10. Do you have a bed partner or roommate? (please check one)

1) No bed partner or roommate	2) Partner/roommate in other room	3) Partner/roommate in same room, but not in same bed	4) Partner/roommate in same bed
☐	☐	☐	☐

11. Ask him/her how often, in the past month, you have had (please check one per question):

		1) Not during the past month	2) Less than once a week	3) Once or twice a week	4)Three or more times a week
(a)	Loud snoring	☐	☐	☐	☐
(b)	Long pauses between breaths while asleep	☐	☐	☐	☐
(c)	Legs twitching or jerking	☐	☐	☐	☐
(d)	Episodes of disorientation or confusion during sleep	☐	☐	☐	☐
(e)	Other restlessness while you sleep, please describe_____				

Sleep Score Questionnaire

On average, how many day/nights during the last month have you had, or been told do the following WHILE ASLEEP OR TRYING TO SLEEP? (✓one per row)

	(0) Never	(1) Rarely (less than once a week)	(2) Sometimes (1-2 times per week)	(3) Frequently (3-4 times per week)	(4) Always (5-7 times per week)	(888) Do not know
1. Wheeze or whistle from your chest	☐	☐	☐	☐	☐	☐
2. Chest pain while in bed	☐	☐	☐	☐	☐	☐
3. Needed to wake up from sleep to use toilet 2 or more times	☐	☐	☐	☐	☐	☐
4. Lying awake during your sleep time feeling worried, depressed or sad.	☐	☐	☐	☐	☐	☐
5. Lying awake during your sleep time with thoughts racing through your mind	☐	☐	☐	☐	☐	☐

Over the last month, how frequently have you experienced DISTURBED SLEEP because of the following? (✓one per row)

	(0) Never	(1) Rarely (less than once a week)	(2) Sometimes (1-2 times per week)	(3) Frequently (3-4 times per week)	(4) Always (5-7 times per week)	(888) Do not know
6. Pain or physical discomfort	☐	☐	☐	☐	☐	☐
7. Noise	☐	☐	☐	☐	☐	☐
8. Heartburn during sleep time	☐	☐	☐	☐	☐	☐
9. Indigestion during sleep time	☐	☐	☐	☐	☐	☐

10. In what position do you normally sleep? (✓one)

1) My back_____
2) My side_____
3) My stomach_____
4) My back and side_____

5) My stomach and side_____
6) My stomach and back_____
7) All positions_____
8) Sitting up_____
888) Don't know_____

11.   **If you snore, over the LAST ONE MONTH, has your snoring usually been (✓one):**

    1)   Only slightly louder than heavy breathing_____

    2)   About as loud as mumbling or talking _____

    3)   Louder than talking _____

    4)   Extremely loud - can be heard through a closed door_____

    5)   I haven't snored in the last month_____

  888)   Don't know_____

12. How often do you drive? (✓one.)

    0)   Never_____   3)   3-4 days per week_

    1)   Rarely, less than one day per week___   4)   5-7 days per week____

    2)   1-2 days per week _____

13.   What is the average amount of time you spend driving on days that you drive? (✓one)

    0) Less than 10 minutes_____   3) At least 1 hour but less than 2 hours_

    1) At least 10 minutes but less than 1/2 hour   4) At least 2 hours_____

    2) At least 1/2 hour but less than 1 hour

For the following, please circle the number that best describes your feelings:

14.   How would you rate the overall quality of your life?

    1   **2**   **3**   **4**   **5**   **6**

Perfect Health   Miserable

15.   Are you hopeful that your quality of life will improve soon?

    1   **2**   **3**   **4**   **5**   **6**

Things will   Things will

Totally Improve   Never Improve

16. During the past month, how would you rate your level of energy?

    1   **2**   **3**   **4**   **5**   **6**

Fresh as a Daisy   Tired to Death

SMOKING HISTORY

17.   Have you ever smoked cigarettes? (NO means less than 20 packs in a lifetime, or less than one cigarette a day over a year) 1) YES____   2) NO _____

_____

If NO skip to 23, if YES continue:

18.   How old were you when you first started regular cigarette smoking?

   Age in years _____

19.   On average over the entire time you smoked, how many cigarettes did you smoke per day?

   Cigarettes per day_____

20.   Over the last month, have you smoked at least one cigarette per day?

   1) YES_____   2) NO_____

21.   If YES to question 20, approximate number of cigarettes per day_____

22.   If NO to question 20, How old were you when you stopped smoking?

   Age in years   _____

USE OF ALCOHOL

23.   Current Use: During the past month how many drinks **per week** have you had?

   Wine_____   Beer_____   Spirits_____   Total_____

24.   During the past month, on how many **days per week** have you had at least one drink?

   Please circle 1 of the following numbers: **0   1   2   3   4   5   6   7**

MEDICATIONS

25. **Have you ever taken the following?   (Please circle Yes or No)**

Blood Pressure Medicine	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Breathing Pills for Lung	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Breathing Pills for Heart	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Breathing Sprays	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Heart Pills	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Water Pills	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Sleeping Pills	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Antihistamine and/or decongestants	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Thyroid Medicine	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO
Tranquilizers	1) YES	2) NO	If yes, do you still take this?	1) YES	2) NO

26.   Other medications currently taking: __

_____

27.   Have you experienced a change in weight of greater than 10 lbs. over the last three (3) months?   1) YES _____   2) NO _____

a. If yes, how much? (+ or -) _____lbs.

b. If yes, was this a result of intentionally changing your diet? 1) YES ____   2) NO _____

FOR WOMEN ONLY:

28. Have you reached menopause? (circle one)

       1) YES   How old were you then: _____ (years)
       2) NO
       If NO, please answer questions 29 and 30:

29.   Approximately how many days in your menstrual cycle? (From first day of one period to first day of next period):   _____ Days

30.   When was your last period? (✓**one):**

       1) Currently menstruating_____   5) One month ago   _____
       2) One week ago   _____   6) More than one month but
                                                   less than six months ago_____
       3) Two weeks ago   _____   7) More than 6 months ago_____
       4) Three weeks ago   _____

THE EPWORTH SLEEPINESS SCALE

In contrast to just feeling tired, how likely are you to doze off or fall asleep in the following situations? (This refers to your usual life in recent times. Even if you have not done some of these things recently, try to work out how they would have affected you). Use the following scale to choose the most appropriate number for each situation:

0 = Would never doze
1 = Slight chance of dozing
2 = Moderate chance of dozing
3 = High chance of dozing

SITUATION   CHANCE OF DOZING

Sitting and reading   _____

Watching TV   _____

Sitting inactive in a public place   _____
(i.e. a theater or a meeting)

As a passenger in a car   _____
for an hour without break

Lying down to rest in the afternoon   _____
when circumstances permit

Sitting and talking to someone   _____

Sitting quietly after lunch without alcohol   _____

In a car, while stopping for a few minutes   _____
in traffic

FUNCTIONAL OUTCOMES OF SLEEP QUESTIONNAIRE (FOSQ)

Some people have difficulty performing everyday activities when they feel tired or sleepy. The purpose of this questionnaire is to find out if you generally have difficulty carrying out certain activities because you are too sleepy or tired. In this questionnaire, when the words "sleepy" or "tired" are used, it means the feeling that you can't keep your eyes open, your head is droopy, that you want to "nod off", or that you feel the urge to take a nap. These words do not refer to the tired or fatigued feeling you may have after you have exercised.

DIRECTIONS: Please put a (✓) in the box for your answer to each question. Select only **one** answer for each question. Please try to be as accurate as possible. All information will be kept confidential.

	(0) I don't do this activity for other reasons	(4) No difficulty	(3) Yes, a little difficulty	(2) Yes, moderate difficulty	(1) Yes, extreme difficulty
1. Do you have difficulty concentrating on the things you do because you are sleepy or tired?	☐	☐	☐	☐	☐
2. Do you generally have difficulty remembering things, because you are sleepy or tired?	☐	☐	☐	☐	☐
3. Do you have difficulty finishing a meal because you become sleepy or tired?	☐	☐	☐	☐	☐
4. Do you have difficulty working on a hobby (for example, sewing, collecting, gardening) because you are sleepy or tired?	☐	☐	☐	☐	☐
5. Do you have difficulty doing work around the house (for example, cleaning house, doing laundry, taking out the trash, repair work) because you are sleepy or tired?	☐	☐	☐	☐	☐

	(0) I don't do this activity for other reasons	(4) No difficulty	(3) Yes, a little difficulty	(2) Yes, moderate difficulty	(1) Yes, extreme difficulty
6. Do you have difficulty operating a motor vehicle for short distances (less than 100 miles) because you become sleepy or tired?	☐	☐	☐	☐	☐

7.

Do you have difficulty operating a motor vehicle for <u>long</u> distances (greater than 100 miles) because you become sleepy or tired?

☐ ☐ ☐ ☐ ☐

8.

Do you have difficulty getting things done because you are too sleepy or tired to drive or take public transportation?

☐ ☐ ☐ ☐ ☐

9.

Do you have difficulty taking care of financial affairs and doing paperwork (for example, writing checks, paying bills, keeping financial records, filling out tax forms, etc.) because you are sleepy or tired?

☐ ☐ ☐ ☐ ☐

10.

Do you have difficulty performing employed or volunteer work because you are sleepy or tired?

☐ ☐ ☐ ☐ ☐

11.

Do you have difficulty maintaining a telephone conversation, because you become sleepy or tired?

☐ ☐ ☐ ☐ ☐

12.

Do you have difficulty visiting with your family or friends in <u>your</u> home because you become sleepy or tired?

☐ ☐ ☐ ☐ ☐

13.

Do you have difficulty visiting with your family or friends in <u>their</u> home because you become sleepy or tired?

☐ ☐ ☐ ☐ ☐

14.

Do you have difficulty doing things for your family or friends because you are too sleepy or tired?

☐ ☐ ☐ ☐ ☐

	(4) No	(3) Yes, a little	(2) Yes, moderately	(1) Yes, extremely
15. Has your relationship with family, friends or work colleagues been affected because you are sleepy or tired?	☐	☐	☐	☐

In what way has your relationship been affected? _____

	(0) I don't do this activity for other reasons	(4) No difficulty	(3) Yes, a little difficulty	(2) Yes, moderate difficulty	(1) Yes, extreme difficulty
16. Do you have difficulty exercising or participating in a sporting activity because you are too sleepy or tired?	☐	☐	☐	☐	☐
17. Do you have difficulty watching a movie or videotape because you become sleepy or tired?	☐	☐	☐	☐	☐
18. Do you have difficulty enjoying the theater or a lecture because you become sleepy or tired?	☐	☐	☐	☐	☐
19. Do you have difficulty enjoying a concert because you become sleepy or tired?	☐	☐	☐	☐	☐
20. Do you have difficulty watching TV because you are sleepy or tired?	☐	☐	☐	☐	☐
21. Do you have difficulty participating in religious services, meetings or a group or club, because you are sleepy or tired?	☐	☐	☐	☐	☐

	(0) I don't do this activity for other reasons	(4) No difficulty	(3) Yes, a little difficulty	(2) Yes, moderate difficulty	(1) Yes, extreme difficulty
22. Do you have difficulty being as active as you want to be in the <u>evening</u> because you are sleepy or tired?	☐	☐	☐	☐	☐
23. Do you have difficulty being as active as you want to be in the <u>morning</u> because you are sleepy or tired?	☐	☐	☐	☐	☐
24. Do you have difficult being as active as you want to be in the <u>afternoon</u> because you are sleepy or tired?	☐	☐	☐	☐	☐
25. Do you have difficulty keeping pace with others your own age because you are sleepy or tired?	☐	☐	☐	☐	☐

	(1) Very Low	(2) Low	(3) Medium	(4) High	
26. How would you rate your general level of activity?	☐	☐	☐	☐	

	(0) No intimate or sexual relationship	(4) No	(3) Yes, a little	(2) Yes, moderately	(1) Yes, extremely
27. Has your intimate or sexual relationship been affected because you are sleepy or tired?	☐	☐	☐	☐	☐

	(0) I don't engage in sexual activity for other reasons	(4) No	(3) Yes, a little	(2) Yes, moderately	(1) Yes, extremely
28.	☐	☐	☐	☐	☐
29.	☐	☐	☐	☐	☐
30.	☐	☐	☐	☐	☐

Thank you for completing this questionnaire.

THE STANFORD SLEEPINESS SCALE

Please (circle) the number next to the statement that best describes how sleepy you feel <u>right now</u>.

1. Feeling active and vital; alert; wide awake

2. Functioning at a high level, but not at peak; able to concentrate

3. Relaxed; awake not at full alertness; responsive

4. A little foggy; not a peak; let down

5. Fogginess; beginning to lose interest in remaining awake; slowed down

6. Sleepiness; prefer to be lying down; fighting sleep; woozy

7. Almost in reverie; sleep onset soon; lost struggle to remain awake

J.7

KAROLINSKA SLEEPINESS SCALE

WE WOULD BE GRATEFUL IF YOU COULD ANSWER THE FOLLOWING QUESTION ABOUT A DISORDER CALLED SLEEP APNEA:

a.Have you ever suspected, or been told, that you might have sleep apnea?

☐ ☐
Yes No

b.Have you ever had a sleep study to assess whether you might have sleep apnea?

☐ ☐
Yes No

c.Have you ever been treated for sleep apnea?

☐ ☐
Yes No

If No, this questionnaire is completed. Thank you.

If Yes, indicate which of the following treatment applies:

☐ ☐ ☐ ☐
CPAP Intra-oral Surgery Other
device

d.Are you still being treated for sleep apnea?

☐ ☐
Yes No

If Yes, indicate what the current treatment is:

☐ ☐ ☐
CPAP Intra-oral Other
device

e.Have you ever stopped treatment for sleep apnea?

☐ ☐
Yes No

If Yes, please indicate reason: _____

_____

_____

WE WOULD BE GRATEFUL IF YOU COULD ANSWER THE FOLLOWING QUESTIONS ABOUT YOUR EMPLOYMENT AS A COMMERCIAL DRIVER:

a.Are you currently employed as a commercial driver?

☐ ☐
Yes No

b.What type of commercial vehicle to you drive (check all that apply):

1) Local truck

☐

2) Van

☐

3) Bus

☐

4) Construction vehicle

☐

5) Tractor-Trailer

☐

6) Other (specify): _____

☐

c.In the last month, how days/week have you driven a commercial vehicles on average?

_____

d.In the last month, how many hours/day, on average, did you drive a commercial vehicle?

___ ___

e.In the last month, which of the following best describes what you do?

☐   ☐   ☐

Drive only   Drive only   Drive both

days   at night   day and night

f. In the last month, which of the following best describes your daily shift pattern?

☐   ☐   ☐

Regular   Regular, but rotates   Irregular depending on job

g.In the last month, how many miles do you estimate you have driven?

___ ___ , ___ ___ ___ miles

FMCSA

Federal Motor Carrier Safety Administration

DEPARTMENT OF TRANSPORTATION · UNITED STATES OF AMERICA

For more information on the Federal Motor Carrier Safety Administration and the Office of Research and Technology, check our website at www.fmcsa.dot.gov.