

SCIENTIFIC INVESTIGATIONS

## Relationships Between the Pittsburgh Sleep Quality Index (PSQI), Epworth Sleepiness Scale (ESS), and Clinical/Polysomnographic Measures in a Community Sample

Daniel J. Buysse, M.D.<sup>1</sup>; Martica L. Hall, Ph.D.<sup>1</sup>; Patrick J. Strollo, M.D.<sup>1</sup>; Thomas W. Kamarck, Ph.D.<sup>1</sup>; Jane Owens, Ph.D.<sup>1</sup>; Laisze Lee, M.S.<sup>1</sup>; Steven E. Reis, M.D.<sup>2</sup>; Karen A. Matthews, Ph.D.<sup>1,3</sup>

From the Sleep Medicine Institute and Neuroscience Clinical and Translational Research Center (University of Pittsburgh) and the Pittsburgh Mind Body Center (University of Pittsburgh and Carnegie Mellon University). <sup>1</sup>Department of Psychiatry, <sup>2</sup>Department of Medicine, <sup>3</sup>Department of Psychology, University of Pittsburgh, Pittsburgh, PA

**Study Objectives:** 1) To characterize PSQI and ESS scores, and their relationship to each other, in an adult community sample; 2) To determine whether PSQI and ESS scores, in combination with each other, were associated with distinct demographic, clinical, and sleep characteristics.

**Methods:** The PSQI, ESS, clinical rating scales, sleep diaries, actigraphy, and home polysomnography were collected from 187 community-dwelling adults (mean age 59.5 years, 47.1% women, 41.2% African Americans) as part of a study investigating novel cardiovascular risk factors. Correlations, cluster analysis, principal components analysis, MANOVA, ANOVA, and regressions were used to characterize the relationships between the PSQI, ESS, and other study variables

**Results:** Mean PSQI score was 6.3 (3.4), and mean ESS score was 8.2 (3.9). PSQI and ESS correlated weakly with each other ( $r = 0.16$ ,  $p = 0.03$ ), but segregated from each other on principal components analysis. Groups of participants categorized by either cluster analysis of PSQI and ESS scores, or by scores above or below traditional cut-off values, differed from each other on psychological/stress symp-

toms and quantitative and qualitative sleep diary measures, but not on actigraphic or polysomnographic measures. Specifically, higher PSQI scores were associated with female sex, greater psychological distress, and greater sleep disturbance on sleep diaries.

**Conclusions:** The PSQI and ESS measure orthogonal dimensions of sleep-wake symptoms, but neither is related to objective sleep measures. The PSQI is more closely related to psychological symptom ratings and sleep diary measures than the ESS. These instruments are not likely to be useful as screening measures for polysomnographic sleep abnormalities.

**Keywords:** Sleep, sleep quality, sleepiness, Pittsburgh Sleep Quality Index, Epworth Sleepiness Scale

**Citation:** Buysse DJ; Hall ML; Strollo PJ; Kamarck TW; Owens J; Lee L; Reis E; Matthews KA. Relationships between the Pittsburgh Sleep Quality Index (PSQI), Epworth Sleepiness Scale (ESS), and clinical/polysomnographic measures in a community sample. *J Clin Sleep Med* 2008;4(6):563-571.

Sleep quality and daytime sleepiness are salient and clinically relevant dimensions of sleep-wake function. Poor sleep quality and insomnia symptoms have been associated with worse health, increased health care costs and utilization, absenteeism from work, and increased risk for psychiatric disorders, including depression.<sup>1</sup> Daytime sleepiness has been associated with increased risk of motor vehicle accidents, worse physical health, and increased mortality risk.<sup>2</sup> Although sleep and sleepiness can be measured by objective means such as polysomnography (PSG) and the multiple sleep latency test (MSLT), these methods are often impractical as clinical screening or research tools. Self-report questionnaires are most commonly used to assess sleep quality and daytime sleepiness. Many different instruments have been developed to measure sleep quality, insomnia,

and daytime sleepiness (for review, see<sup>3</sup>), but 2 of the most widely-used are the Pittsburgh Sleep Quality Index (PSQI)<sup>4</sup> and the Epworth Sleepiness Scale (ESS).<sup>5,6</sup> A search of the ISI Web of Knowledge Citation Index in January 2008 identified over 900 publications citing the PSQI, and over 1500 citing the ESS. Despite their widespread use, however, relatively little attention has been paid to how the PSQI and ESS relate to each other, or to other clinical and sleep measures.

The PSQI is a 19-item self-rated questionnaire for evaluating subjective sleep quality over the previous month. The 19 questions are combined into 7 clinically-derived component scores, each weighted equally from 0-3. The 7 component scores are added to obtain a global score ranging from 0-21, with higher scores indicating worse sleep quality. The clinical and psychometric properties of the PSQI have been formally evaluated by several research groups.<sup>4,7-9</sup> The PSQI has a sensitivity of 89.6% and specificity of 86.5% for identifying cases with sleep disorder, using a cut-off score of 5. Validity is further supported by similar differences between groups using PSQI or polysomnographic sleep measures. The PSQI has been translated into 48

Submitted for publication February, 2008

Accepted for publication July, 2008

Address correspondence to: Dr. Buysse: E-1127 WPIC, 3811 O'Hara St., Pittsburgh, PA 15213; Tel: (412) 246-6413; Fax: (412) 246-5300; E-mail: buyssedj@upmc.edu

languages and has been used in a wide range of population-based and clinical studies.

The ESS consists of 8 self-rated items, each scored from 0-3, that measure a subject's habitual "likelihood of dozing or falling asleep" in common situations of daily living. No specific time frame is specified. The ESS score represents the sum of individual items, and ranges from 0-24. Values >10 are considered to indicate significant sleepiness. The ESS is sensitive to change in clinical status, as evidenced by improvements following treatment of sleep apnea with continuous positive airway pressure (CPAP).<sup>10</sup> Psychometric analyses of the ESS support its internal consistency and unidimensionality, although factor analyses have indicated some variation in the number of identifiable factors.<sup>11-13</sup> The ESS has been translated into 52 languages and has also been used in a wide range of population-based and clinical studies.

Despite their widespread use, we found few published studies that systematically examined whether the PSQI and ESS measure orthogonal dimensions of sleep-wake function. No significant correlation was found between PSQI and ESS in a small study of patients with sporadic adult-onset ataxia,<sup>14</sup> a study of patients with Parkinson disease and healthy controls,<sup>15</sup> a study of insomnia and control subjects,<sup>16</sup> or a study of hypersomnolence in truck drivers.<sup>17</sup> We found no other data on how scores for each scale relate to each other; how, in aggregate, they relate to other important clinical measures; or whether distinctive subgroups can be identified by the combination of PSQI and ESS scores. For example, it would be reasonable to hypothesize that low scores on both instruments would be associated with the most "normal" clinical features, that high scores on both would be associated with the most "abnormal" clinical features, and that high scores on one instrument (but not the other) would be associated with different types of clinical features.

The current analyses address these issues in a sex and race-balanced community sample of middle-aged men and women. Specifically, goals for the present paper were: (1) To characterize the distributions of PSQI and ESS scores, and their relationship to each other; and (2) To determine whether subgroups of participants characterized by PSQI and ESS scores differed in terms of demographic, clinical, sleep, and other physiological characteristics.

## METHODS

### Overview

The PSQI, ESS, and other self-report and physiological data were collected from a community sample of mid-life men and women, including approximately equal numbers of Caucasians and African Americans. Participants were initially recruited as part of a larger epidemiological study, Heart Strategies Concentrating on Risk Evaluation (HeartSCORE), a single-site prospective community-based participatory research cohort study investigating the mechanisms accountable for population disparities in cardiovascular risk and attempting to decrease this risk via a community-based intervention. HeartSCORE assessed biological markers, psychological factors, and health behaviors relevant to cardiovascular disease and classified individuals into Framingham Risk groups, a classification which

is based on smoking, cholesterol, blood pressure, age group, and sex.<sup>18</sup> The current sleep study protocol (SleepSCORE) was an independently funded study that recruited participants and used data from the parent HeartSCORE study. SleepSCORE included 10 days of sleep diary and actigraphy data; retrospective clinical questionnaires regarding sleep-wake, psychological, psychosocial, and stress-related experiences and symptoms; two 24-hour period of ambulatory blood pressure monitoring and urine collection for catecholamines; and 2 nights of home PSG. Both HeartSCORE and SleepSCORE studies were approved by the University of Pittsburgh Biomedical Institutional Review Board, and all participants provided written informed consent.

### Participants

Participants for the current study were recruited from the larger HeartSCORE sample, as described above. Eligibility criteria for HeartSCORE included age 45 to 75 years, residence in the greater Pittsburgh metropolitan area, and absence of comorbid conditions expected to limit life expectancy to < 5 years. Exclusion criteria for the sleep study included pregnancy; use of continuous positive airway pressure treatment for sleep disordered breathing; medication for sleep problems on a regular basis; night work; medication for diabetes; and prior diagnosis of stroke, myocardial infarction, or interventional cardiology procedures. For these analyses we included the first 187 participants out of a targeted enrollment of 225. By design, we recruited approximately equal numbers of men and women, Caucasians and African Americans, and those with high/moderate vs. low Framingham Risk equation scores from the parent study. Participants were compensated \$200 for study completion.

### Measures

The PSQI and ESS, described above, constituted the major outcome variables for this study. We examined potential clinical and physiological correlates of PSQI and ESS scores grouped into 8 domains, designed to ensure a multimodal assessment of sleep-wake function, as well as a broad range of psychological and physiological characteristics relevant to cardiovascular risk and sleep:

Demographic and clinical variables included age, sex, race, and body mass index (BMI). Coronary heart disease (Framingham) risk categories were defined on the basis of participant age, cholesterol, blood pressure, and diabetes and smoking status.<sup>18</sup>

Self-report measures were collected to characterize participants' sleep, psychological symptoms, and stress. Specific constructs and measures were selected on the basis of previous research suggesting their role as potential risk factors for cardiovascular disease; examination of these risk factors was one of the major aims of SleepSCORE. Self-report measures included the Ford Insomnia Response to Stress Test (FIRST),<sup>19</sup> a dispositional questionnaire assessing the tendency to sleep poorly at times of stress; the Perceived Stress Scale (PSS),<sup>20</sup> a measure of perceived stress in the last month; the Ongoing Stress Scale,<sup>21</sup> a measure of the number of ongoing stressors that lasted longer than 12 months and were rated as very stress-

ful; the Center for Epidemiological Studies Depression (CESD) scale,<sup>22</sup> which was scored without the sleep item to prevent spurious associations; the Spielberger Trait Anxiety Inventory (STAI)<sup>23</sup>; the Spielberger Trait Anger Scale<sup>23</sup>; the Life Orientation Test (LOT), a measure of optimism<sup>24</sup>; 27 items from the Cook-Medley Hostility Inventory,<sup>25</sup> which measure cynical attitudes, angry affect, and aggressive responding and have been associated with cardiovascular risk; and the Unfair Treatment Scale,<sup>26</sup> which measures exposure to a range of unpleasant, rude circumstances in the past several weeks.

The Pittsburgh Sleep Diary<sup>27</sup> was also collected for 10 consecutive days and nights as a prospective, ecologically valid measure of sleep self-report. Two domains of sleep diary variables were considered in these analyses: quantitative sleep parameters (sleep latency, total sleep time, wake after sleep onset [WASO], sleep efficiency) and qualitative variables (sleep quality, feeling rested, and overall feeling in the morning).

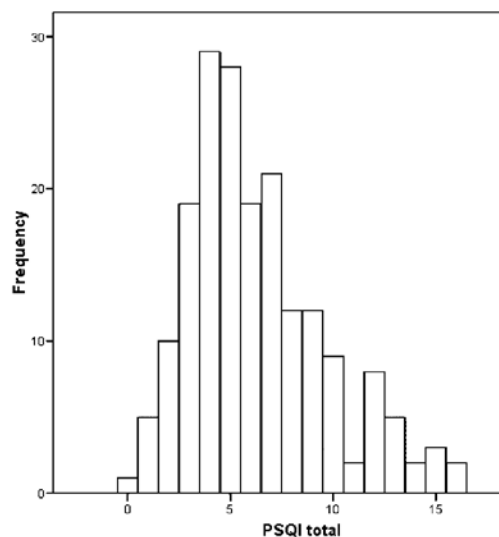
Wrist actigraphy was used to provide an objective longitudinal assessment of rest-activity patterns, from which sleep-wake patterns can be inferred. Actigraphy data were collected in one-minute epochs using Actiwatch 64 devices (Respironics, Inc.) for 10 consecutive days. Sleep-wake times from participants' sleep diaries were entered to calculate summary actigraphy variables using validated MiniMitter software algorithms. We chose 2 specific variables, total sleep time and fragmentation index, to represent sleep amount and sleep continuity in these analyses.

Home PSG was collected with Compumedics Siesta monitors for 2 consecutive nights, concurrent with actigraphy and sleep diary. The PSG montage included bilateral central and occipital electroencephalogram (EEG) channels, bilateral electrooculograms (EOG), bipolar submental electromyograms (EMG), and one channel of electrocardiogram (EKG). Additional channels for sleep disordered breathing were recorded on the first night, including nasal pressure, oral-nasal thermistors, inductance plethysmography, and fingertip oximetry. Periodic limb movements were assessed using bilateral anterior tibialis EMG. PSG technologists scored sleep records in 20-second epochs using standard sleep stage scoring criteria<sup>28</sup> and American Academy of Sleep Medicine<sup>29</sup> definitions for apnea and hypopnea events. PSG outcomes included subdomains of sleep continuity (total sleep time, sleep latency, WASO, number of awakenings, sleep efficiency), sleep architecture (percentages of stage 1, stage 2, stage 3+4, and REM), and sleep disorders (apnea-hypopnea index, desaturation index, periodic limb movement arousal index). In order to be considered usable, each PSG had to include  $\geq 4$  hours of concurrent data for sleep staging (EEG, EOG, EMG) and sleep disordered breathing (oximetry plus one additional channel). If a study did not meet these criteria, a repeat night was scheduled. Based on these criteria, usable PSG data were ultimately obtained in 98.9% of study participants.

### Statistical Analysis

The aims of our analyses, and therefore our statistical methods, were primarily descriptive and exploratory in nature. For Aim 1, descriptive statistics were used to characterize PSQI and ESS scores in the sample, and Pearson correlation coefficient

### A: FREQUENCY DISTRIBUTION OF PSQI SCORES



### B: FREQUENCY DISTRIBUTION OF ESS SCORES

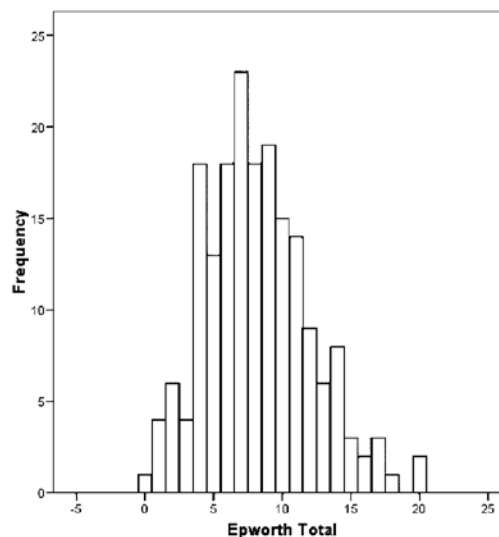


Figure 1—Frequency distributions of the PSQI (A) and ESS (B).

was used to examine the relationship between these scores. We conducted an exploratory factor analysis using the seven component scores of the PSQI and the 8 items of the ESS to determine whether the two instruments measure orthogonal dimensions. Specifically, we conducted a principal components analysis with Varimax rotation, and examined components with eigenvalues greater than 1.0. Variables with factor loadings  $>0.40$  were considered to load on a particular factor.

For Aim 2, we used 2 methods to characterize subgroups within the overall sample on the basis of their conjoint PSQI-ESS scores. First, we conducted an exploratory cluster analysis using PSQI and ESS scores to empirically identify 4 subject groups, using K-means cluster analysis. This analysis makes no a priori assumptions about specific PSQI or ESS scores that define subgroups. Second, we used the published cut-off scores for each instrument ( $> 5$  on the PSQI,  $> 10$  on the ESS) to define 4 subgroups that had elevated scores on each, normal scores

**Table 1**—Principal Components Analysis of PSQI Components and ESS Items<sup>1</sup>

| Component / Item                | Component Number |       |       |       |       |
|---------------------------------|------------------|-------|-------|-------|-------|
|                                 | 1                | 2     | 3     | 4     | 5     |
| PSQI Sleep Duration             | —                | —     | 0.811 | —     | —     |
| PSQI Sleep Disturbances         | —                | 0.732 | —     | —     | —     |
| PSQI Sleep Latency              | —                | 0.678 | —     | —     | —     |
| PSQI Daytime Dysfunction        | —                | 0.522 | —     | —     | —     |
| PSQI Habitual Sleep Efficiency  | —                | —     | 0.795 | —     | —     |
| PSQI Sleep Quality              | —                | 0.608 | —     | —     | —     |
| PSQI Medications to Sleep       | —                | —     | —     | —     | 0.905 |
| ESS Item 1 (Sitting reading)    | —                | —     | —     | 0.693 | —     |
| ESS Item 2 (Watching TV)        | —                | —     | —     | 0.849 | —     |
| ESS Item 3 (Inactive in public) | 0.612            | —     | —     | —     | —     |
| ESS Item 4 (Car passenger)      | 0.680            | —     | —     | —     | —     |
| ESS Item 5 (Lying down)         | 0.636            | —     | —     | —     | —     |
| ESS Item 6 (Sitting talking)    | 0.677            | —     | —     | —     | —     |
| ESS Item 7 (Sitting quietly)    | 0.539            | —     | —     | —     | —     |
| ESS Item 8 (Car in traffic)     | 0.714            | —     | —     | —     | —     |

<sup>1</sup>Values shown are weightings on five rotated components (factors) from principal components analysis with eigenvalues >1.0. Only component weightings >0.5 are shown.

on each, or one elevated and one normal score. This analysis provides a clinical frame of reference for understanding potential subgroups. For both methods, we contrasted the resultant groups on clinical and sleep domains identified above. We examined the PSQI and ESS groups according to sex, gender, and Framingham risk score using chi-square tests. Groups were then compared using multivariate analysis of variance for each domain after controlling for sex, race, and Framingham risk groups. When the MANOVA result was significant, we conducted individual ANOVAs on each individual measure within the domain, again controlling for sex, race, and Framingham risk. Significant ANOVA results were followed by post hoc contrasts using Tukey's b test to determine which groups were different from each other. As discussed below, results using the cluster analysis-defined groups and clinical cut-off-defined groups were quite similar. Therefore, we present results of the cluster analysis-defined groups in greater detail, since they were defined empirically based on data in this participant sample.

As a final analysis to address Aim 2, we also conducted regression analyses using the PSQI, ESS, and their cross-product as independent variables, and each of the individual clinical and laboratory measures as dependent variables. This analysis permits a direct examination of whether the PSQI, ESS, or some nonlinear combination of the 2 are related to the dependent variables.

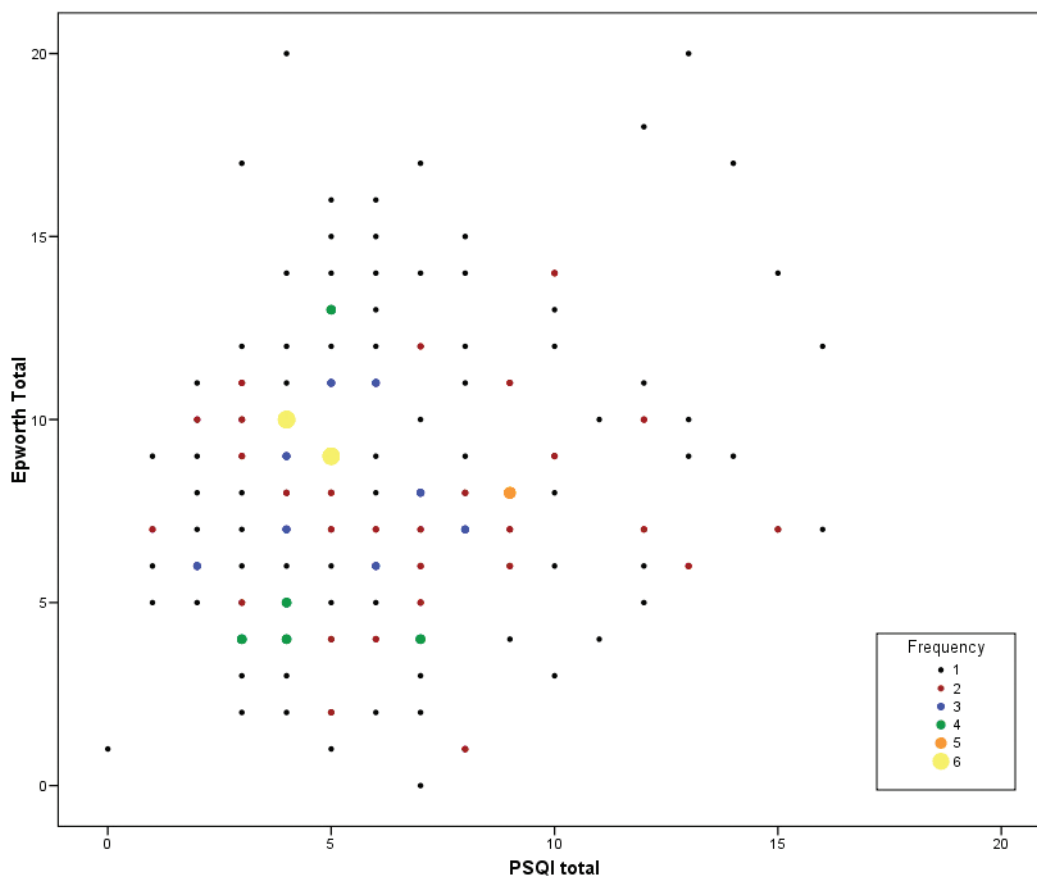
We addressed the issue of multiple comparisons by conducting omnibus multivariate tests (MANOVA) followed by univariate tests (ANOVAs). MANOVAs were conducted for 7 of the outcome domains: Self-report measures, quantitative sleep diary parameters, qualitative sleep diary variables, wrist actigraphy, and PSG measures of sleep continuity, sleep architecture, and sleep disorders (apnea, PLMs). To further correct for multiple comparisons, the  $\alpha$  level for the 7 MANOVAs was set at  $0.05/7 = 0.007$ . Given the exploratory nature of the follow-up ANOVAs, we simply report the unadjusted p values. Because the regression analyses were conducted on each of the 40 dependent measures rather than multivariate measures, we adopted a more conservative  $\alpha$  value of 0.01.

## RESULTS

The sample of 187 subjects had a mean age of 59.5 (SD 7.2) and included 88 (47.1%) women and 77 (41.2%) African Americans. Mean PSQI score for the sample was 6.3 (SD 3.4, median 6), and mean ESS score was 8.2 (SD 3.9, median 8). Frequency histograms for the PSQI and ESS (Figure 1) show approximately normal distributions for each, with slight right skewness. Ninety-five subjects (50.8%) had PSQI scores > 5, indicating poor sleep quality, and 48 subjects (25.7%) had ESS scores > 10, indicating daytime sleepiness. A scatter plot of PSQI and ESS scores (Figure 2) shows a weak relationship between the 2 scales. The Pearson correlation coefficient was 0.16 ( $p = 0.03$ ), indicating less than 3% shared variance between the instruments. The "Daytime Dysfunction" component of the PSQI may measure a similar construct to the ESS, as suggested by a correlation coefficient of 0.34 ( $p < 0.001$ ). When the Daytime Dysfunction component was subtracted from the PSQI global score, the correlation coefficient between PSQI and ESS was 0.10 ( $p = 0.16$ ). Principal components analysis for the 7 component scores of the PSQI and the 8 items of the ESS further supported the independence of these measures. Five factors had eigenvalues greater than 1.0 and accounted for 61% of the cumulative variance. There was no overlap of PSQI components and ESS questions on these components when considering values of 0.5 or greater for individual items (Table 1).

K-means cluster analysis of PSQI and ESS scores converged on a 4-cluster solution. The 4 groups of subjects could broadly be characterized by high/low PSQI and ESS scores (Figure 2; Table 2). Thus, Cluster 1 ( $n = 73$ ) was characterized by low PSQI and ESS scores; Cluster 2 ( $n = 37$ ) had low PSQI and high ESS scores; Cluster 3 ( $n = 54$ ) had high PSQI and low ESS scores; and Cluster 4 ( $n = 23$ ) had high PSQI and ESS scores. ANOVAs indicated significant differences between clusters for both PSQI and ESS scores, and post hoc tests showed each cluster to be significantly different from each other cluster on both the PSQI and ESS. Analyses based on the 4 empirically defined groups from cluster analysis showed no significant differences





**Figure 2**—Scatterplot of PSQI and ESS scores. Different symbol shading and sizes correspond to the number of subjects at a particular PSQI-ESS score.

in age or body mass index. Chi-square analysis for sex distribution indicated a significant difference between groups ( $\chi^2 = 17.5$ ,  $p = 0.001$ ), with women overrepresented in the clusters with high PSQI scores. There was no significant difference in race among the 4 clusters ( $\chi^2 = 6.4$ ,  $p = 0.10$ ), or in the distribution of Framingham risk categories ( $\chi^2 = 4.0$ ,  $p = 0.27$ ).

MANOVAs for other clinical and sleep outcomes, using groups defined by cluster analysis and controlling for gender, race, and Framingham risk group, are presented in Table 2. Only those domains that showed significant differences among groups are shown. The 4 groups differed significantly from each other on self-report clinical measures, quantitative sleep diary measures, and qualitative sleep diary measures. In particular, the groups differed on 5/10 clinical self-report measures, with the most “normal” scores being associated with the low PSQI/low ESS cluster. Likewise, the low PSQI/low ESS group had the most positive qualitative sleep ratings and the least disturbed quantitative sleep measures on the sleep diary. The 4 cluster-defined groups did not significantly differ from each other on measures derived from actigraphy, PSG sleep continuity, architecture, or apnea/PLMs. For instance, mean AHI in the 4 groups was  $13.6 \pm 16.4$  (Cluster 1),  $16.0 \pm 14.7$  (Cluster 2),  $12.3 \pm 14.7$  (Cluster 3), and  $12.7 \pm 13.8$  (Cluster 4).

We also divided the sample into 4 groups on the basis of commonly used PSQI and ESS cut-off scores. Seventy-two subjects (38.5%) had PSQI score  $\leq 5$  and ESS score  $\leq 10$ ; 20 subjects (10.7%) had PSQI score  $\leq 5$  but ESS score  $> 10$ ; 67 subjects

(35.8%) had PSQI score  $> 5$  but ESS score  $\leq 10$ ; and 28 subjects (15.0%) had both PSQI  $> 5$  and ESS  $> 10$ . Analyses using clinical cut-off scores were very similar to those using cluster analysis. Considering the 4 cut-off defined PSQI/ESS groups together, we again observed a significant effect for sex ( $\chi^2 = 14.08$ ,  $p = 0.003$ ), but not for race or Framingham risk groups. In particular, chi-square analyses indicated a significant sex effect for PSQI (higher proportion of women in high PSQI groups;  $\chi^2 = 7.42$ ,  $p = 0.006$ ), but not for ESS. We also examined differences among PSQI/ESS groups, as defined by traditional cut-off scores, on clinical and physiological domains, and again found very similar results to those using empirically defined clusters. Significant differences among groups were found for self-report psychological symptom measures (Ford Insomnia Response to Stress Test, Perceived Stress Scale, Spielberger Anxiety Scale, Pessimism Scale of the Life Optimism Test), and for both quantitative and qualitative sleep diary measures, but not for actigraphy or PSG domains of sleep disorders (apnea, periodic limb movements), sleep continuity, or sleep architecture.

Finally, we performed regression analyses to examine relationships between PSQI, ESS, and their interaction on the dependent measures of this study. These analyses make use of PSQI and ESS scores as continuous variables, rather than as categorical ones. The results of these analyses were once again quite similar to those using the clinical PSQI/ESS groups. PSQI scores were significantly related ( $p < 0.004$ ) to all self-report symptom measures except ongoing stress, depression, and un-

**Table 2**—Comparison of PSQI / ESS Subgroups Using Empirically-Defined Clusters

| Domain / Variable                         | PSQI / ESS Clusters <sup>2</sup> |                            |                            |                          | Statistical Results <sup>1</sup><br>MANOVA,<br>ANOVA, or $\chi^2$ |
|---|----------------------------------|----------------------------|----------------------------|--------------------------|---|
|   | Cluster 1<br>N = 73              | Cluster 2<br>N = 37        | Cluster 3<br>N = 54        | Cluster 4<br>N = 23      |   |
| PSQI and ESS                              |                                  |                            |                            |                          |   |
| PSQI                                      | 3.6 (1.3) <sup>a</sup>           | 5.4 (1.6) <sup>b</sup>     | 8.1 (2.0) <sup>c</sup>     | 12.2 (2.3) <sup>d</sup>  | < 0.001   |
| ESS                                       | 6.8 (2.4) <sup>a</sup>           | 13.0 (2.2) <sup>b</sup>    | 5.5 (2.3) <sup>c</sup>     | 11.5 (3.4) <sup>d</sup>  | < 0.001   |
| Demographics                              |                                  |                            |                            |                          |   |
| Age                                       | 60.1                             | 60.1                       | 59.0                       | 58.2                     | n.s.  |
| Sex (N, Female / Male)                    | 23/50                            | 16/21                      | 37/17                      | 12/11                    | $\chi^2 = 0.001$  |
| Race (N, African American / White)        | 26/47                            | 12/25                      | 25/29                      | 14/9                     | n.s.  |
| Body Mass Index                           | 29.7                             | 28.3                       | 29.2                       | 31.2                     | n.s.  |
| Self-report                               |                                  |                            |                            |                          |   |
| MANOVA < 0.001                            |                                  |                            |                            |                          |   |
| Ford Insomnia Response to Stress Test     | 16.4 (5.4) <sup>a</sup>          | 20.1 (7.0) <sup>b</sup>    | 20.5 (5.9) <sup>b</sup>    | 22.6 (5.6) <sup>b</sup>  | < 0.0001  |
| Perceived Stress Scale                    | 3.2 (2.1) <sup>a</sup>           | 3.9 (2.5) <sup>a</sup>     | 4.0 (2.8) <sup>a</sup>     | 5.3 (2.4) <sup>b</sup>   | 0.007   |
| Ongoing Stress Scale                      | 0.33 (.82)                       | 0.46 (.90)                 | 0.44 (.92)                 | 0.43 (.90)               | n.s.  |
| Center for Epidemiologic Studies          |                                  |                            |                            |                          |   |
| Depression (less sleep item) <sup>3</sup> | 9.3 (9.2)                        | 9.6 (10.0)                 | 9.2 (9.3)                  | 12.6 (9.5)               | n.s.  |
| Spielberger Anxiety <sup>3</sup>          | 4.6 (4.2) <sup>a</sup>           | 6.4 (5.0) <sup>a,b</sup>   | 6.7 (6.0) <sup>a,b</sup>   | 8.6 (4.7) <sup>b</sup>   | 0.008   |
| Spielberger Anger <sup>3</sup>            | 4.8 (3.6)                        | 5.7 (4.2)                  | 5.5 (4.7)                  | 6.4 (2.7)                | n.s.  |
| Life Optimism Test                        | 17.9 (3.4)                       | 18.0 (3.9)                 | 15.7 (4.7)                 | 15.9 (4.2)               | 0.008   |
| Cook-Medley Hostility Inventory Total     | 7.8 (4.4) <sup>a</sup>           | 7.8 (4.3) <sup>a</sup>     | 9.3 (4.2) <sup>a,b</sup>   | 10.3 (3.6) <sup>b</sup>  | 0.014   |
| Unfair Treatment Scale                    | 7.3 (4.1)                        | 7.5 (4.7)                  | 7.2 (4.3)                  | 9.9 (6.1)                | n.s.  |
| Sleep Diary Quantitative                  |                                  |                            |                            |                          |   |
| MANOVA < 0.001                            |                                  |                            |                            |                          |   |
| Latency (min) <sup>3</sup>                | 15.0 (9.3) <sup>a</sup>          | 13.7 (8.5) <sup>a</sup>    | 26.4 (17.2) <sup>b</sup>   | 30.0 (20.1) <sup>b</sup> | < 0.0001  |
| Total Sleep Time (hours)                  | 7.4 (0.9)                        | 7.2 (0.9)                  | 7.3 (1.0)                  | 7.1 (1.3)                | n.s.  |
| Wake After Sleep Onset (min)              | 15.9 (19.1) <sup>a</sup>         | 21.7 (17.8) <sup>a,b</sup> | 20.0 (16.1) <sup>a,b</sup> | 36.9 (35.6) <sup>b</sup> | 0.004   |
| Sleep Efficiency (%)                      | 92.7 (5.4) <sup>a</sup>          | 92.1 (5.7) <sup>a,b</sup>  | 89.3 (4.7) <sup>b,c</sup>  | 82.7 (15.1) <sup>c</sup> | < 0.0001  |
| Sleep Diary Qualitative                   |                                  |                            |                            |                          |   |
| MANOVA < 0.001                            |                                  |                            |                            |                          |   |
| Rested                                    | 2.61 (0.52) <sup>a</sup>         | 2.24 (0.49) <sup>b</sup>   | 2.18 (0.40) <sup>b,c</sup> | 2.01 (0.61) <sup>c</sup> | < 0.001   |
| How Feel                                  | 2.51 (0.50) <sup>a</sup>         | 2.24 (0.55) <sup>b</sup>   | 2.10 (0.53) <sup>b</sup>   | 1.83 (0.49) <sup>b</sup> | < 0.001   |
| Sleep quality                             | 2.56 (0.46) <sup>a</sup>         | 2.24 (0.50) <sup>b</sup>   | 2.14 (0.43) <sup>b,c</sup> | 1.92 (0.52) <sup>c</sup> | < 0.001   |

<sup>1</sup>n.s. = Not significant ( $p > 0.05$ ).  $p$  values are for ANOVA unless otherwise specified. ANOVA results reported only within outcome domains with significant MANOVA result.

<sup>2</sup>Post hoc comparisons (Tukey's  $b$  test) reported only for variables with significant ANOVA result. Group values followed by the same superscript do not differ from each other.

<sup>3</sup>Statistical analysis conducted on transformed data. Values in table are shown in original units.

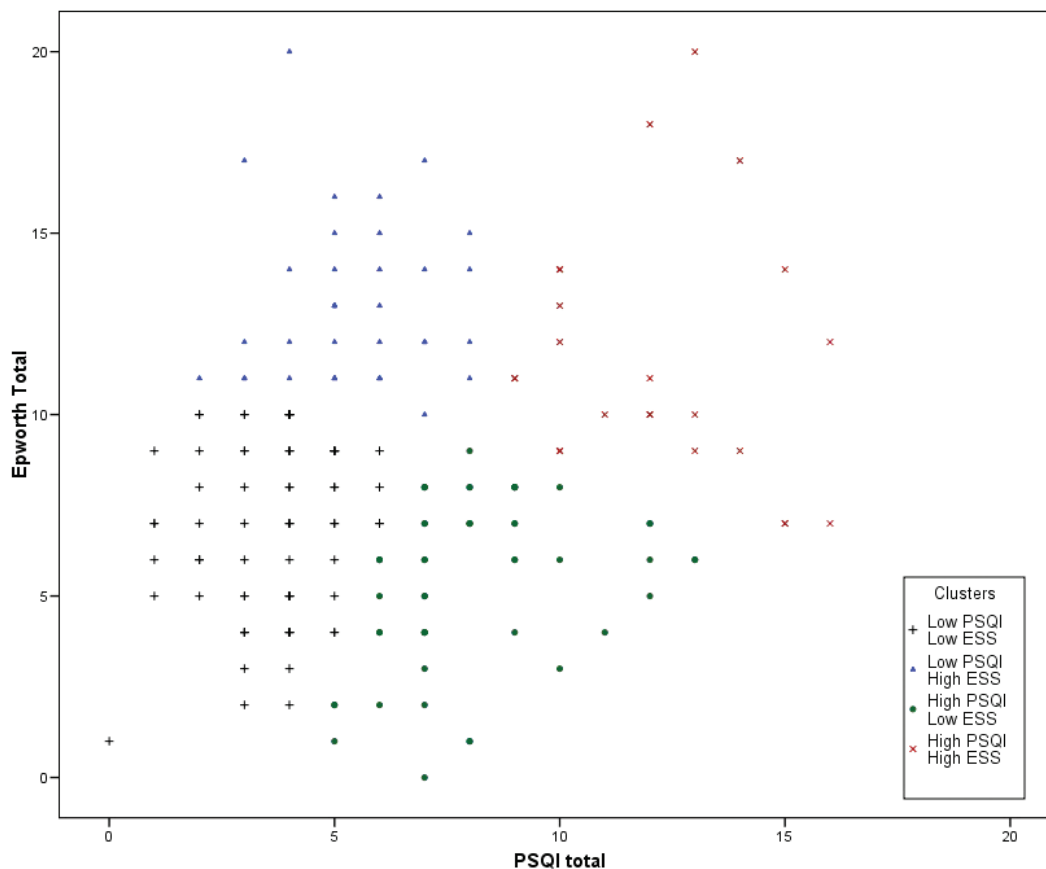
fair treatment, and to all qualitative and quantitative sleep diary measures except total sleep time. PSQI scores were not related to measures from actigraphy or polysomnography. Neither ESS scores nor the interaction of PSQI and ESS scores were significantly related to any of the outcomes at the  $p < 0.01$  level.

## DISCUSSION

This study is the first to examine relationships between the PSQI and ESS in a community sample of middle-aged and older adults. The 2 scales were only weakly correlated with each other, and deleting the "Daytime Dysfunction" component of the PSQI made the correlation nonsignificant. Exploratory factor analysis suggested that the 2 scales are essentially orthogonal. Previously published cut-off scores and cluster analysis were used to define subgroups of individuals based on their combination of PSQI and ESS scores. These subgroups differed from each other in quantitative and qualitative sleep diary reports, as well as a number of psychological self-report measures. Most of these observed differences were associated with PSQI scores, rather than ESS scores or the combination of PSQI and ESS scores. Regression analyses confirmed this observation. No dif-

ferences were found among PSQI/ESS subgroups in terms of actigraphy or PSG outcomes. We conclude that the PSQI and ESS, two of the most widely used self-report measures of sleep and sleepiness, capture orthogonal dimensions of sleep-wake experience, and that both of these dimensions are distinct from what is measured with actigraphy and PSG.

In some sense, it may seem obvious that the PSQI and the ESS should only weakly correlate with each other: The PSQI primarily assesses nighttime sleep quality, and the ESS the tendency to doze in various situations during the day. However, one could also argue that responses to these 2 scales would be affected by the individual's overall bias toward more or less symptom reporting. Indeed, the PSQI global score and ESS often correlate with self-report measures of depression, anxiety, stress, and fatigue,<sup>7,9,30,31</sup> which supports such a hypothesis. The current findings and previous reports, however, suggest that individuals' reports of sleep-related and sleepiness-related symptoms are distinct. For instance, the "Daytime Dysfunction" component of the PSQI generally has lower item-total correlations than other components,<sup>4,9,30,32</sup> and empirical examination of the factor structure of the PSQI identified distinct factors for perceived sleep quality, sleep efficiency, and daily disturbances



**Figure 3**—Histogram of PSQI-ESS Scores by Empirically Defined Clusters. Scatterplot of PSQI and ESS scores showing the distribution of subjects identified with K-means cluster analysis. Different symbols correspond to the four clusters.

(which includes the “Daytime Dysfunction” component).<sup>8</sup> By contrast, previous studies of the ESS have supported its unidimensional structure,<sup>5,11,12</sup> with different items contributing in a hierarchical fashion.<sup>11,33-35</sup>

Participant groups based on the combination of PSQI and ESS scores did not differ from each other on objective measures of rest-activity patterns, measured by actigraphy, or sleep, measured by polysomnography. This is consistent with previous studies, which have shown weak or inconsistent associations between the PSQI or ESS and objective measures. For instance, mean values for habitual sleep patterns derived from the PSQI typically differ from those measured by PSG, and PSQI-PSG correlations are weak.<sup>4,7,32,36</sup> In a similar way, early studies reported significant correlations between the ESS and the multiple sleep latency test (MSLT),<sup>6,33</sup> but other studies have shown weaker correlations with PSG, the MSLT, or the maintenance of wakefulness test (MWT).<sup>31,37,38</sup> This relative lack of association may, in part, be explained by the fact that the PSQI and ESS assess habitual patterns of sleep and sleepiness, whereas PSG, MSLT, and MWT assess sleep and sleepiness on a discrete occasion. Furthermore, the PSQI evaluates aspects of sleep (e.g., sleep “quality”) that are not directly captured with PSG, and we did not obtain an objective measure of daytime sleepiness that might directly relate to the ESS. It would appear, however, that using the PSQI and ESS with traditional cut-off scores is not an efficient mechanism for identifying community subjects with significant sleep apnea, periodic limb movements, or disturbed sleep continuity.

The PSQI was more strongly related than the ESS to self-report measures of psychological symptoms and stress. As noted above, correlations with psychological symptoms have previously been reported for both the PSQI and the ESS, but the strength of these associations has generally been stronger for the PSQI. The reason for this difference is not clear. It seems plausible that symptoms of impaired sleep quality, depression, anxiety, and stress might be linked to each other by some form of underlying “hyperarousal,” as hypothesized in the pathophysiology of insomnia.<sup>39</sup> For instance, we have previously shown significant relationships between disturbed sleep, altered sympathovagal tone, and symptoms of stress in healthy adults and individuals with depression and insomnia.<sup>40-42</sup> It seems less plausible that hyperarousal would manifest as daytime sleepiness. Alternatively, the observed associations with PSQI could be explained by sensitivity to negative experiences, and sleepiness may not be experienced in the same negative way.

PSQI/ESS subgroups were related to sex, weakly related to race, but not related to age. The absence of age differences likely related to the relatively homogeneous sample recruited. Sex differences have been reported inconsistently for both the PSQI<sup>4,7,9</sup> and the ESS<sup>6,43</sup> in previous studies. The increased prevalence of insomnia in women compared to men is consistent with our observation of elevated PSQI scores in women, although the actual sources of this difference are unknown. Unlike a previous epidemiological study, we did not find that a higher proportion of men had elevated ESS scores,<sup>43</sup> but substantial differences in

study samples, selection criteria, and procedures may account for this.

To our knowledge, only one other study has evaluated sex and race differences in PSQI and ESS scores.<sup>44</sup> In general, this study found small effect sizes associated with sex, and moderate effect sizes associated with race for both the PSQI and ESS; we found sex and race differences associated only with the PSQI. However, the sample reported by Knutson and colleagues was much larger ( $n = 610$ ) than our own, which may account for our failure to find ESS differences. African Americans in our sample also reported higher overall levels of stress and psychological symptoms and had lower incomes,<sup>45</sup> all factors that have previously been associated with subjective sleep complaints and insomnia.<sup>1,46</sup> This may help to explain the PSQI finding in our study and the Knutson study, but does not explain the discrepancy in findings with regard to ESS.

Strengths of the current study include a relatively large community sample with equal sex and race distributions, and the use of multiple self-report and objective outcomes. The mean PSQI and ESS scores in our sample were similar to values reported in another recent racially-diverse community sample (PSQI = 5.8 and ESS = 7.3).<sup>44</sup> In addition, we found a convergent pattern of results using multiple methods to analyze PSQI and ESS relationships with each other, and with other clinical and sleep measures. Nevertheless, our findings should be viewed as exploratory in nature. The large number of comparisons of statistical comparisons may have led to type I errors, although the use of multivariate statistics and multiple confirmatory approaches mitigates this concern. In addition, we excluded potential participants who had more serious medical and psychiatric disorders, which may limit generalizability to clinical samples. Finally, we used home PSG rather than laboratory-based PSG. Although we had a very high success rate and good quality for home PSG, it is possible that laboratory studies may have yielded different results. Indeed, Edinger and colleagues found significant differences in PSG measures when collected in the home and laboratory setting among a group of insomnia sufferers and control subjects.<sup>47</sup>

In conclusion, we found that the PSQI and ESS measure orthogonal symptom domains, and that subgroups based on these two scales are related to other self-report measures of sleep and distress, but not to objective measures of rest-activity or sleep. It does not appear that clinical groups based on the combination of these scales provide more information than the scales individually.

### ACKNOWLEDGMENTS

The authors wish to thank the staffs of the HeartSCORE and SleepSCORE studies for their exemplary efforts in recruitment and data collection, as well as the staff of the Neuroscience Clinical and Translational Research Center, who collected and scored the PSG data.

Supported by NIH research grants HL076379, HL076852, RR00052, RR024153 and the Pennsylvania Department of Health (Contract ME-02-384). The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions.

### DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Buysse consults for Actelion, Arena, Cephalon, Eli Lilly, GlaxoSmith-Kline, Merck, Neurocrine, Neurogen, Pfizer, Respiroics, Sanofi-Aventis, Sepracor, Servier, Somnus Therapeutics, Stress Eraser, Takeda, and Transcept. Dr. Strollo has received research support from ResMed and Respiroics. The other authors have indicated no financial conflicts of interest.

### REFERENCE LIST

1. Buysse DJ, Germain A, Moul D, et al. Insomnia. In: Buysse DJ, ed. *Sleep Disorders and psychiatry*. Arlington, VA: American Psychiatric Publishing, Inc; 2005:29-75.
2. Roehrs T, Carskadon MA, Dement WC, et al. Daytime sleepiness and alertness. In: Kryger MH, Roth T, Dement WC, eds. *Principles and practices of sleep medicine*, 4th ed. Philadelphia: Elsevier Saunders; 2005:39.
3. Moul DE, Hall M, Pilkonis PA, et al. Self-report measures of insomnia adults: rationales, choices, and needs. *Sleep Med Rev* 2004;8:177-98.
4. Buysse DJ, Reynolds CF, Monk TH, et al. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193-213.
5. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. *Sleep* 1992;15:376-81.
6. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14:540-5.
7. Buysse DJ, Reynolds CF, Monk TH, et al. Quantification of subjective sleep quality in healthy elderly men and women using the Pittsburgh Sleep Quality Index (PSQI). *Sleep* 1991;14:331-8.
8. Cole JC, Motivala SJ, Buysse DJ, et al. Validation of a 3-factor scoring model for the Pittsburgh Sleep Quality Index in older adults. *Sleep* 2006;29:112-6.
9. Carpenter JS, Andrykowski MA. Psychometric evaluation of the Pittsburgh Sleep Quality Index. *J Psychosom Res* 1998;45:5-13.
10. Johns MW. Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale. *Chest* 1993;103:30-6.
11. Hagell P, Broman JE. Measurement properties and hierarchical item structure of the Epworth Sleepiness Scale in Parkinson's disease. *J Sleep Res* 2007;16:102-9.
12. Violani C, Lucidi F, Robusto E, et al. The assessment of daytime sleep propensity: a comparison between the Epworth Sleepiness Scale and a newly developed Resistance to Sleepiness Scale. *Clin Neurophysiol* 2003;114:1027-33.
13. Nguyen AT, Baltzan MA, Small D, et al. Clinical reproducibility of the Epworth Sleepiness Scale. *J Clin Sleep Med* 2006;2:170-4.
14. Abele M, Klockgether T. Health-related quality of life in sporadic adult-onset ataxia. *Mov Disord* 2007;22:348-52.
15. Fabbrini G, Barbanti P, Aurilia C, et al. Excessive daytime sleepiness in de novo and treated Parkinson's disease. *Mov Disord* 2002;17:1026-30.
16. Tsai PS, Wang SY, Wang MY, et al. Psychometric evaluation of the Chinese version of the Pittsburgh Sleep Quality Index (CP-SQI) in primary insomnia and control subjects. *Qual Life Res* 2005;14:1943-52.
17. de Pinho RS, Silva-Junior FP, Bastos JP, et al. Hypersomnolence and accidents in truck drivers: A cross-sectional study. *Chronobiol Int* 2006;23:963-71.
18. Wilson PW, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation* 1998;97:1837-47.



19. Drake C, Richardson G, Roehrs T, et al. Vulnerability to stress-related sleep disturbance and hyperarousal. *Sleep* 2004;27:285-91.
20. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983;24:385-96.
21. Bromberger JT, Matthews KA. A longitudinal study of the effects of pessimism, trait anxiety, and life stress on depressive symptoms in middle-aged women. *Psychol Aging* 1996;11:207-13.
22. Radloff LS. The CES-D Scale: A self-report depression scale for research in the general population. *Appl Psychol Measure* 1977;1:385-401.
23. Spielberger CD, Gorsuch RL, Lushene RE. Manual for the state-trait anxiety inventory. Palo Alto, CA: Consulting Psychologists Press; 1970.
24. Scheier MF, Carver CS. Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health Psychol* 1985;4:219-47.
25. Cook WW, Medley DM. Proposed hostility and pharisaic-virtue scales for the MMPI. *J Appl Psychol* 1954;38:414-8.
26. Williams DR, Yu Y, Jackson JS, et al. Racial differences in physical and mental health: Socioeconomic status, stress, and discrimination. *J Health Psych* 1997;2:335-51.
27. Monk TH, Reynolds CF, Kupfer DJ, et al. The Pittsburgh Sleep Diary. *J Sleep Res* 1994;3:111-20.
28. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. NIH Publication 204. Washington, DC: U.S. Government Printing Office, Department of Health Education and Welfare; 1968.
29. American Academy of Sleep Medicine Task Force, Flemons WW, Buysse D, et al. Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. *Sleep* 1999;22:667-89.
30. Beck SL, Schwartz AL, Towsley G, et al. Psychometric evaluation of the Pittsburgh Sleep Quality Index in cancer patients. *J Pain Symptom Manage* 2004;27:140-8.
31. Olson LG, Cole MF, Ambrogetti A. Correlations among Epworth Sleepiness Scale scores, multiple sleep latency tests and psychological symptoms. *J Sleep Res* 1998;7:248-53.
32. Doi Y, Minowa M, Uchiyama M, et al. Psychometric assessment of subjective sleep quality using the Japanese version of the Pittsburgh Sleep Quality Index (PSQI-J) in psychiatric disordered and control subjects. *Psychiatry Res* 2000;97:165-72.
33. Johns MW. Sleepiness in different situations measured by the Epworth Sleepiness Scale. *Sleep* 1994;17:703-10.
34. Johns MW. Sleep propensity varies with behaviour and the situation in which it is measured: the concept of somnificity. *J Sleep Res* 2002;11:61-7.
35. Kingshott R, Douglas N, Deary I. Mokken scaling of the Epworth Sleepiness Scale items in patients with the sleep apnoea/hypopnoea syndrome. *J Sleep Res* 1998;7:293-4.
36. Backhaus J, Jughanns K, Broocks A, et al. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. *J Psychosom Res* 2002;53:737-40.
37. Chervin RD, Aldrich MS, Pickett R, et al. Comparison of the results of the Epworth Sleepiness Scale and the Multiple Sleep Latency Test. *J Psychosom Res* 1997;42:145-55.
38. Sangal RB, Mitler MM, Sangal JM. Subjective sleepiness ratings (Epworth sleepiness scale) do not reflect the same parameter of sleepiness as objective sleepiness (maintenance of wakefulness test) in patients with narcolepsy. *Clin Neurophysiol* 1999;110:2131-5.
39. Perlis ML, Smith MT, Pigeon WR. Etiology and pathophysiology of insomnia. In: Kryger MH, Roth T, Dement WC, eds. Principles and practice of sleep medicine, 4th ed. Philadelphia: Elsevier Saunders; 2005:714-25.
40. Hall M, Buysse DJ, Nowell PD, et al. Symptoms of stress and depression as correlates of sleep in primary insomnia. *Psychosom Med* 2000;62:227-30.
41. Hall M, Vasko R, Buysse DJ, et al. Acute stress affects heart rate variability during sleep. *Psychosom Med* 2004;66:56-62.
42. Hall M, Thayer JF, Germain A, et al. Psychological stress is associated with heightened physiological arousal during NREM sleep in primary insomnia. *Behav Sleep Med* 2007;5:178-93.
43. Gander PH, Marshall NS, Harris R, et al. The Epworth Sleepiness Scale: influence of age, ethnicity, and socioeconomic deprivation. Epworth Sleepiness scores of adults in New Zealand. *Sleep* 2005;28:249-53.
44. Knutson KL, Rathouz PJ, Yan LL, et al. Stability of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Questionnaires over 1 year in early middle-aged adults: the CARDIA study. *Sleep* 2006;29:1503-6.
45. Mezick EJ, Matthews KA, Hall M, et al. The influence of race and socioeconomic status on sleep: Pittsburgh SleepSCORE project. *Psychosom Med* 2008;70:410-6.
46. Hall M, Bromberger JT, Matthews KA. Socioeconomic status as a correlate of sleep in African-American and Caucasian women. In: Adler NE, Marmot M, McEwen BS, et al, eds. Socioeconomic status and health in industrial nations: social, psychological, and biological pathways. New York: Annals of the New York Academy of Sciences; 1999:427-30.
47. Edinger JD, Fins AI, Sullivan RJ, et al. Sleep in the laboratory and sleep at home: comparisons of older insomniacs and normal sleepers. *Sleep* 1997;20:1119-26.