Stability of the Pittsburgh Sleep Quality Index and the Epworth Sleepiness Questionnaires Over 1 Year in Early Middle-Aged Adults: The CARDIA Study

Kristen L. Knutson, PhD[1]; Paul J. Rathouz, PhD[2]; Lijing L. Yan, PhD[3,4]; Kiang Liu, PhD[3]; Diane S. Lauderdale, PhD[2]

[1]Section of Endocrinology, Department of Medicine, University of Chicago, Chicago, IL; [2]Department of Health Studies, University of Chicago, Chicago, IL; [3]Department of Preventive Medicine, Northwestern University, Chicago, IL; [4]Department of Health Economics and Management, Guanghua School of Management, Peking University, Beijing, China

Study Objectives: To describe the stability of the Pittsburgh Sleep Quality Index (PSQI) and the Epworth Sleepiness Scale (ESS) scores over 1 year among a population-based sample of black and white early middle-aged adults.

Participants: More than 600 participants, aged 38 to 50 years, from the Chicago site of the Coronary Artery Risk Development in Young Adults (CARDIA) Study.

Methods: The PSQI and ESS were completed twice, approximately 1 year apart, between 2003 and 2005. Seven PSQI 4-level component scores, a global PSQI score, and the ESS scores were calculated. A PSQI global score greater than 5 was classified as poor quality sleep, and an ESS score greater than 10 was classified as high daytime sleepiness.

Results: The mean ±SD PSQI score was 5.7 ±3.1 in Year 1 and 5.9 ±3.1 in Year 2. The mean ESS score was 7.4 ±4.3 in Year 1 and 7.2 ±4.2 in Year 2. The Pearson correlation coefficient for the PSQI score in both years in the full sample was .68 and ranged from .54 among black men to .72 among black women. The Pearson correlation coefficient for the ESS score in both years in the full sample was .76 and ranged from .70 among black men to .80 among white men. In the full sample, 76% had the same PSQI dichotomous classification, and 85% had the same ESS dichotomous classification in both years.

Conclusions: These results suggest that the PSQI and ESS are stable measures of sleep quality and sleepiness over the past year in early middle-aged adults.

Keywords: Survey methods, longitudinal, reliability

Citation: 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(11);1503-1506.

INTRODUCTION

THE PITTSBURGH SLEEP QUALITY INDEX (PSQI) AND EPWORTH SLEEPINESS SCALE (ESS) ARE WIDELY USED IN CLINICAL AND NONCLINICAL RESEARCH. The PSQI is designed to assess sleep quality during the past month and contains 19 self-rated questions from which 7 component scores are calculated and summed into a global score.[1] Higher scores represent worse sleep quality: component scores range from 0 to 3, and global scores range from 0 to 21. The PSQI has been used, for example, to measure sleep quality among truck drivers[2] and to test the effects of a drug on sleep quality in a randomized placebo-controlled trial.[3]

The ESS is an 8-item questionnaire designed to assess general level of daytime sleepiness, and scores on this instrument range from 0 to 24, with higher scores indicating greater sleepiness.[4] The ESS has been used in studies that have examined daytime sleepiness in medical interns[5] and in patients with multiple sclerosis,[6] and it was a main outcome measure of the effects of didgeridoo playing in patients with moderate obstructive sleep apnea.[7]

Reliability has been tested for both scales and found to be good. A previous study determined within-subject reliability among 91 subjects who completed the PSQI on 2 occasions an average of 28 days apart. The Pearson correlation coefficient for the global score was 0.85, and correlations for the component scores ranged from 0.65 to 0.84.[1] Another study examined test-retest reliability in a sample of 76 insomniacs over a period of 2 days to several weeks and observed a correlation of 0.87 for the global score.[8] A test-retest reliability analysis of the ESS was conducted in 87 medical students using a 5-month interval, and the Pearson correlation was 0.82.[9]

It is unknown, however, whether the constructs measured by these scales are stable over longer periods of time. Furthermore, it is unknown whether these measurements of sleep quality and sleepiness and/or the stability of these instruments varies by race or sex. Previous studies have observed differences in sleep architecture, disturbance, and behavior based on ethnicity and sex.[10-12] The aim of this analysis was to describe the stability of PSQI and the ESS scores over 1-year period among a population-based sample of black and white early middle-aged adults and to examine whether stability varied by sex and race.

METHODS

These data are from an ancillary study to Coronary Artery Risk Development in Young Adults (CARDIA), an on-going, prospective, multicenter cohort study of the evolution of cardiovascular risk factors among adults. In the original CARDIA cohort in 1985 to 1986 (N = 5115), participants were aged 18 to 30 years and were balanced by sex, race (black and white), and education. The ancillary study included participants from the Chicago site of CARDIA who participated in the Year 15 clinical exam and who were not pregnant at that time (total eligible 814).

The PSQI and ESS were sent to participants on 2 occasions approximately 1 year apart between 2003 and 2005 (mean interval ±

Disclosure Statement

This was not an industry supported study. Drs. Knutson, Rathouz, Yan, Liu, and Lauderdale have indicated no financial conflicts of interest.

Submitted for publication March 13, 2006
Accepted for publication July 3, 2006

Address correspondence to: Kristen L. Knutson, PhD, Department of Health Studies, University of Chicago, 5841 S. Maryland Ave, MC2007, Chicago, IL 60637; Tel: (773) 834-1973; Fax: (773) 702-1979; E-mail: kknutson@health.bsd.uchicago.edu

Table 1—PSQI and ESS Scores in Both Years and Pearson and Intraclass Correlation Coefficients Between the Scores in Each Year

	Full Sample	White women	White men	Black women	Black men
PSQI, no.	610	187	166	167	90
ESS, no.	609	186	166	165	92
Scores, mean ± SD					
PSQI Year 1	5.7 ±3.1	5.1 ±2.8	5.0 ±2.3	6.9 ±3.8	6.1 ±3.0
PSQI Year 2	5.9 ±3.1	5.5 ±3.1	5.3 ±2.6	6.7 ±3.6	5.9 ±2.8
ESS Year 1	7.4 ±4.3	6.3 ±3.7	7.0 ±3.8	8.4 ±4.7	8.4 ±4.7
ESS Year 2	7.2 ±4.2	6.6 ±4.0	6.7 ±3.8	8.1 ±4.5	7.4 ±4.4
Pearson correlation (95%CI)					
PSQI Year 1 & 2	.68 (0.63, 0.72)	.66 (0.57, 0.73)	.67 (0.57, 0.74)	.72 (0.63, 0.78)	.54 (0.38, 0.67)
ESS Year 1 & 2	.76 (0.73, 0.79)	.78 (0.71, 0.83)	.80 (0.73, 0.84)	.75 (0.67, 0.81)	.70 (0.58, 0.79)
Intraclass Correlation Coefficient					
PSQI Year 1 & 2	.81 (0.78, 0.84)	.79 (0.72, 0.84)	.80 (0.73, 0.85)	.83 (0.77, 0.88)	.70 (0.55, 0.80)
ESS Year 1 & 2	.86 (0.84, 0.88)	.87 (0.83, 0.91)	.89 (0.85, 0.92)	.86 (0.80, 0.89)	.83 (0.74, 0.88)

PSQI refers to the Pittsburgh Sleep Quality Index; ESS, Epworth Sleepiness Scale; CI, confidence interval

SD was 340 ± 72 days). Seven PSQI 4-level component scores, a global PSQI score, and the ESS scores were calculated.[1,4] As recommended by their developers, a PSQI global score greater than 5 was classified as poor quality sleep,[1] and an ESS Score greater than 10 was classified as high daytime sleepiness.[4]

Means and standard deviations were calculated for the global score of PSQI and the ESS score for the full sample and for each race-sex group (white women, white men, black women, black men). Analyses of variance with Hochberg posthoc tests were calculated to test for differences in means between race-sex groups. Cohen d values were calculated to assess effect size between each race-sex pair comparison. Generally, Cohen d values greater than 0.8 are defined as having "large" effect sizes, with values above 0.5 having "moderate" effects, and values above 0.2 having "small" effects.[13] Pearson correlation coefficients and intraclass correlation coefficients were calculated to compare the global PSQI scores and ESS scores within subjects over the 2 test times for the full sample and for each race-sex group. Pearson correlation coefficients between groups were compared using Fisher Z transformation of the coefficients, and 95% confidence intervals (CI) for the difference in Pearson r were computed. K coefficients were calculated to compare the dichotomous classifications and PSQI component scores in the full sample. The 95% CI were calculated for the proportions, correlations, and κ coefficients. All statistical analyses were performed using SPSS 13.0 (SPSS Inc., Chicago, IL).

RESULTS

The final sample included 610 individuals aged 38 to 50 years who completed the PSQI at both years and 609 who completed the ESS at both years. Table 1 presents the means and standard deviations of the PSQI and ESS scores in both years for the full sample. Fifty-four percent (95% CI: 50%, 58%) of all participants had a follow-up PSQI score within 1 point of their first PSQI score, and 48% (95% CI: 44%, 52%) of all participants had a follow-up ESS score within 1 point of their first ESS score. Table 1 also presents the Pearson correlation coefficients and the intraclass correlation coefficients for the PSQI and ESS scores (all $p < .001$). According to the Pearson correlation coefficients, 46% of the variance in the PSQI score at Year 2 is explained by the PSQI score at Year 1, and 58% of the variance in the ESS score at Year 2 is explained by the ESS score at Year 1. For both instruments, there is high within-subject reliability based on the intraclass correlation coefficients, both of which are above .80.

Means and correlation coefficients for each race-sex group are also presented in Table 1. According to Hochberg posthoc tests, the PSQI and ESS means for white women and white men did not differ significantly, nor did these means differ significantly between black women and black men ($p > .05$). The means for black women did differ significantly from the means for white women and white men for both measures at both time points. Mean ESS and PSQI scores at Year 1 differed between black men and white men but not at Year 2. Only the mean ESS score in Year 1 differed significantly between white women and black men. The Cohen d values for the race-sex comparisons are presented in Table 2. Moderate effects of race sex (Cohen $d > 0.4$) were observed for a few comparisons, although none of the comparison had a large effect (Cohen $d > 0.8$). Thus, despite significant differences in mean values, effect sizes between race-sex groups were not large. The percentages of subjects with a follow-up score within 1 point of their first score ranged from 42% among black males to 61% among white males for PSQI score and from 39% among black women to 55% among white men for ESS score. All of the Pearson correlation coefficients were greater than .5 for the race-sex groups. Only 1 comparison of the Pearson correlation coefficients between race-sex groups was significant. This was the comparison between the PSQI correlation among black women and black men (95% CI for difference in Pearson r: 0.03, 0.50; $p = .01$).

Table 2—Cohen d Values for Race-Sex Comparisons of Mean PSQI and ESS Scores

	PSQI		ESS	
Comparison	**Year 1**	**Year 2**	**Year 1**	**Year 2**
White women vs white men	0.06	0.07	0.17	0.03
White women vs black women	0.58	0.39	0.52	0.36
White women vs black men	0.32	0.12	0.52	0.19
White men vs black women	0.64	0.46	0.35	0.16
White men vs black men	0.38	0.19	0.35	0.33
Black women vs black men	0.26	0.27	0.00	0.17

PSQI refers to the Pittsburgh Sleep Quality Index; ESS, Epworth Sleepiness Scale.

The comparison of the dichotomous categories for the 2 questionnaires in the full sample indicated that 76% (95% CI: 73%, 79%) had the same PSQI classification, either as "poor quality sleep" or not in both years, and 85% (95% CI: 82%, 88%) had the same ESS classification, either as "high daytime sleepiness" or not in both years. The κ coefficient was 0.53 (95% CI: 0.46, 0.59; $p < 0.001$) for the PSQI classification and 0.53 (95% CI: 0.44, 0.61; $p<.001$) for the ESS classification. By race-sex group, 80% of white women, 77% of white men, 75% of black women, and 72% of black men had the same PSQI classification in both years. Eighty-seven percent of white women, 88% of white men, 79% of black women, and 84% of black men had the same ESS classification in both years.

For the 7 PSQI components, the percentages with the same scores at both years in the full sample were 55% for sleep latency ($\kappa = 0.35$; 95% CI for κ: 0.29, 0.41; $p < .001$), 57% for sleep duration ($\kappa = 0.31$; 95% CI for κ: 0.25, 0.37; $p < .001$), 62% for daytime dysfunction ($\kappa = 0.36$; 95% CI for κ: 0.30, 0.43; $p < .001$), 63% for sleep quality ($\kappa = 0.36$; 95% CI for κ: 0.30, 0.42; $p < .001$), 66% for sleep efficiency ($\kappa = 0.20$; 95% CI for κ: 0.13, 0.26; $p < .001$), 76% for sleep disturbances ($\kappa = 0.35$; 95% CI for κ: 0.28, 0.43; $p < .001$), and 81% for medication use ($\kappa = 0.42$; 95% CI for κ: 0.34, 0.49; $p < .001$).

DISCUSSION

Approximately half of the sample had follow-up PSQI global scores within 1 point of their first score, and three quarters of the sample would be similarly classified using the PSQI cutoff for poor quality sleep. Of the component scores, medication use and sleep disturbances showed the greatest stability in a year. Medication use was not very common in this sample, since over 80% of subjects in both years reported never taking medicine to help them sleep, and, thus, stability was fairly high. The sleep-disturbance component was based on the greatest number of questions (9 questions), which likely contributed to the stability of this component, since any single question did not contribute as much to the final score. The other components are based on 1 (sleep quality, sleep duration), 2 (daytime dysfunction, sleep latency), or 3 (sleep efficiency) questions. Approximately half of the sample also had follow-up ESS scores within 1 point of their first score, and 85% of the sample would be similarly classified using the ESS cutoff for high daytime sleepiness. Pearson correlations and intraclass correlation coefficients suggest fairly high reliability within subjects over 1 year. Finally, comparison of the 4 race-sex groups demonstrated that blacks had higher mean PSQI and ESS scores, indicating poorer sleep quality and higher daytime sleepiness than whites. Previous analysis of actigraphy data collected among this sample indicated strong race-sex differences even after adjustment for demographic and socioeconomic variables.[14] The correlation coefficients generally did not differ between race-sex groups, which indicates similar stability of these instruments between these groups, particularly between blacks and whites.

We found that a nontrivial percentage of the population is differently classified for poor sleep quality (24%) and high daytime sleepiness (15%) 1 year later. For the PSQI classification, 11% were no longer classified as having poor quality sleep in Year 2, and 13% became poor sleepers. For the ESS score, 9% were no longer classified with high daytime sleepiness in Year 2, and 6% moved into the high-daytime-sleepiness category. Thus, the numbers of individuals who report improvements in their sleep quality or sleepiness scores are similar to those who report deterioration. Comparison with previous reliability studies that retested the same subjects much sooner after the initial test on smaller samples indicates that our correlation for global PSQI scores 1 year apart was significantly lower than the scores reported by the previous study (95% CI for difference in r: 0.20, 0.57).[1] This is likely largely due to the difference in the interval between testing and suggests that the difference in our sample in the 2 administrations partly reflects short-term reliability of the instrument and partly reflects true changes over time in the construct the instrument is measuring. The correlation for the 2 ESS scores in our sample is similar to that reported by a previous study that repeated the measurement after a 5-month interval (95% CI for difference in Pearson r: -0.07, 0.37),[9] which suggests that the difference in scores within subjects is due to reliability of the instrument and that the construct is fairly stable over a 1-year period.

Our reliability estimates suggest that sleep quality and daytime sleepiness as assessed by the PSQI and ESS are fairly stable over a 1-year period in a population-based sample of early middle-aged black and white adults. For the majority of patients or subjects aged 38 to 50 who complete either of these questionnaires, it can be reasonably assumed that this score is representative of a 1-year time period. However, up to one quarter of the sample may experience a change in sleep quality or sleepiness over a 1-year period that would result in a different classification. These changes are likely due to various life events, such as changes in family status, health, or employment. This study did not collect this type of data at the same time as the sleep measurements, and we cannot test potential predictors of changes in sleep quality or sleepiness. Future researchers or clinicians may consider assessing potentially important life events. The PSQI and ESS are fairly stable measures of sleep quality and sleepiness over the past year in early middle-aged adults.

REFERENCES

1. Buysse DJ, Reynolds III CF, Monk TH, Berman SB, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Res 1989;28:193-213.
2. Souza JC, Paiva T, Reimao R. Sleep habits, sleepiness and accidents among truck drivers. Arq Neuropsiquiatr 2005;63:925-30.
3. Johnson DA, Orr WC, Crawley JA, et al. Effect of esomeprazole on nighttime heartburn and sleep quality in patients with GERD: a randomized, placebo-controlled trial. Am J Gastroenterol 2005;100:1914-22.
4. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540-5.
5. Rosen IM, Gimotty PA, Shea JA, Bellini LM. Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy, and burnout among interns. Acad Med 2006;81:82-5.
6. Heesen C, Nawrath L, Reich C, Bauer N, Schulz KH, Gold SM. Fatigue in multiple sclerosis: an example of cytokine mediated sickness behaviour? J Neurol Neurosurg Psychiatry 2006;77:34-9.
7. Puhan MA, Suarez A, Cascio CL, Zahn A, Heitz M, Braendli O. Didgeridoo playing as alternative treatment for obstructive sleep apnoea syndrome: randomised controlled trial. BMJ 2006;332:266-70.
8. Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. J Psychosom Res 2002;53:737-40.
9. Johns MW. Reliability and factor analysis of the Epworth Sleepiness Scale. Sleep 1992;15:376-81.

10. Jean-Louis G, Kripke DF, Ancoli-Israel S, Klauber MR, Sepulveda RS. Sleep duration, illumination, and activity patterns in a population sample: effects of gender and ethnicity. Biol Psychiatry 2000;47:921-7.
11. Profant J, Ancoli-Israel S, Dimsdale JE. Are there ethnic differences in sleep architecture? Am J Hum Biol 2002;14:321-6.
12. Jean-Louis G, Magai CM, Cohen CI, et al. Ethnic differences in self-reported sleep problems in older adults. Sleep 2001;24:926-33.
13. Cohen J. Statistical Power Analysis for the Behavioral Sciences Hillsdale, NJ: Lawrence Erlbaum Associates; 1987.
14. Lauderdale D, Knutson K, Yan L, et al. Objectively measured sleep characteristics among early middle-aged adults: The CARDIA study. Am J Epidemiol 2006; 164:5-16.