Sleep Apnea Syndrome*

A Critical Review of the Apnea Index as a Diagnostic Criterion

David T. R. Berry, M.A.; Wilse B. Webb, Ph.D.; and A. Jay Block, M.D., F.C.C.P.

The utility of the apnea index (number of apneic events per hour of sleep) in diagnosing sleep apnea syndrome is reviewed. Data from currently extant reports indicate that many otherwise normal, aging subjects may be classified as having sleep apnea syndrome based on the current diagnostic threshold of five apneic episodes per hour. A \( \chi^2 \) analysis suggests a relationship between age and level of sleep apnea. Several other reports indicate that use of a threshold of five apneic episodes per hour does not reliably predict increased health risk or somnolence in aging subjects. Adjustment of the apnea index, based on studies of aging normal subjects and of aging patients with sleep apnea syndrome is necessary to ensure reliable results in clinical and research applications.

During the past decade the sleep apnea syndrome has become widely recognized as a clinical entity. According to a recent nosology of sleep disorders published by the Association of Sleep Disorder Centers, sleep apneas are pauses in nocturnal breathing lasting ten seconds or longer, whereas sleep apnea syndrome is "a potentially lethal condition characterized by multiple obstructive or mixed apneas with repetitive episodes of inordinately loud snoring and excessive daytime sleepiness..." Since the syndrome is defined clinically, the number of apneas is only one factor among a group of abnormalities, albeit one that can be measured and quantified. Because of the ease of quantification of apneas, increased importance has been attributed to the apnea index in diagnosing the sleep apnea syndrome. It seems appropriate, therefore, to evaluate the utility of this index for diagnosis.

Gullemiault et al\(^4\) have published the most commonly accepted threshold for considering sleep apnea activity to be pathologic. Initially, sleep apnea syndrome was diagnosed (presumably symptomatically) in a group of patients who had no fewer than 45 episodes of apnea recorded per night. Recordings for apnea were also made in 20 asymptomatic subjects between 40 and 60 years of age. The men had a mean of seven apneic events per night, and the women had 2.1. The highest individual number of apneic episodes recorded in this group was 25. These authors\(^5\) proposed: "A sleep apnea syndrome is diagnosed if during 7 h of nocturnal sleep, at least 30 apneic episodes are observed..." Use of this score with the entire sample of both normal and ill subjects resulted in correct classification of symptomatically diagnosed sleep apnea syndrome in the ill patients and determined that the normal subjects were free of sleep apnea syndrome. In the same report an apnea index (number of apneic episodes per hour of sleep) of five or greater was proposed as a more sensitive threshold for diagnosing sleep apnea syndrome. These criteria are widely used according to recent clinical and research publications.

With the adoption of these thresholds, nocturnal apnea became implicitly conceptualized as a sign of sleep apnea syndrome. Clinically, sleep apnea syndrome might be diagnosed when apneic activity exceeded either of these cutting scores. Meehl\(^6\) indicates that any cutting score used to diagnose a condition should be evaluated for predictive validity. The overall "hit rate" (true and false positives as well as true and false negatives) must be considered in establishing the validity of a cutting score. That is, given known disease or health, how often does the cutting score correctly classify a subject? As regards the cutting score for sleep apnea, how often does the score assign a diagnosis of sleep apnea syndrome to those who by other clinical criteria would not be classified as having sleep apnea syndrome? Excessive false classifications might render the cutting score suspect and indicate a need for the score to be adjusted.

One method of evaluating this proposed criterion of five apneic episodes per hour as an index of sleep apnea syndrome would be to determine whether such an index classifies persons as having sleep apnea syndrome who otherwise do not meet the clinical criteria described in the introduction. A review of the literature discloses a number of measurements of sleep apnea in subjects who were described as normal. These subjects represent an ideal population for eval...
ating false-positive rates of the current apnea index.

**Materials and Methods**

In our search for reports on apnea levels in normal subjects, we examined the Sleep Research abstracts from 1973 to 1982. We selected journal reports publishing quantitative data on sleep apneas in at least ten normal persons. Normal subjects were those so designated in the methods section of each report. Logically, this excludes subjects with the serious symptoms of sleep apnea syndrome.

Two of the earlier studies identified in the literature review were apparently subsamples of more extensive later reports and hence are referred to in the later reports.

Webb studied 20 normal men (mean age, 44 years). Block et al studied 30 asymptomatic men (mean age, 38 years) and 19 asymptomatic women (mean age, 29 years). In a follow-up study, Block et al studied 20 postmenopausal women (mean age, 59 years). Carskadon and Dement studied 40 elderly men and women who were reportedly the healthiest in their age group. In one major study, McGinty et al examined 26 healthy men (mean age, 64 years), but failed to present apnea data alone. Bixler et al studied an age-stratified sample of 100 healthy subjects. Krieger et al examined one young and one old group of normal subjects. The young group consisted of ten young men (mean age, 24 years) and ten young women (mean age, 24 years); the older group included nine women (mean age, 64 years) and 11 men (mean age, 67 years). Smallwood et al studied one group of ten young men (mean age, 25 years) and an older group of 24 men and six women (mean ages, 60 and 66 years, respectively).

**Results**

Table 1 summarizes the data on levels of sleep apnea gleaned from these reports. Where possible, subsamples from studies with wide variance in age have been presented separately. This allowed a rank ordering of the subsamples by age (eg, Bixler et al). Some of the information presented in this table was calculated from tabled data on the original report; a footnote marks this extraction. The total percentages of subjects experiencing any episodes of sleep apnea and percentages of subjects experiencing significant levels of sleep apnea as judged by the criteria we were testing are listed separately.

Table 1 indicates that false-positive rates (ie, apnea index greater than five in these normal subjects) are not very high in younger healthy subjects; however, a dramatic increase in false-positives apparently takes place in samples including large numbers of subjects past 60 years of age. To explore this hypothesis, we obtained data on age and apnea activity in individual subjects from the cited published reports and from Drs. Robert Smallwood and A. Jay Block. A x2 test of independence of the two factors of age (older or younger than 60 years) and level of sleep apnea activity (above or below the threshold value of five apneic episodes per hour) was highly significant (x2 = 35.4; p<0.001), resulting in a rejection of the hypothesis of independence of age and diagnostically significant level of sleep apnea.

**Discussion**

These results suggest a relationship between increasing age and the level of sleep apnea activity. Because these subjects were normal and healthy, it is difficult to view this as a pathologic process. That is, it does not seem logical to assume that subjects selected for health have such high rates of serious illness; rather, the current cutting scores must be incorrectly classifying these subjects. Thus, the current cutting scores have high false-positive rates in normal elderly populations.

Are there other convergent data bringing up the utilization of this criterion in aging subjects into question? Ancoli-Israel et al studied 24 elderly subjects who had sleep and breathing complaints. Comparing findings in the subjects with sufficient apnea to diagnose sleep apnea syndrome with findings in the remaining subjects, they found no significant differences between the two groups on reports of heart disease, hypertension, obesity, or sinus problems. Thus, an apnea index greater than five failed to predict significant differences in these two groups on symptoms supposedly intimately associated with the sleep apnea syndrome. Krieger and Kurtz observed 20 elderly subjects for three years and found that initial apnea index classifications of five or greater did not predict an increase in health risk across this period. Carskadon et al studied 24 elderly subjects for signs of daytime sleepiness. Calculations from their tabled data indicate that those subjects with more than five respiratory disturbances...
per hour were not significantly sleepier than the remaining subjects (based on comparison of mean multiple sleep latency scores for each group). Thus, classification of elderly subjects by these cutting scores does not reliably distinguish them on health variables or daytime somnolence—two presumed components of the sleep apnea syndrome.

An additional source of concern in applying these cutting scores to aging subjects lies in the original procedure for validation. Apparently, subjects over 60 years of age were not included in the original contrast (normal) sample. Failure to cross-validate cutting scores on samples other than those included in the original calculations may generate spurious results when the scores are used to classify subjects in the new population. Because aging subjects were not included in the original sample for validation, application of these cutting scores to elderly subjects may be statistically problematic.

Thus, several lines of evidence suggest that current diagnostic criteria for the sleep apnea syndrome diagnostic criteria (more than 30 apneic episodes per night or more than five apneic episodes per hour of sleep) may be inappropriate for aging subjects. First, large percentages of otherwise healthy aging subjects may be classified as having sleep apnea syndrome under the current apnea index. In more general samples of elderly subjects, these criteria do not predict relative health or somnolence accurately. Finally, the lack of aging subjects in the original sample for validation makes application of these cutting scores questionable in this group.

Further research is needed to establish reliable cutting scores for assistance in the diagnosis of sleep apnea syndrome in the elderly. Until this is available, clinicians should be aware of the possibility of improperly classifying aging subjects by the apnea index alone. Researchers as well should be cognizant of possible heterogeneous groups formed using this crite-

rion in aging subjects.

ACKNOWLEDGMENTS: The helpful reviews and suggestions of C. M. Carswell and Dr. C. Belar are gratefully acknowledged.

REFERENCES
1 Association of Sleep Disorder Centers. Diagnostic classification of sleep and arousal disorders. 1st ed. Sleep 1979; 2:1-137
14 Ancoli-Israel S, Kripke D, Mason W, Messin S. Sleep apnea and nocturnal myoclonus in a senior population. Sleep 1981; 4:349-58