

Open camera or QR reader and  
scan code to access this article  
and other resources online.



# Sleep Self-Report and Actigraphy Measures in Healthy Midlife Women: Validity of the Pittsburgh Sleep Quality Index

Rochelle S. Zak, MD,<sup>1</sup> Jennifer Zitser,<sup>2,3</sup> Holly J. Jones,<sup>4</sup> Catherine L. Gilliss,<sup>5</sup> and Kathryn A. Lee<sup>5</sup>

## Abstract

**Background:** Validity of the Pittsburgh Sleep Quality Index (PSQI) has not been established for midlife women before menopause, and evidence suggests that two-factor or three-factor models may be more informative than the PSQI global score derived from its seven components. We hypothesized that the PSQI and its factor structure would be valid in premenopausal women.

**Materials and Methods:** We performed a validation study of the PSQI against wrist actigraphy in a community-based convenience sample of 71 healthy premenopausal women (aged 40–50 years). For convergent validity, PSQI and its component scores were compared with homologous actigraphy measures. For discriminant validity, characteristics known to affect sleep quality were compared, including body mass index, exercise, menopausal status, menopausal symptoms, and depressive symptoms measured with the Center for Epidemiological Studies–Depression (CES-D) Scale.

**Results:** The PSQI global score and Components 1 (quality) and 5 (disturbance) were correlated ( $p < 0.05$ ) with actigraphy-measured wake after sleep onset. The PSQI global score and Components 1 (quality) and 7 (daytime dysfunction) were correlated with CES-D scores. PSQI Components 2 (onset latency) and 4 (efficiency) were not congruent with homologous actigraphy measures, while component 3 (duration) was congruent with actigraphy duration. The single-factor PSQI global score had a higher McDonald's omega (0.705) and Cronbach's alpha (0.702) than the two-factor or three-factor models.

**Conclusions:** The PSQI global score is a valid measure of sleep quality in healthy midlife women, performing better than two-factor or three-factor models. However, overlapping CES-D and PSQI scores warrant further clinical assessment and research to better differentiate poor sleep quality from depression.

**Keywords:** menopause, insomnia, sleep onset latency, sleep disturbance, depression, factor analysis

## Introduction

**P**OOOR SLEEP IS a prevalent complaint among women during midlife, 40–60 years of age and a cardinal feature of the menopausal transition.<sup>1–3</sup> While it is often assumed that sleep

complaints are a consequence of hormonal changes, other factors such as body weight or mental health are likely involved.<sup>4,5</sup> Sleep quality, however, is a complex construct captured by objective measures of sleep duration or fragmentation and by a subjective sense of adequate or restful

<sup>1</sup>Sleep Disorders Center, University of California, San Francisco, California, USA.

<sup>2</sup>Department of Neurology, Memory and Aging Center, Global Brain Health Institute, University of California, San Francisco, San Francisco, California, USA.

<sup>3</sup>Movement Disorders Unit, Department of Neurology, Tel-Aviv Sourasky Medical Center, Affiliated to the Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

<sup>4</sup>College of Nursing, University of Cincinnati, Cincinnati, Ohio, USA.

<sup>5</sup>Department of Family Health Care Nursing, School of Nursing, University of California, San Francisco, California, USA.

sleep.<sup>6</sup> Objective sleep measures are expensive and resource-intensive and do not fully capture the subjective experience; thus, a validated questionnaire that adequately reflects a woman's perception of her sleep experience during midlife transition is critical. One self-report measure frequently used in research to measure sleep quality is the Pittsburgh Sleep Quality Index (PSQI). The PSQI consists of seven unique components that are summed to produce a global sleep quality score. It was initially validated against polysomnography (PSG) in three small groups of adults.<sup>6</sup> Because analysis by age or sex was not possible in the initial study, validation studies expanded to different populations using PSG or actigraphy as objective measures to address these potential correlates.<sup>7-9</sup> However, subsequent validation studies have yielded conflicting results, finding the PSQI congruent with actigraphy in a large sample of postmenopausal women (mean age=83 years)<sup>8</sup> but not in a small sample of older breast cancer survivors (mean age=58 years).<sup>9</sup> Other researchers postulated that a single global quality score may not fully capture the multidimensional nature of sleep quality and explored reconfigurations of the seven components of the PSQI using confirmatory factor analysis to support either two dimensions (sleep efficiency and sleep quality) or three dimensions (sleep efficiency, sleep quality, and daytime dysfunction).<sup>9-13</sup>

To the best of our knowledge, PSQI validation studies have not focused on midlife women before menopause. To address this gap, we sought to affirm convergent and discriminant validity in a community-based sample of healthy women in the late premenopausal and early perimenopausal stages. To test for convergent validity, we aimed to establish validity of the PSQI global score and its components using homologous actigraphy sleep parameters (sleep onset latency [SOL], sleep duration, wake after sleep onset [WASO], number of awakenings). We hypothesized that PSQI-reported values would have robust correlations with actigraphy values. To assess discriminant validity, the PSQI global score and component scores were examined for associations with three health characteristics known to impact sleep quality: body weight, menopausal symptoms, and depressive symptoms. Our secondary objective was to explore whether a two-factor or three-factor version of the PSQI would be a more informative measure of sleep quality for midlife premenopausal women than the global score derived from its seven components. We hypothesized that a two-factor or three-factor model would enhance the utility of the PSQI for research on midlife women's sleep quality beyond the single global score.

## Materials and Methods

### *Design and sample*

This cross-sectional validation study is based on PSQI data from a sample of women living in the San Francisco Bay Area who participated in the University of California San Francisco Midlife Women's Health Study. Latina, African American, and Caucasian women aged 40-50 years were recruited if they were healthy and experiencing regular menstrual cycles. Women were excluded if taking hormone therapy, currently pregnant, or had a major health problem such as cancer, stroke, major depressive disorder or sleep disorder. Recruitment details were previously reported.<sup>14,15</sup> The university's Committee on Human Research approved

the study and participants provided informed consent before data collection, which included demographic information, clinical measures, and questionnaires delineated below. The analysis for this validation study is focused on a subsample of women who also consented to participate in the sleep component with actigraphy and sleep diaries.

### *Measures*

**Demographic and clinical measures.** Demographic data included age, race/ethnicity, education, income, employment, and marital status. Weight and height were obtained to calculate body mass index (BMI), and a first-morning urine sample was obtained and analyzed for follicle-stimulating hormone (FSH) levels. FSH values and menstrual cycle regularity were used to categorize reproductive stage based on the Stages of Reproductive Aging Workshop criteria.<sup>16</sup> Premenopausal stage was designated as urinary FSH  $\leq 2.5$  IU/dL and no change in menstrual cycles; perimenopausal stage was designated as FSH  $> 2.5$  IU/dL or a change in menstrual cycle pattern during the past 6 months.

**Pittsburgh Sleep Quality Index.** The PSQI is a 19-item self-report assessment of sleep quality based on seven components of sleep (quality, onset latency, duration, efficiency, disturbance, use of sleep medication, and daytime dysfunction).<sup>6</sup> Participants are asked to consider typical nights during the previous month without distinguishing between weeknights and weekends. Each component is scored from 0 to 3, with higher scores indicating worse sleep. The seven component scores are summed for a global score ranging from 0 to 21. Scores above 5 indicate poor sleep quality.

**Wrist actigraphy sleep continuity measures.** Participants wore an actigraph (Mini-Motionlogger, AAM-32; Ambulatory Monitoring, Inc., Ardsley, NY) on their nondominant wrist to estimate sleep and wake time based on continuous movement counts sampled in 30-second epochs using zero-crossing mode. Wrist actigraphy has been validated with overnight PSG measures of sleep duration and wake time.<sup>17</sup> To control for potential weekend variability and reduce burden, participants were instructed to wear the actigraph continuously for two consecutive weekdays, and press the event marker when turning out the light and when awakening in the morning. To reduce researcher scoring bias, sleep epochs were determined using the Cole-Kripke algorithm from an automatic sleep scoring program (Action4<sup>®</sup> Software Program; Ambulatory Monitoring, Inc.).

Sleep duration was estimated from the first epoch of sleep to final awakening after subtracting intervening wake epochs. WASO, the sum of all wake epochs after sleep onset until final wake, was standardized as a percentage of the woman's sleep duration. SOL was minutes to the first epoch of sleep after pressing the event marker, and number of awakenings was determined using the algorithm that coded a wake period as  $\geq 2$  minutes of consecutive wake time. The two nights (with intraclass correlations of 0.74 for sleep duration, 0.72 for WASO, and 0.81 for number of awakenings) were averaged.

**Reclining and sleeping hours on weekdays and weekends.** The Paffenbarger Physical Activity Questionnaire (PPAQ) asks participants to estimate the number of hours per

day, on weekdays and weekends, spent in various activities categorized from vigorous to reclining/sleeping.<sup>18,19</sup> Hours spent reclining or sleeping on weekdays and weekends were used for comparisons with PSQI sleep duration. Detailed physical activity results for the total sample are reported elsewhere.<sup>14</sup>

**Depressive symptoms.** The 20-item Center for Epidemiological Studies–Depression (CES-D) Scale is a valid screening measure for depressive symptoms.<sup>20</sup> Participants indicate how often they experienced a particular symptom in the past week, from 0 (rarely/none) to 3 (5–7 days). Scores can range from 0 to 60, with a score  $\geq 16$  considered a risk factor for depression.<sup>20</sup>

*Statistical analysis*

Frequencies, percentages, and means  $\pm$  standard deviations (SDs) are used to describe the sample. Group mean differences were tested with independent sample *t*-tests or Mann–Whitney *U* tests for nonparametric comparisons of medians when appropriate. Cramer’s *V* ( $\phi$   $\Phi$ ) tests were used to compare dichotomous variables, with interpretations for small ( $<0.30$ ), medium ( $0.30$ – $0.49$ ), and large ( $>0.50$ ) effect sizes. Paired *t*-tests were used to test the first hypothesis that there would be: (a) no significant within-subject difference between PSQI continuous measures and their concept-equivalent actigraphy values (criteria: paired  $t < 1.96$ ,  $p > 0.05$ ); and (b) the two values would be highly correlated (criteria: Pearson  $r > 0.40$ ,  $p \leq 0.05$ ). For PSQI component scores that range from 0 to 3, Spearman *rho* was used to test convergent and discriminant validity with actigraphy and health measures. Cohen’s *d* (effect size in SD units) values were calculated

to evaluate for clinically meaningful mean differences (small  $0.20$ – $0.49$ , medium  $0.50$ – $0.79$ , and large  $\geq 0.80$ ). Statistical significance was set at  $p \leq 0.05$ .

Principal component analysis (PCA) with oblimin rotation was used to explore the factor structure in this small sample. Factors were limited to two-factor and three-factor models based on the published literature.<sup>9–13</sup> The Omega program for SPSS<sup>21</sup> was used to calculate McDonald’s omega ( $\omega$ ) and Cronbach’s alpha ( $\text{Cr } \alpha$ ) coefficients to examine internal consistency reliabilities for the original PSQI and the two models. All analyses were performed using SPSS.

**Results**

*Demographic and clinical characteristics*

Sample characteristics are presented in Table 1. By study design and community-based sampling strategies, the sample was racially/ethnically diverse and most perceived their health as “very good” or “excellent.” The mean age was  $43 \pm 2.4$  years; 50 of the 71 women were in late premenopausal reproductive stage and 21 were in early perimenopause. As expected, the only significant difference between these two groups was FSH level, but more than half (54%) of the sample reported experiencing hot flashes or night sweats, with no significant difference between the two groups. BMI averaged  $28.2 \pm 7.0$  (overweight category), and CES-D scores averaged  $13 \pm 9.6$  with 34% at risk for depression (score  $\geq 16$ ). On the PPAQ, participants reported reclining/sleeping about 30 minutes more on weekends than weekdays (Table 2).

*PSQI sleep characteristics*

For the overall sample, the PSQI global score averaged  $6.4 \pm 3.1$  (range 2–17) and internal consistency was

TABLE 1. PARTICIPANT CHARACTERISTICS BY REPRODUCTIVE STAGE

Characteristic	Total sample (n = 71)	Late premenopausal (n = 50)	Early perimenopausal (n = 21)	Statistic, p-value
	Mean $\pm$ SD	Mean $\pm$ SD	Mean $\pm$ SD	
Age (years)	43.3 $\pm$ 2.4	43.1 $\pm$ 2.2	43.8 $\pm$ 2.8	NS <i>d</i> = 0.28
Income adequacy (6–30)	20.5 $\pm$ 4.8	20.6 $\pm$ 5.2	20.2 $\pm$ 3.8	NS <i>d</i> = 0.09
Follicle-stimulating hormone (IU/dL)	1.25 $\pm$ 1.86	0.86 $\pm$ 1.70	2.18 $\pm$ 1.93	<i>t</i> = 2.7, <i>p</i> = 0.011
BMI (kg/m <sup>2</sup> )	28.2 $\pm$ 7.0	27.3 $\pm$ 5.35	30.5 $\pm$ 9.88	NS <i>d</i> = 0.40
Exercise (days/week)	3.2 $\pm$ 1.7	3.3 $\pm$ 1.8	3.0 $\pm$ 1.3	NS <i>d</i> = 0.19
CES-D score (0–60)	13.2 $\pm$ 9.6	13.6 $\pm$ 9.6	12.2 $\pm$ 9.9	NS <i>d</i> = 0.14
	n (%)	n (%)	n (%)	
Depression risk (CES-D $\geq 16$ )	23 (34)	17 (35)	6 (32)	NS $\Phi$ = 0.030
Hot flashes or night sweats	38 (54)	25 (50)	13 (62)	NS $\Phi$ = 0.109
Race/ethnicity				NS $\Phi$ = 0.189
African American	21 (30)	12 (24)	9 (43)	
Caucasian	29 (40)	22 (44)	7 (33)	
Latina	21 (30)	16 (32)	5 (24)	
Has partner/spouse	38 (54)	26 (52)	12 (57)	NS $\Phi$ = 0.113
Employed for pay	58 (87)	43 (88)	15 (83)	NS $\Phi$ = 0.057
Has child at home	44 (66)	31 (63)	13 (79)	NS $\Phi$ = 0.084

BMI, body mass index; CES-D, Center for Epidemiological Studies–Depression Scale; *d*, effect size in SD units; NS, *t*-test for mean differences not significant;  $\phi$   $\Phi$ , Cramer’s *V*; SD, standard deviation.

TABLE 2. SLEEP CHARACTERISTIC COMPARISONS BY REPRODUCTIVE STAGE

Characteristic	Total sample (n=71)	Late premenopausal (n=50)	Early perimenopausal (n=21)	Statistic, p-value
	Mean ± SD	Mean ± SD	Mean ± SD	
<b>PSQI</b>				
Global score (0–21)	6.4 ± 3.1	6.5 ± 3.3	6.1 ± 2.6	NS <i>d</i> =0.13
Poor global sleep quality (>5), <i>n</i> (%)	40 (56)	27 (54)	13 (62)	NS $\Phi$ =0.091
Component 1 quality (0–3)	1.1 ± 0.74	1.0 ± 0.72	1.3 ± 0.75	NS <i>d</i> =0.27
Component 2 onset latency (0–3)	1.1 ± 0.83	1.1 ± 0.78	1.1 ± 0.99	NS <i>d</i> =0.000
Component 3 duration (0–3)	1.0 ± 0.68	0.9 ± 0.66	1.1 ± 0.74	NS <i>d</i> =0.29
Component 4 efficiency (0–3)	0.4 ± 0.79	0.5 ± 0.88	0.2 ± 0.43	NS <i>d</i> =0.43
Component 5 disturbance (0–3)	1.4 ± 0.55	1.5 ± 0.58	1.2 ± 0.38	<i>t</i> =2.2, <i>p</i> =0.034 <i>d</i> =0.61
Component 6 sleep medication (0–3)	0.4 ± 0.87	0.4 ± 0.91	0.3 ± 0.77	NS <i>d</i> =0.12
Component 7 daytime dysfunction (0–3)	1.0 ± 0.72	1.1 ± 0.70	0.8 ± 0.77	NS <i>d</i> =0.41
<b>Actigraphy</b>				
Time in bed (hours)	7.9 ± 1.1	7.9 ± 1.1	7.8 ± 1.1	NS <i>d</i> =0.09
Sleep duration (hours)	7.0 ± 1.2	7.1 ± 1.1	6.8 ± 1.4	NS <i>d</i> =0.24
Sleep efficiency (%)	89.2 ± 10.4	89.7 ± 9.5	87.8 ± 12.2	NS <i>d</i> =0.17
Sleep onset latency (minutes)	11.7 ± 7.6	11.7 ± 7.0	11.6 ± 9.1	NS <i>d</i> =0.01
Number of awakenings	13 ± 8.3	12 ± 8.2	15 ± 8.5	<i>t</i> =2.00, <i>p</i> =0.055 <i>d</i> =0.37
Wake after sleep onset (%)	9.0 ± 11.2	8.2 ± 10.8	11.1 ± 12.1	NS <i>d</i> =0.26
Median	5	4	6	MWU <i>p</i> =0.080
<b>Recline/sleep</b>				
Weekdays (hours)	7.3 ± 1.3	7.4 ± 1.3	7.0 ± 1.2	NS <i>d</i> =0.32
Weekends (hours)	8.0 ± 1.4	8.0 ± 1.4	7.8 ± 1.6	NS <i>d</i> =0.13

MWU, Mann–Whitney *U*; PSQI, Pittsburgh Sleep Quality Index.

acceptable (Cronbach's alpha=0.71). As shown in Table 2, poor sleep quality (PSQI >5) was evident for more than half of the sample (56%), yet sleep efficiency (Component 4) was not particularly problematic and use of sleep medication (Component 6) was very low (both ≤0.5 on the 0–3 scales). Sleep disturbance (Component 5) was higher (worse) for premenopausal women compared with perimenopausal women (*p*=0.034; *d*=0.61), yet the trend of better sleep on

actigraphy was evident for premenopausal women, as indicated by fewer wake episodes (*p*=0.055, *d*=0.37) and less WASO (*p*=0.080, *d*=0.26).

#### Convergent validity

PSQI continuous measures. PSQI items asking about number of hours and minutes for sleep are compared with

TABLE 3. WITHIN-SUBJECT COMPARISONS BETWEEN SELF-REPORT PITTSBURGH SLEEP QUALITY INDEX AND ACTIGRAPHY (*n* = 71)

Sleep characteristic	Mean ± SD	Pearson correlation ( <i>r</i> ) with PSQI	Paired <i>t</i> statistic ( <i>p</i> value)	Cohen's <i>d</i>
<b>Time in bed (hours)</b>				
PSQI	7.5 ± 1.1			
Actigraphy	7.9 ± 1.1	0.44 <sup>a</sup>	2.6 (0.01)	0.439
Recline/sleep weekdays	7.3 ± 1.3	0.54 <sup>a</sup>	1.3 (0.21) <sup>b</sup>	0.219
Recline/sleep weekends	8.0 ± 1.4	0.34 <sup>a</sup>	2.9 (0.004)	0.490
<b>SOL (minutes)</b>				
PSQI	19.3 ± 16.9			
Actigraphy	11.7 ± 7.8	0.24	3.5 (0.001)	0.591
<b>Sleep duration (hours)</b>				
PSQI	6.7 ± 1.1			
Actigraphy	7.0 ± 1.2	0.40 <sup>a</sup>	2.3 (0.028)	0.388
<b>Sleep efficiency (%)</b>				
PSQI	89.7 ± 11.8			
Actigraphy	89.2 ± 9.1	0.05	0.3 (0.81) <sup>b</sup>	0.050

<sup>a</sup>Pearson correlation *p*-value ≤0.05; measure significantly correlated with PSQI measure.

<sup>b</sup>Paired samples *t*-test *p*-value >0.05; measure not significantly different from PSQI measure.  
SOL, sleep onset latency.

actigraphy measures in Table 3. Each woman's PSQI time in bed was significantly correlated ( $r$ 's 0.34–0.54) with her actigraphy time in bed and her PPAQ reclining/sleeping on weekdays and weekends. However, PSQI time in bed was significantly shorter than actigraphy time in bed (paired  $t = 2.6, p = 0.01$ ), and only PPAQ weekday reclining/sleeping hours did not differ from PSQI hours ( $p > 0.05$ ). PSQI SOL was significantly longer and more variable ( $19.3 \pm 16.9$  minutes) than actigraphy-recorded SOL ( $11.7 \pm 7.8$  minutes), and these two measures were not significantly correlated. In contrast, PSQI sleep duration was correlated with actigraphy duration ( $r = 0.40$ ) but was also significantly shorter ( $6.7 \pm 1.1$  hours) than actigraphy duration ( $p = 0.028$ ). Finally, each woman's sleep efficiency calculated from her PSQI responses did not differ from her actigraphy value; these two continuous measures were not correlated.

PSQI global score and component scores. Table 4 presents Spearman  $\rho$  correlations between PSQI component scores and actigraphy measures. The PSQI global score was significantly correlated with actigraphy sleep continuity variables (efficiency and WASO), with higher PSQI global scores supported by worse sleep continuity. PSQI global sleep quality and Component 1 (quality) scores were also inversely related to longer self-reported PPAQ duration of reclining/sleeping on both weekdays and weekends. The PSQI global score was not related to actigraphy time in bed, SOL, or sleep duration. Component 1 was related to actigraphy SOL and WASO. However, as with the continuous measure of SOL, Component 2 (SOL, 0–3 score) was not correlated with actigraphy-recorded SOL.

The PSQI Component 3 (duration) score was inversely related ( $\rho = -0.33, p < 0.01$ ) to actigraphy sleep duration (Table 4), with the lowest Component 3 score ("0") indi-

cating the longest sleep duration (>7 hours). Component 3 was also significantly correlated with actigraphy measures of sleep efficiency and WASO, both of which include sleep duration in their calculations. Lower PSQI Component 3 (duration) scores also had a significant inverse relationship with longer durations of PPAQ self-reported reclining/sleeping on both weekdays and weekends.

The Component 4 (efficiency) score was not correlated with any actigraphy measure, including sleep efficiency. PSQI Component 5 (disturbance) was significantly correlated with actigraphy-measured WASO, indicating that the higher the PSQI self-report of sleep disturbance, the higher the percentage of WASO recorded with actigraphy. PSQI Component 6 (medication) and Component 7 (daytime dysfunction) were unrelated to any actigraphy measure.

*Discriminant validity: PSQI and health correlates associated with sleep quality*

Neither FSH level nor experience of menopausal symptoms was related to any PSQI parameter. In contrast, both FSH and menopausal symptoms were correlated with actigraphy number of awakenings ( $r = 0.382$  and  $r = 0.408$ , respectively;  $p \leq 0.001$ ). As shown in Table 4, menopausal stage was negatively related to PSQI Component 5 (disturbance), such that women in the premenopausal stage had more disturbed sleep than women in the perimenopausal stage. This relationship is supported by the significant group mean difference seen only for Component 5 in Table 2. As expected, BMI was correlated with the PSQI global score, indicating that the higher the BMI, the worse the overall global sleep quality; however, BMI was not significantly correlated with PSQI Component 1 (quality) or any other PSQI component.

CES-D scores were related to the PSQI global score, Component 1 (quality), and Component 5 (disturbance) with

TABLE 4. PITTSBURGH SLEEP QUALITY INDEX COMPONENT SCORES AND RELATIONSHIPS (SPEARMAN  $\rho$ ) WITH ACTIGRAPHY AND CLINICAL CHARACTERISTICS ( $n = 71$ )

PSQI	Total score	Comp 1 quality	Comp 2 latency	Comp 3 duration	Comp 4 efficiency	Comp 5 disturbance	Comp 6 medication	Comp 7 dysfunction
<b>Actigraphy measures</b>								
Time in bed	-0.001	0.044	0.192	-0.225	0.142	0.089	-0.198	0.074
Sleep onset latency	0.229	0.267 <sup>a</sup>	0.225	0.128	0.044	0.121	0.086	0.076
Total sleep time	-0.098	-0.007	0.167	-0.330 <sup>b</sup>	0.083	-0.029	-0.175	0.018
Sleep efficiency	-0.309 <sup>a</sup>	-0.225	-0.080	-0.383 <sup>b</sup>	-0.090	-0.230	-0.084	-0.175
WASO%	0.344 <sup>b</sup>	0.256 <sup>a</sup>	0.097	0.342 <sup>b</sup>	0.159	0.276 <sup>a</sup>	0.075	0.167
Wakes (number)	0.175	0.028	0.097	0.138	0.049	0.224	0.017	0.116
<b>Clinical measures</b>								
Menopause stage (pre = 0; peri = 1)	0.011	0.179	-0.004	0.089	-0.122	-0.258 <sup>a</sup>	0.005	-0.174
Hot flashes or night sweats (no = 0, yes = 1)	0.027	0.048	0.206	-0.110	-0.090	0.035	-0.034	-0.196
FSH	0.018	0.027	-0.049	-0.018	-0.049	-0.020	-0.020	-0.014
Age	-0.080	0.134	-0.115	-0.066	-0.178	-0.049	-0.159	-0.176
BMI	0.267 <sup>a</sup>	0.136	0.029	0.239	-0.010	0.177	0.178	0.124
Exercise days/week	-0.183	-0.244	0.192	-0.233	-0.022	-0.013	0.032	-0.389 <sup>b</sup>
Depressive symptoms	0.356 <sup>b</sup>	0.324 <sup>b</sup>	0.022	0.144	0.038	0.322 <sup>b</sup>	0.054	0.599 <sup>b</sup>
Recline weekdays	-0.366 <sup>b</sup>	-0.255 <sup>a</sup>	-0.005	-0.648 <sup>b</sup>	-0.235	0.088	-0.313 <sup>a</sup>	-0.118
Recline weekends	-0.407 <sup>b</sup>	-0.372 <sup>b</sup>	-0.120	-0.512 <sup>b</sup>	-0.307 <sup>a</sup>	-0.065	-0.252 <sup>a</sup>	-0.137

<sup>a</sup> $p \leq 0.05$ .

<sup>b</sup> $p \leq 0.01$ .

Comp, component of the PSQI; FSH, follicle-stimulating hormone; WASO, wake after sleep onset.

TABLE 5. INTERNAL CONSISTENCY OF PITTSBURGH SLEEP QUALITY INDEX FACTOR STRUCTURE MODELS AND RELATIONSHIPS (SPEARMAN  $\rho$ ) WITH ACTIGRAPHY MEASURES ( $n=71$ )

Factor structure models	$\omega$ (Cr $\alpha$ )	Time in bed (A)	Sleep onset (A)	Total sleep time (A)	Sleep efficiency (A)	WASO (A)	Number of wakes (A)
Single-factor model							
PSQI (C 1–7)	0.705 (0.702)	0.027	0.203	–0.074	–0.307 <sup>a</sup>	0.332 <sup>b</sup>	0.171
PSQI (without C 6)	0.716 (0.712)	0.077	0.189	–0.028	–0.293 <sup>a</sup>	0.319 <sup>b</sup>	0.173
PSQI (C 1–6)	0.682 (0.677)	0.025	0.204	–0.071	–0.296 <sup>a</sup>	0.323 <sup>b</sup>	0.162
PSQI (C 1–5)	0.700 (0.697)	0.075	0.185	–0.023	–0.281 <sup>a</sup>	0.310 <sup>a</sup>	0.164
Two-factor model							
Factor 1 (C 1, 2, 5, 6)	0.584 (0.573)	0.100	0.273 <sup>a</sup>	0.031	–0.231	0.263 <sup>a</sup>	0.167
Factor 2 (C 3, 4, 7)	0.624 (0.607)	–0.050	0.112	–0.155	–0.295 <sup>a</sup>	0.305 <sup>a</sup>	0.137
Three-factor model							
Factor 1 (C 3, 4)	–(0.619)	–0.094	0.127	–0.213	–0.329 <sup>b</sup>	0.337 <sup>b</sup>	0.138
Factor 2 (C 1, 2, 6)	0.505 (0.498)	–0.025	0.258 <sup>a</sup>	–0.057	–0.174	0.186	0.039
Factor 3 (C 5, 7)	–(0.396)	0.083	0.136	–0.005	–0.234	0.258 <sup>a</sup>	0.191

<sup>a</sup> $p < 0.05$ .

<sup>b</sup> $p < 0.01$ .

A, actigraphy measure; C, PSQI component; Cr- $\alpha$ , Cronbach's alpha coefficient;  $\omega$ , McDonald's omega with minimum of 3 items.

modest  $\rho$  values (range 0.32–0.36). The CES-D score was highly correlated with Component 7 (daytime dysfunction) score ( $\rho=0.599$ ,  $p < 0.001$ ). When the CES-D item about restless sleep was removed from the CES-D, correlations were somewhat attenuated ( $\rho=0.28$ – $0.31$ , and  $0.58$ , respectively) but remained significant. In contrast, there were no relationships between CES-D score and any actigraphy measure (all Pearson  $r$  coefficients  $< 0.10$ ).

#### Factor structure

PCA was performed first forcing a two-factor model and then forcing a three-factor model. Both models had absolute factor loadings  $> 0.40$ . However, with McDonald's omega and Cronbach's alpha  $< 0.70$ , neither model had acceptable internal consistency compared with the original PSQI single-factor model with its seven components (Table 5). When Component 6 (sleep medication) was removed from the PSQI single-factor model due to low endorsement, internal consistency omega improved slightly from 0.705 to 0.716. Omega did not improve when Component 7 (daytime dysfunction) was removed from the PSQI single-factor model because of its overlap with depressive symptoms, or when Components 6 and 7 were removed simultaneously (Table 5). Given limited internal consistencies of the two-factor and three-factor models, confirmatory fit indices and further validations of these alternative factor structures were not performed. Each model and its correlations with actigraphy measures are shown in Table 5 with less robust correlations for the two-factor and three-factor models.

#### Discussion

To the best of our knowledge, this is the first study demonstrating validity of the PSQI in a diverse sample of healthy premenopausal women between the ages of 40 and 50 years. Our hypothesis regarding convergent validity of the PSQI global score was supported by robust correlations with actigraphy values. Hypotheses regarding discriminant validity were only partially supported. There was no association between experience of menopausal symptoms and any PSQI

component. In contrast, depressive symptom scores were correlated with the PSQI global score as well as three component scores, and BMI and PSQI global score were correlated. Finally, contrary to our hypothesis, the single-factor PSQI global score was more internally consistent than either a two-factor or three-factor version.

The PSQI global sleep quality score was congruent with actigraphy sleep efficiency and WASO but not sleep duration. These findings emphasize the need to consider sleep duration or quantity as distinct from the dimension of sleep quality. Our finding that women underestimated sleep duration by about 20 minutes compared with actigraphy contradicts Jackson et al.<sup>22</sup> who reported overestimates in sleep duration. The estimates in their sample did not differ by sex, but self-reported weekday sleep duration was overestimated by 46–48 minutes compared with weekday actigraphy, and overestimated by 49–73 minutes compared with one-night home PSG values (6.0 hours).<sup>22</sup>

Our results of a significant correlation for the PSQI global score with actigraphy WASO support and extend findings from a large study of older women ( $\geq 70$  years) showing a small but significant correlation ( $\rho=0.14$ ,  $p < 0.001$ ) with WASO based on three nights of actigraphy.<sup>8</sup> A small correlation ( $\rho=0.17$ ) was also reported for WASO based on 5 days of actigraphy that included two weekend nights, but this correlation did not reach statistical significance in the small sample of 62 breast cancer survivors who were primarily postmenopausal.<sup>9</sup> Our correlation was also stronger than the correlation ( $\rho=0.198$ ) reported for a mixed-gender sample of 59 older adults (59–75 years) based on WASO from 7 days of actigraphy.<sup>23</sup> Our more robust correlation may reflect our focus on weekdays or our standardization of WASO as a percentage of sleep duration rather than absolute minutes of WASO.

Given the correlation between PSQI global scores and actigraphy WASO and sleep efficiency and that the PSQI is a composite score of multiple components, we sought to establish whether analysis of the individual PSQI components could yield additional insights for clinicians or researchers beyond what can be inferred from the PSQI global score.

### *Convergent validity for PSQI component scores*

As expected, Component 1 (quality) was associated with actigraphy-measured SOL and WASO, but not with actigraphy sleep duration. Component 2 (SOL) was not significantly related to actigraphy SOL but was more robust than correlations reported by others.<sup>9,23</sup> Actigraphy underestimated SOL compared with self-reported SOL from participants, which could reflect the inability of actigraphy to pick up quiet wakefulness.<sup>24</sup> SOL is often skewed in research samples, and its relationship to sleep quality is not necessarily linear when reporting a few minutes to fall asleep could indicate chronic short sleep duration and not falling asleep within 30 minutes may reflect chronic insomnia. It would be interesting in future studies to compare the PSQI response of “cannot get to sleep within 30 minutes” on three or more nights/week with the response to the PSQI question of typical minutes to fall asleep in the past month. Because sleep efficiency includes SOL, it was not surprising that Component 4 (efficiency) was unrelated to actigraphy sleep efficiency and the absence of an association between these two measures has been reported in other cohorts.<sup>23</sup>

Component 3 (sleep duration 0–3 scale) was related to actigraphy sleep duration and supports relationships seen in prior studies.<sup>22,23</sup> Whether duration is in continuous hours or a 0–3 range (Component 3), these findings indicate that sleep duration should be considered distinctly different from sleep quality. Our results support the suggestion by Jackson et al.<sup>22</sup> that caution be taken when dichotomizing sleep duration into adequate and inadequate categories based on self-report measures that not only differ on weekdays and weekends but differ significantly from objective measures.

Underestimating time in bed and sleep duration by self-report compared with actigraphy may be explained by how a person responds to PSQI items with a general time frame of the past month, as we observed differences in PPAQ responses for weeknights and weekends. Discrepancies may also reflect either the predetermined time frames for Component 3 (duration), actigraphy’s tendency to underestimate quiet wakefulness,<sup>24</sup> or the different levels of awareness, expectations, and distress associated with sleep.<sup>25</sup>

Component 5 (disturbance) was significantly correlated with actigraphy WASO as expected but did not reach statistical significance for actigraphy-recorded number of awakenings. This is not particularly surprising since Component 5 scores reflect the frequency of awakenings for a variety of specified reasons, whereas WASO reflects the accumulated amount of wake time during the night after falling asleep, irrespective of the reason. Of additional interest, Component 5 was worse in premenopausal women than perimenopausal women. Our prior work suggests that premenopausal women have more frequent awakenings to urinate and perhaps these awakenings are perceived as more disturbing.<sup>25</sup> Conversely, aspects of the menopausal transition may alter perception of sleep disturbance or change expectations about sleep fragmentation resulting in subjective improvement in sleep quality.

### *Discriminant validity*

Our hypotheses were only partially supported for discriminant validity of the PSQI with health indicators known to correlate with sleep quality. PSQI global scores were not differentiated by reproductive stage or menopausal symp-

toms. In fact, our premenopausal group reported more sleep disturbance (Component 5) yet showed less sleep disturbance by actigraphy measures compared with perimenopausal women. This finding may be influenced by the precise reason for awakening, duration of awakenings, or types of sleep disturbance experienced by midlife women<sup>25</sup> and should be explored in future studies.

BMI correlated with the global score but was unrelated to any single PSQI component. The correlation between BMI and the PSQI global score reflects the known relationship between obesity and poor sleep quality.<sup>5</sup> This finding supports the need to consider the influence of BMI on midlife women’s sleep quality in future studies.

CES-D depression scores were correlated with both the PSQI global score and Component 7 (daytime dysfunction), supporting the suggestion that the PSQI global score also measures some aspect of mood.<sup>8,23</sup> In contrast, we found no significant correlations between CES-D scores and actigraphy measures of sleep continuity. Future research should consider modified versions of sleep and depression questionnaires with items that are independent of the comorbid condition. For example, the PSQI item about enthusiasm in Component 7 (daytime dysfunction) could be excluded or rephrased to better reflect daytime dysfunction exclusively due to sleepiness. It may be necessary to remove Component 7 from the PSQI scoring algorithm for a valid sleep quality measure or control for depression in the analyses.<sup>26</sup> Finally, to minimize the collinearity between measures, it may be necessary to remove any sleep-related item from a depression measure or remove any depression-related item from a sleep measure.

### *Factor structure*

Contrary to the hypothesis in our secondary aim, the single-dimension PSQI global score was more internally consistent than either a two-factor or three-factor model. Although the original seven PSQI components could be recombined into a two-factor or three-factor structure, only the original single PSQI score had an acceptable McDonald’s omega ( $\geq 0.70$ ) in our sample, particularly when removing Component 6 (medications). Fewer items will often yield lower internal consistency, but our low omega may also be due to an all-female sample. In a sample of adults with type 2 diabetes, women had a lower Cronbach’s alpha (0.67) compared with men (0.72), suggesting that PSQI factor structures may be sex dependent.<sup>27</sup> Our three-factor model was similar to the Otte et al.<sup>11</sup> model derived from their sample of women with hot flashes in late perimenopausal and early postmenopausal stages, and the Cole et al.<sup>12</sup> model in their sample of older men and women. Morris et al.<sup>27</sup> also favored a three-factor model, but Component 6 (medication) was a single factor. Researchers who proposed a two-factor or three-factor PSQI measure suggest that Component 6 be eliminated.<sup>11,13</sup> Our findings support this suggestion, as omitting Component 6 only slightly improved the omega and was unrelated to any actigraphy measure. However, it should be noted that very few women in our sample reported using sleep medication in the past month, and further replication of this finding is warranted.

### *Limitations*

This study is limited by the small sample size, and findings should be interpreted with caution due to less representation

from perimenopausal women. Our small sample may have limited the ability to detect statistical significance, as indicated by small-to-medium effect sizes ( $d$ -values  $<0.80$ ) with  $p$ -values  $>0.05$ . In addition, actigraphy data were collected for two nights that reflected weekdays specifically rather than the PSQI past-month time frame, resulting in correlations weaker than would be expected with measures obtained simultaneously. However, our correlations were of similar magnitude as reported in other studies, our sample was healthy with little change expected over the course of a few weeks, and our methodology was similar to other PSQI validation studies.<sup>7–9</sup> It should also be noted that women were enrolled in the study between 1997 and 2001, and this may compromise historical validity. The modest correlations between actigraphy and PSQI components may also be a consequence of their 0–3 response range, and keeping PSQI parameters as continuous variables for future validation studies may not only be warranted but also more clinically relevant. Finally, our small sample size was not sufficient for stable factor structures using confirmatory factor analysis, and confirmatory fit indices such as root mean square error of approximation should be considered in future research with larger samples.

### Conclusions

Our findings support the PSQI global score as a valid approximation of sleep quality in healthy women (40–50 years) before menopause when ovarian hormones are still fluctuating. PSQI convergent validity was demonstrated, justifying its use to evaluate sleep quality in premenopausal and perimenopausal women. As researchers move toward shorter versions of the PSQI,<sup>28</sup> it would be prudent to minimize the attention on daytime dysfunction in favor of more attention on reasons for sleep disturbance. The finding that a PSQI global score and three of its component scores were influenced by depressive symptoms suggests that either sleep quality is inextricably linked to depressive symptoms or that more nuanced questions are needed in both types of questionnaires. Finally, we determined that the single-factor PSQI had greater internal consistency than a two-factor or three-factor model and was comparable when Components 6 (medication) and 7 (daytime dysfunction) were removed. While our cohort was not at high risk for depression, further research is needed on unique differences between depressive symptoms and complaints of poor sleep quality.

### Author Disclosure Statement

No competing financial interests exist.

### Funding Information

This work was supported by the National Institutes of Health (NIH) (grant nos. R01 NR04259, T32 NR00788) with a supplement from the NIH Office of Research on Women's Health and by the National Black Nurses' Association.

### References

- Baker FC, Lampio L, Saaresranta T, Polo-Kantola P. Sleep and sleep disorders in the menopausal transition. *Sleep Med Clin* 2018;13:443–456.
- Kravitz HM, Avery E, Sowers M, et al. Relationships between menopausal and mood symptoms and EEG sleep measures in a multi-ethnic sample of middle-aged women: The SWAN sleep study. *Sleep* 2011;34:1221–1232.
- Santoro N. Perimenopause: From research to practice. *J Womens Health (Larchmont)* 2016;25:332–339.
- Baker FC, Willoughby AR, Sassoon SA, Colrain IM, de Zambotti M. Insomnia in women approaching menopause: Beyond perception. *Psychoneuroendocrinology* 2015;60:96–104.
- Shaver J, Giblin E, Lentz M, Lee K. Sleep patterns and stability in perimenopausal women. *Sleep* 1988;11:556–561.
- Buysse DJ, Reynolds CF, 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res* 1989;28:193–213.
- Mollayeva T, Thurairajah P, Burton K, Mollayeva, S, Shapiro CM, Colantonio A. The Pittsburgh Sleep Quality Index as a screening tool for sleep dysfunction in clinical and non-clinical samples: A systematic review and meta-analysis. *Sleep Med Rev* 2016;25:52–73.
- Beaudreau SA, Spira AP, Stewart A, et al. Validation of the Pittsburgh Sleep Quality Index and the Epworth sleepiness scale in older black and white women. *Sleep Med* 2012;13:36–42.
- Fontes F, Goncalves M, Maia S, Pereira S, Severo M, Lunet N. Reliability and validity of the Pittsburgh Sleep Quality Index in breast cancer patients. *Support Care Cancer* 2017;25:3059–3066.
- Otte JL, Rand KL, Carpenter JS, Russell KM, Champion VL. Factor analysis of the Pittsburgh sleep quality index in breast cancer survivors. *J Pain Symptom Manage* 2013;45:620–627.
- Otte JL, Rand KL, Landis CA, et al. Confirmatory factor analysis of the Pittsburgh sleep quality index in women with hot flashes. *Menopause* 2015;22:1190–1196.
- Cole J, Motivala S, Buysse D, Oxman M, Levin M, Irwin M. Validation of a 3-factor scoring model for the Pittsburgh Sleep Quality Index in older adults. *Sleep* 2006;29:112–116.
- Nicassio P, Ormseth S, Custodio M, Olmstead R, Weisman M, Irwin M. Confirmatory factor analysis of the Pittsburgh Sleep Quality Index in rheumatoid arthritis patients. *Behav Sleep Med* 2014;12:1–12.
- Choi J, Guiterrez Y, Gilliss C, Lee KA. Physical activity, weight, and waist circumference in midlife women. *Health Care Women Int* 2012;33:1086–1095.
- Gilliss CL, Lee KA, Gutierrez Y, et al. Recruitment and retention of healthy minority women into community-based longitudinal research. *J Womens Health Gender Based Med* 2001;10:77–85.
- Harlow SD, Gass M, Hall JE, et al. Executive summary of the Stages of Reproductive Aging Workshop +10: Addressing the unfinished agenda of staging reproductive aging. *J Clin Endocrinol Metab* 2012;97:1159–1168.
- Marino M, Li Y, Rueschman MN, et al. Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep* 2013;36:1747–1755.
- Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of physical activities: An update of activity codes and MET intensities. *Med Sci Sports Exerc* 2000;32(9 Suppl.):S498–S516.
- Lee IM, Paffenbarger RS, Hsieh CC. Time trends in physical activity among college alumni, 1962–1988. *Am J Epidemiol* 1992;135:915–925.



20. Radloff LS. The CES-D scale: A self-report depression scale for research in the general population. *Appl Psychol Meas* 1977;1:385–401.
21. Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. *But. Commun Methods Meas* 2020;14:1–24.
22. Jackson CL, Patel SR, Jackson WB, Lutsey PL, Redline S. Agreement between self-reported and objectively measured sleep duration among white, black, Hispanic, and Chinese adults in the United States: Multi-ethnic study of atherosclerosis. *Sleep* 2018;41:1–12.
23. Grandner MA, Kripke DF, Yoon I, Youngstedt SD. Criterion validity of the Pittsburgh Sleep Quality Index: Investigation in a non-clinical sample. *Sleep Biol Rhythms* 2006;4:129–139.
24. Sadeh A. The role and validity of actigraphy in sleep medicine: An update. *Sleep Med Rev* 2011;15:259–267.
25. Jones HL, Zak R, Lee KA. Sleep disturbances in midlife women at the cusp of the menopausal transition. *J Clin Sleep Med* 2018;14:1127–1133.
26. Lampio L, Polo-Kantola P, Himanen S, et al. Sleep during menopausal transition: A 6-year follow-up. *Sleep* 2017;40:1–9.
27. Morris JL, Rohay J, Chasens ER. Sex differences in the psychometric properties of the Pittsburgh Sleep Quality Index. *J Womens Health* 2018;27:278–282.
28. Sancho-Domingo C, Carballo JL, Coloma-Carmona A, Buysse DJ. Brief version of the Pittsburgh Sleep Quality Index (B-PSQI) and measurement invariance across gender and age in a population-based sample. *Psychol Assess* 2021;33:111–121.

Address correspondence to:  
*Rochelle S. Zak, MD*  
*Sleep Disorders Center*  
*University of California*  
*2330 Post Street, Suite 420*  
*San Francisco, CA 94115*  
*USA*

*E-mail: rochelle.zak@ucsf.edu*