Sources of Variability in Epidemiological Studies of Sleep Using Repeated Nights of In-Home Polysomnography: SWAN Sleep Study

Huiyong Zheng, Ph.D.[1]; MaryFran Sowers, Ph.D.[1]; Daniel J. Buysse, M.D.[2]; Flavia Consens, M.D.[3]; Howard M. Kravitz, D.O., M.P.H.[4]; Karen A. Matthews, Ph.D.[5]; Jane F. Owens, Dr.P.H.[2]; Ellen B. Gold, Ph.D.[6]; Martica Hall, Ph.D.[2]

[1]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI; [2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA; [3]Department of Neurology and Anesthesiology and Pain Medicine, University of Washington, Seattle, WA; [4]Departments of Psychiatry and Preventive Medicine, Rush University Medical Center, Chicago, IL; [5]Departments of Epidemiology and Psychiatry, University of Pittsburgh, Pittsburgh, PA; [6]Division of Epidemiology, Department of Public Health Sciences, University of California, Davis, CA

SCIENTIFIC INVESTIGATIONS

Study Objective: To quantify sources of night-to-night variability.
Methods: This project was conducted in 285 middle-aged African American, Caucasian, and Chinese women from the Study of Women's Health Across the Nation (SWAN) Sleep Study living in Chicago, the Detroit area, Oakland, and Pittsburgh. The study used 3 repeated nights of in-home polysomnography (PSG) measures. Night 1 data included assessment of sleep staging, sleep apnea, and periodic limb movements, while Nights 2 and 3 focused on sleep staging.
Results: Mean total sleep time (TST) increased substantially from 365 minutes on Night 1 to 391 minutes and 380 minutes, respectively, on Nights 2 and 3. Mean percent sleep efficiency (SE%) for the 3 nights were 83%, 85%, and 85%, respectively. Night 1 sleep values were significantly different than Nights 2 and 3 measures except for S2 (%), S1 (min), and Delta (S3+4)%. Nights 2 and 3 differences in variability were negligible. Obesity, past smoking, and financial strain measures were associated with greater Night 1 vs. Night 2 or Night 3 differences. We concluded that there was significant Night 1 vs. Nights 2 and 3 variability and, though relatively modest, it was sufficient to bias estimates of association. Additionally, personal characteristics including smoking, obesity, and financial strain increased night-to-night variability.
Conclusions: This reports adds new information about between and within person sources of variation with in-home PSG and identifies elements that are essential in the design and planning of future sleep studies of multi-ethnic groups in social and physiological transition states such as the menopause.
Keywords: Polysomnography (PSG), In-home PSG, sleep methods, variability, sleep behaviors, first night effect (FNE)
Citation: Zheng H; Sowers MF; Buysse DJ; Consens F; Kravitz HM; Matthews KA; Owens JF; Gold EB; Hall M. Sources of variability in epidemiological studies of sleep using repeated nights of in-home polysomnography: SWAN Sleep Study. *J Clin Sleep Med* 2012;8(1):87-96.

P olysomnography (PSG) is widely used in clinical and epidemiological research settings to provide objective sleep measures.[1-6] PSG can be conducted using in-home settings or sleep laboratories considering the advantages afforded by variability in the closeness of monitoring, the ability to correct technical problems in a timely manner, and control temperature, noise and other environmental factors to minimize systematic bias. Use of PSG in the home or laboratory setting has substantial costs for data acquisition and the time required for data processing while potentially imposing physical and psychological burdens on the participant.[7] Given these considerations, it is important to ascertain how much and what kind of data must be collected, determine if more than a single night's data collection is required to describe sleep behaviors; and, identify personal characteristics associated with substantially increased within-person variation in sleep behaviors.

Considering the number of data collection nights that are needed to provide unbiased estimates of sleep characteristics is often made more complex because of the impact of the "first night" effect (FNE), generated from changes in the sleep envi-

BRIEF SUMMARY

Current Knowledge/Study Rationale: Use of repeated polysomnography (PSG) in the home or laboratory to describe sleep behavior has substantial costs in money and time for data acquisition and data processing while potentially imposing physical and psychological burdens on the participant. In order to reduce the costs and relieve the burdens for designing and planning of future sleep studies, this report uses 3 nights' in-home PSG data from SWAN Sleep Study to evaluate night-to-night variation, information redundancy, and identify personal characteristics associated with substantially increased within-person variation in sleep behaviors.

Study Impact: Through the evaluation of sources of night-to-night variation with in-home PSG, this reported identifies elements that are essential in the design and planning of future sleep studies of multi-ethnic groups in social and physiological transition states such as the menopause. Two nights of in-home PSG assessment with an appropriate sample size can provide robust parameter estimates of sleep duration, continuity, and architecture in community samples; the identified personal characteristics associated with greater variability between first and second night measures includes smoking, obesity, and financial strain.

ronment, the presence of sleep monitoring instrumentation, and any potential psychological uneasiness of being observed.[8-12] The FNE has been associated with less total sleep time (TST), lower sleep efficiency (SE), more intermittent waking time, and longer REM latency (RL)[9] in clinical or in-home studies.[13,14]

The SWAN Sleep Study evaluated sleep characteristics in 368 African American, Caucasian, and Chinese women across the menopause transition using 3 nights of in-home PSG. Sleep stage scoring and electrocardiograms were used on all study nights while sleep disordered breathing and leg movements (Night 1) as well as skin temperature and snoring sensors (Night 2 only) were used on selected nights. We evaluated: (1) the magnitude of night-to-night variability on PSG-processed sleep measures; (2) the loss of information if PSG studies of women were restricted to 1, 2, or 3 nights; and (3) sources of within-person variation in the 3 nights of study.

METHODS

The SWAN Sleep Study is a comprehensive study of sleep nested within the ongoing, larger parent longitudinal cohort SWAN study and conducted at 4 of the 7 clinical sites. This 2003 to 2005 time frame overlapped the 5th – 7th annual core SWAN protocol examinations.

SWAN Study Design and Participants

SWAN, a community-based, multisite cohort study of the menopausal transition, enrolled 3,302 women, aged 42-52 years, at its 1996 baseline.[15] Each clinical site recruited Caucasian women. Also recruited were African American women in Boston, Chicago, Detroit area, and Pittsburgh, Chinese women in Oakland, Japanese women in Los Angeles, and Hispanic women in Newark. Women were excluded from cohort enrollment if they were pregnant, using exogenous hormones in the 3 months prior to the baseline interview, had not had menstrual bleeding in the 3 months prior to the baseline interview, or had a hysterectomy. Institutional review boards approved the study, and women gave signed, written informed consent to participate.

SWAN Sleep Study Design and Participants

The SWAN Sleep Study was a nested cross-sectional study of sleep patterns at mid-life.[16,17] A cohort of 370 was enrolled, including 328 pre- and peri-menopausal and 42 postmenopausal African American, Caucasian, and Chinese women, aged 48 to 59 years, from the Chicago, Detroit area, Oakland CA, and Pittsburgh SWAN sites. Women with surgical menopause (< 1%) or using hormone therapy (approximately 23% of the cohort by SWAN follow-up visit 5) were excluded. Exclusion criteria also included factors that could affect sleep including ongoing treatment for cancer or rotating or night shift employment (exclusion rates for these measures were between 1% and 3%). Two of the 370 women had no PSG study and were excluded from this analysis.

Sleep Study Protocol

The sleep protocol was initiated within 7 days of the beginning of the follicular phase of the menstrual cycle in women who were still menstruating. Three consecutive nights of in-

home PSG studies used the Vitaport 3 (VP3) PSG monitor (Temec, Netherlands). Night 1 included a sleep disorders screening PSG montage with 2 channels of electroencephalography (EEG) (C4/A1, C3/A2), bilateral electro-oculograms (EOG), bipolar submental electromyograms (EMG), and one channel of electrocardiogram (EKG). Sleep disordered breathing was assessed using nasal pressure and oral-nasal thermistors to measure airflow; impedance plethysmography characterized chest and abdominal wall movements; and fingertip oximetry (Nonin X-pod model 3012) to measure oxyhemoglobin saturation. Bilateral anterior tibialis EMG was used to assess periodic leg movements (PLM), and characteristics of restless legs were quantified by self-reported questionnaire.[18] On Nights 2 and 3, a sleep staging montage was deployed, which included the EEG, EOG, submental EMG, and EKG channels but not nasal pressure, airflow, oximetry, respiratory effort, or anterior tibialis measurements. Because studies were conducted overnight in participants' homes, technicians were not present to replace sensors and electrodes during the studies. PSG study failure was defined as follows: for the sleep screening night the PSG had to include ≥ 4 h of scorable data for sleep staging and oximetry, concurrent with scorable data for ≥ 1 of the following: nasal pressure cannula, thermistor, or inductance plethysmography belt. For sleep staging PSGs, scorable data were required for 100% of the recording time for at least one EEG channel, one EOG channel, and the EMG channel. In the SWAN Sleep Study,

the overall PSG failure rate was 6.25% (i.e., $1 - \dfrac{1035}{368 \times 3}$, the de-

nominator $368 \times 3 = 1104$ is the total number of *expected* PSG studies for 368 women who participated in PSG studies and the numerator 1035 is the total number of *scorable* nights including repeat studies conducted when initial studies were inadequate), which compares favorably with that of other in-home PSG studies, such as the 5% to 9% failure rate reported in the Sleep Heart Health Study.[19]

Sleep was visually scored in 20-sec epochs on each night using standardized scoring criteria.[16,20] This study was initiated prior to the recent publication of the American Academy of Sleep Medicine's scoring criteria. Rechtschaffen and Kales criteria recommend either 20- or 30-sec scoring epochs. The University of Pittsburgh laboratory used 20-sec epochs for 2 reasons. First, 20-sec epochs provide slightly finer-grained measures of sleep and wakefulness with less potential misclassification (since each epoch can receive only one stage score, up to 50% of an epoch may be another stage). Second, algorithms for quantitative EEG measurement with power spectral analysis used 4-sec epochs, and alignment with visually scored sleep data was more precise if scoring epochs were multiples of 4 seconds. Measures of sleep duration included time in bed and time spent asleep (TST). Time in bed was calculated as time from reported lights out (with confirmation of PSG signals consistent with reduced activity) to time of reported awakening from sleep (again with confirmation of PSG signals consistent with increased activity). TST was calculated as total minutes scored as stages 1 to 4 of NREM sleep and REM sleep. Sleep continuity was quantified by measures of sleep latency (SL [time in minutes from beginning of the recording period to the first consecutive 10 min of stage 2 or

stage 3-4 sleep interrupted by ≤ 2 min of stage 1 or wakefulness]); wakefulness after sleep onset (WASO [total minutes of wakefulness between sleep onset and verified awakening in the morning]), and sleep efficiency (SE [time spent asleep/time in bed × 100]). Measures of sleep architecture included minutes and percent of time spent asleep spent in NREM stages 1, 2, and 3 + 4, and REM sleep.

Sociodemographic Information

Race/ethnicity was determined by self-designation as African American, Caucasian, or Chinese. Other sociodemographic variables included age (continuous variable), marital status (single/never married, married or living as married, separated/widowed/divorced), and educational attainment (high school graduate or less, some college, college graduate, graduate studies). A 3-level response to a question about difficulty in paying for basics (very, somewhat, or not very difficult) including food, shelter, and health care was used as an indicator of financial strain. Study site designation was included in statistical models.

Physical and Mental Health Variables

Self-perceived overall health was coded as excellent, very good/good, fair/poor. Body mass index (BMI) was computed as measured weight in kilograms divided by height in meters squared. Depressive symptoms were assessed with the Center for Epidemiologic Studies Depression (CES-D, depressed vs. not depressed) Scale administered at the closest annual core SWAN visit preceding the Sleep Study.

Menopause transition status was designated as using annual Core SWAN data into one of the following 4 categories: premenopausal (no change in menstrual bleeding regularity); early perimenopausal (menses in the preceding 3 months with an increase in bleeding irregularity); late perimenopausal (menses in the previous 12 months, but not the previous 3 months; and postmenopausal (≥ 12 months of amenorrhea).[21]

Daily medication use (prescription and over-the-counter), recorded at Sleep Study protocol inception and from daily diaries was coded according to the World Health Organization Anatomical Therapeutic Chemical (ATC) classification.[22] Physical activity was measured at the annual core SWAN visits assessing 3 domains (sports, leisure, and household activities) and was treated as a continuous variable. Responses about smoking frequency, alcohol consumption, and caffeine consumption were determined from the daily sleep diaries. Smoking behavior was classified as current, past, and never. Current smokers were those who reported smoking ≥ 7 cigarettes in the 2-week period initiated by the sleep protocol.

Data Analysis

Of the 370 participants enrolled in the SWAN Sleep Study, 368 had PSG data: 364 completed Night 1 (sleep screening PSG), 342 completed Night 2 (sleep staging PSG), and 329 completed Night 3 (sleep staging PSG). These numbers include individuals who repeated PSGs when the initial study failed, yielding an overall PSG success rate of 93.8% (364 + 342 + 329 = 1035 successful studies, versus 368 × 3 = 1104 expected PSG studies; 1035 / 1104 = 0.938). Aggregating these data, 365 women had at least one sleep-staging PSG (i.e., Night 2

or Night 3); 361 had the sleep-screening PSG (Night 1) plus at least one sleep-staging PSG; 306 had both sleep-staging PSGs; and 303 had all 3 nights (including repeat studies for initial study failures). All 3 nights were completed in the protocol-planned order by 285 women (sleep screening PSG, sleep staging PSG 1, sleep staging PSG 2), which formed the dataset for these analyses. Data from 83 women who had at least one PSG were excluded due to non-scorable (n = 65) night(s) or failing to follow the temporal sequence (n = 18). The latter case pertained, for instance, to women whose initial screening study failed, and was repeated on another night.

Univariate statistics were computed for continuous variables and frequencies were determined for categorical variables. Variables with highly skewed distributions were transformed or categorized. Statistical significance was based on p-values from 2-sided tests at a value of p < 0.05.

A one-way repeated measures analysis of variance was used to evaluate the temporal effect of study nights on PSG measures. Orthogonal contrasts were used to compare measures across these 3 nights. The difference between the average of Nights 2 and 3 versus Night 1 was also compared. To evaluate whether having 2 subsequent nights added more information than a single night, a multivariate regression model with random design matrix was used.[23]

$$\mathbf{Y}_i = \xi_0 + \mathbf{\Xi}^T \mathbf{X}_i + \mathbf{e}_i, \ i = 1,2,\ldots,N$$

where N is the total number of subjects. $\mathbf{Y}_i$ was the collection of measurements to be removed (e.g., Night 3) and $\mathbf{X}_i$ was the collection of measurements to be retained (e.g., Night 1 and Night 2). The loss-of-information, defined by normalized mean squared error (MSE) of the residuals, was used to quantify the effect of removing some nights of PSG measurements.

Within-person variation for individual sleep measures was assessed using intraclass correlation coefficients (ICC) with 95% confidence bands. The Bland and Altman approach was used to identify the relationship between the means of 2 nights of sleep measurements and the difference between them.[24,25] A sign rank test was used to evaluate the hypothesis that the mean difference between nights was not equal to zero. Intra-individual (within person) and inter-individual (between persons) variation was calculated and placed in a ratio to describe the relative magnitude of each source of variation.

Stepwise regression analyses were use to relate personal characteristics of study participants with sleep characteristics, with a p-value of 0.05 as the inclusion criterion. Goodness of fit of models was assessed graphically and with the Akaike Information Criterion (AIC).

SAS 9.1 (SAS Institute, Cary, NC) and SAS macro facility were used in performing the statistical analyses and plot the findings.

RESULTS

Characteristics of Study Participants

Characteristics of the total sample with PSG were similar to characteristics of the analytical sample of 285 women (**Table 1**). Women in the analytical sample had a median age of

Table 1—Comparing characteristics of women having the night 1 visit and at least one additional sleep staging night (n = 361) in relation to having 3 consecutive nights in the temporal order specified by the protocol, SWAN Sleep Study, 2003 to 2005

Variable	With Night 1 and at least one additional sleep staging night, N = 361	With 3 consecutive nights in temporal order, N = 285
	Median (IQR*)	Median (IQR*)
Age, years	52.0 (3.0)	52.0 (3.0)
Body mass index, kg/m^2	28.1 (10.8)	27.4 (9.3)
Physical Activity, continuous score	7.8 (2.4)	7.8 (2.4)
Apnea-hypopnea Index, events/h	5.0 (10.4)	4.9 (10.1)
Periodic Leg Movement Index, events/h	2.3 (4.4)	2.4 (4.4)
Obesity Status	N (%)	N (%)
BMI < 30	212 (58%)	175 (61%)
BMI ≥ 30	153 (42%)	110 (39%)
Financial Strain (How hard to pay for basics)	15 (4%)	9 (3%)
Very hard	83 (23%)	57 (20%)
Somewhat hard		
Not hard	266 (73%)	218 (77%)
Race/Ethnicity		
African American	136 (37%)	94 (33%)
Chinese	59 (16%)	48 (17%)
Caucasian	170 (47%)	143 (50%)
Education		
≤ High school	61 (17%)	48 (17%)
Some college	115 (32%)	85 (30%)
≥ BS degree	184 (51%)	148 (53%)
Health Status		
Worse	46 (13%)	30 (11%)
Same	106 (30%)	83 (30%)
Better	206 (58%)	168 (60%)
Smoking		
Never	238 (65%)	189 (66%)
Past	87 (24%)	66 (23%)
Current	40 (11%)	30 (11%)
Marital Status		
Single	57 (16%)	44 (16%)
Married	225 (63%)	189 (67%)
Not married	76 (21%)	48 (17%)
Restless Legs Syndrome (RLS)		
Any RLS	80 (22%)	66 (23%)
No RLS	285 (78%)	219 (77%)
CES-D Score		
Not depressed	305 (86%)	240 (87%)
Depressed	48 (14%)	36 (13%)
Taking Sleep Medications		
No	260 (72%)	206 (73%)
Yes	99 (28%)	75 (27%)
Menopausal Status		
Pre- or early perimenopause	240 (66%)	190 (68%)
Late perimenopause	77 (21%)	58 (20%)
Surgical or postmenopause	48 (13%)	37 (13%)

*IQR, interquartile range.

Table 2—Comparisons of selected sleep measurements during 3 consecutive nights with PSG measures, SWAN Sleep Study, 2003 to 2005

Variables	Night1 Mean (SE)	1st order difference*						2nd order difference**	
		Δ_{2-1}*	p	Δ_{3-1}*	p	Δ_{3-2}*	p	Mean	p
TST (Minutes)	364.52 (4.01)	26.16	< 0.0001	15.2	0.005	-11	0.06	-37.15	< 0.0001
$_{log}$SL (Minutes)	2.71 (0.05)	-0.16	0.01	-0.22	0.0003	-0.05	0.72	0.11	0.26
$_{log}$WASO (Minutes)	3.82 (0.04)	-0.14	0.003	-0.15	0.003	-0.01	0.99	0.12	0.08
SM (Percent)	86.93 (0.43)	1.97	0.0001	1.89	0.0007	-0.08	0.99	-2.05	0.01
SE (Percent)	82.79 (0.51)	2.47	< 0.0001	2.12	0.002	-0.35	0.90	-2.82	0.002
DELTA (Percent)	3.32 (0.28)	0.23	0.49	0.14	0.90	-0.09	0.94	-0.32	0.26
DELTA (Minutes)	11.76 (1.01)	1.95	0.001	1.35	0.28	-0.6	0.74	-2.55	0.02
REM (Percent)	23.18 (0.37)	1.87	< 0.0001	1.63	0.0005	-0.24	0.91	-2.11	0.002
REM (Minutes)	85.76 (1.81)	12.83	< 0.0001	9.13	0.0002	-3.7	0.20	-16.54	< 0.0001
NUMA (Counts)	19.74 (0.45)	1.42	0.003	0.81	0.21	-0.61	0.40	-2.03	0.006
NREM (Minutes)	278.76 (3.07)	13.33	0.0003	6.03	0.23	-7.29	0.09	-20.62	0.0004

*1st order difference: Δ_{2-1} = Night 2−Night 1; Δ_{3-1} = Night 3−Night 1; Δ_{3-2} = Night 3−Night 2.
**2nd order difference = (Night 3−Night 2) − (Night 2−Night 1) = Night 1 − 2 Night 2 + Night 3.

Table 3—Percent differences (Δ%) in selected sleep measures between nights with p-value of tests showing the difference is not equal to zero, and the Bland-Altman p-value for detecting possible bias, SWAN Sleep Study, 2003 to 2005

Sleep variables	Night 2 versus Night1			Night 3 versus Night 1		
	Δ%*	Rank Test Δ% = 0	Bland-Altman p_{BA}	Δ%*	Rank Test Δ% = 0	Bland-Altman p_{BA}
TST (Minutes)	7.4	0.0000	0.44	4.0	0.0008	0.99
SL (Minutes)	-13.0	0.006	0.48	-18.9	0.0000	0.06
WASO (Minutes)	-12.7	0.0003	0.72	-12.8	0.0008	0.64
SM (Percent)	2.3	0.0000	0.09	2.3	0.0004	0.15
SE (Percent)	3.1	0.0000	0.12	2.6	0.0004	0.23
DELTA (Percent)	12.3	0.06	0.49	14.2	0.02	0.06
DELTA (Minutes)	18.2	0.0025	0.004	17.9	0.003	0.005
REM (Percent)	9.7	0.0000	0.04	7.1	0.0002	0.13
REM (Minutes)	16.3	0.0000	0.40	10.5	0.0001	0.42
NUMA (Counts)	7.4	0.0008	0.94	3.2	0.13	0.14
NREM (Minutes)	5.0	0.0001	0.33	2.0	0.09	0.84

*Δ%, percent change of night b vs. night a, $100\% \times \dfrac{b - a}{\frac{1}{2}(a + b)}$

52 years (IQR = 3) and a median BMI of 27.5 kg/m², similar to the overall Sleep Study sample.

Information obtained only during Night 1 included the median apnea-hypopnea index (AHI, Night 1), which was 4.9 (IQR = 10.1) episodes/h of sleep, and the median number of periodic leg movements with arousal (PLMAI), which was 2.4 (IQR = 4.4)/h of sleep, respectively.[26,27] Sixty-six (23.2%) women self-identified as having restless leg syndrome (RLS).[18]

## Comparisons of PSG Measurements during 3 Consecutive PSG Nights

Mean TST increased from Night 1 (365 min) to Night 2 (391 min) and Night 3 (380 min) (**Table 2**). Mean SE% for each of the 3 nights were 83%, 85%, and 85%, respectively (**Table 2**). As seen in **Table 2**, when comparing Night 1 to Nights 2 and 3, all but 4 measures (S2 %, S1 [min], S2 [min] and Delta %) were different from each other. Three measures (Delta minutes,

NUMA, and NREM) were statistically different between PSG Nights 1 and 3, but not PSG Nights 1 and 2. No statistically significant differences in the sleep measures were observed between Nights 2 and 3. Delta (%) was the only variable without significant mean differences across the 3 nights.

Agreement and Variation in Data According to PSG Night

The daily difference in values and variation in data according to the different PSG nights was evaluated using the Bland and Altman approach to estimate the bias that can be discerned with repeated assessments (in **Table 3** and **Figure 1**). For best agreement between 2 nights, the mean percent difference (or mean difference, night 2 – night 1) between 2 measures should be close to zero, with no significant correlations between the mean values and differences, i.e., the dispersion of the difference scores should be limited. While the signed rank test indicated that most differences between nights were greater

Figure 1—Bland Altman plots* for total sleep time (TST-Top Figure) and sleep efficiency (SE-Bottom Figure) identifying the mean differences and bias between sleep night 1 with instrumentation for assessing respiration and restless legs versus sleep night 2 without that instrumentation, SWAN Sleep Study, 2003 to 2005
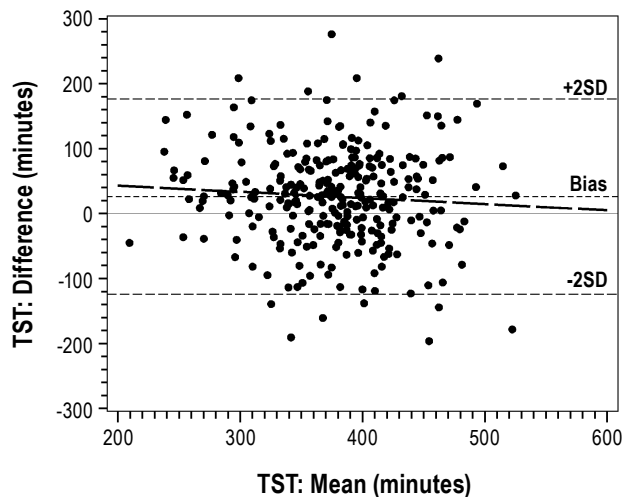
Slope = -0.095; Bias (95% CI) = 26.16 (17.40, 34.91); p value for test of the hypothesis that bias = 0: < 0.0001.
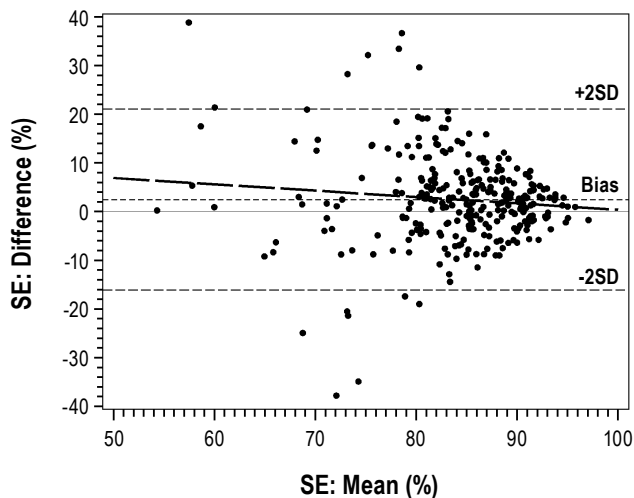
Slope = -0.13; Bias (95% CI) = 2.47 (1.39, 3.55); p value for test of the hypothesis that bias = 0: < 0.0001.

*Mean: (night 1 + night 2)/2; Difference: night 2 − night 1; Bias: mean of the Difference. The short dashed lines are +2SD, bias and -2SD, respectively; long dashed lines are slopes.

than zero (**Table 3**), significant Bland Altman (BA) correlations were observed only in S1 (%), Delta (min), REM (%), RL (min), and RLMA (min) in comparing Night 2 to Night 1, indicating systematic differences between those 2 nights. No significant BA correlations were observed when comparing Nights 2 and 3, indicating no systematic differences between these nights.

Though the BA correlations for some sleep measures (e.g., TST and SE) were not statistically significant, the mean differences shown in **Table 2** and percent change shown in **Table 3** and the BA plots (**Figure 1**) indicated potentially systematic differences between Night 1 and Night 2, especially for SE. TST and SE mean differences between Night 1 and Night 2 were non-zero, indicating a systematic difference between Nights 1 and 2. In contrast, individual PSG measures showed a high degree of agreement of between Nights 2 and 3.

Within- and Between-Person Variation

To describe the variation between nights, **Table 4** shows the intraclass correlation coefficients (ICCs) and their 95% confidence intervals. The highest ICCs were in Delta (%) and Delta (min) and were 0.68 and 0.66, respectively, for Night 1 to Night 3. The ICCs for Delta (%) and Delta (min) were 0.78 and 0.77, respectively, for Night 2 to Night 3. Likewise, the ICCs for Delta (%) and Delta (min) were 0.80 and 0.80, respectively, for Night 1 to Night 2 (data not shown). As seen in **Table 4**, measures of sleep continuity and duration were more likely to have the lower ICCs than selected measures of sleep architecture.

We disaggregated the within- and between-woman variation when comparing data from 2 different nights (**Table 4**). A low within-person variation relative to the between-person variation is generally considered optimal to characterize group differences. Delta % and Delta minutes measures included greater between-person variation relative to the amount of within-person variation. This was associated with markedly greater ratios of inter-individual variation to intra-individual variation (i.e., $\sigma^2_{inter} / \sigma^2_{intra}$); the ratios comparing Night 1 to Night 3 were 2.11 and 1.98, respectively (**Table 4**). Other sleep measures had substantially more within-person variation than between-person variation and lower ratios (i.e., 0.38 [TST] and 0.33% [SE]).

Loss-of-Information: Two Sleep-Staging PSG Nights or One Sleep-Staging PSG Night

When it was identified that there was high correlation between these measures according to night, it was logical to consider how much less variation is explained should the number of study nights be reduced. The amount of information lost (less variation explained) was about 23.6% or 23.5% of the total variation if only "Night 1 + Night 2" or "Night 1 + Night 3" were used. In contrast, removing any 2 of 3 nights could result in the loss of more than half the information.

Participant Characteristics and Night-to-Night Sleep Measures Variability

Characteristics associated with having greater differences in the measures of Night 1 vs. Night 2 (i.e., Night 2 − Night 1) included obesity, financial strain, race/ethnicity, marital status, smoking, and PLMAI (see **Table 5**). While education, menopause status, and physical activity were evaluated, these were not significantly related to differences in measures between Nights 1 and 2 (data not shown). Participant characteristics were not associated with differences in night-to-night variability for Delta (min), NREM (min), NUMA, REM (min), S2 (min), and TST (min).

Table 4—Intraclass correlation coefficients (with upper and lower 95% confidence intervals), intra- and inter-individual variability and the ratio of inter- to intra-individual variability for selected sleep measures, SWAN Sleep Study, 2003 to 2005

Sleep variables	(Night1, Night3)				(Night2, Night3)			
	ICC (95% CI)*	Within σ^2_{intra}**	Between σ^2_{inter}**	Ratio $\sigma^2_{inter}/\sigma^2_{intra}$	ICC (95% CI)*	Within σ^2_{intra}**	Between σ^2_{inter}**	Ratio $\sigma^2_{inter}/\sigma^2_{intra}$
TST (Minutes)	0.28 (0.17,0.38)	3493	1330	0.38	0.29 (0.18, 0.39)	3179	1328	0.42
$_{log}$SL (Minutes)	0.41 (0.31, 0.50)	0.45	0.31	0.68	0.39 (0.29, 0.48)	0.47	0.31	0.65
$_{log}$WASO (Minutes)	0.35 (0.24, 0.45)	0.29	0.15	0.54	0.51 (0.42, 0.59)	0.23	0.24	0.51
SM (Percent)	0.28 (0.17, 0.38)	39	15	0.40	0.43 (0.33, 0.52)	28	21	0.76
SE (Percent)	0.25 (0.14, 0.36)	58	19	0.33	0.37 (0.27, 0.47)	44	25	0.58
S1 (Percent)	0.44 (0.34, 0.53)	17	13	0.77	0.62 (0.54, 0.69)	9	15	1.62
S2 (Percent)	0.49 (0.40, 0.57)	33	31	0.95	0.56 (0.48, 0.63)	25	32	1.28
S1 (Minutes)	0.47 (0.37, 0.56)	197	178	0.90	0.63 (0.55, 0.70)	140	237	1.70
S2 (Minutes)	0.46 (0.36, 0.55)	1482	1271	0.86	0.45 (0.35, 0.54)	1400	1126	0.80
DELTA (Percent)	0.68 (0.61, 0.74)	7	15	2.11	0.78 (0.73, 0.82)	4.5	16	3.57
DELTA (Minutes)	0.66 (0.59, 0.72)	95	187	1.98	0.77 (0.72, 0.81)	67	229	3.41
REM (Percent)	0.31 (0.20, 0.41)	29	13	0.45	0.39 (0.29, 0.48)	22	14	0.64
REM (Minutes)	0.20 (0.09, 0.31)	754	193	0.26	0.31 (0.20, 0.41)	626	284	0.45
NUMA (Counts)	0.49 (0.40, 0.57)	30	28	0.95	0.56 (0.48, 0.63)	26	33	1.26
NREM (Minutes)	0.36 (0.25, 0.46)	1792	1008	0.56	0.36 (0.25, 0.46)	1695	956	0.56
RL (Minutes)	0.37 (0.27, 0.47)	1459	873	0.60	0.43 (0.33, 0.52)	1119	855	0.76
RLMA (Minutes)	0.49 (0.40, 0.57)	791	752	0.95	0.51 (0.42, 0.59)	694	709	1.02

*ICC (95% CI), Intraclass correlation coefficient and 95% CI, ICC = $\sigma^2_{inter} / (\sigma^2_{inter} + \sigma^2_{intra}) = \sigma^2_{between} / (\sigma^2_{between} + \sigma^2_{within})$.
**σ^2_{intra}, intra-individual variability, within subject; σ^2_{inter}, inter-individual variability, between subject.

Table 5—Beta estimates (with standard errors [SE] and p-values for the beta estimate) and standardized beta coefficients for participant characteristics in relation to selected sleep measures between night 2 and night 1, SWAN Sleep Study, 2003 to 2005

Sleep measures	Participant characteristics	Variable levels	Beta estimates			
			Beta	SE	p value	Beta(s)
$_{log}$SL (Minutes)	Financial strain (Difficulty paying for basics?)	very hard	0.31ns	7.96	0.97	0.002
		somewhat hard	10.05*	3.63	0.006	0.17
S1 (Minutes)	BMI	BMI ≥ 30	5.03*	1.91	0.009	0.16
	PLMAI		-0.39*	0.17	0.02	-0.14
S1 (Percent)	BMI	BMI ≥ 30	1.36*	0.55	0.01	0.15
	PLMAI		-0.13*	0.05	0.01	-0.16
DELTA (Percent)	PLMAI		0.09*	0.03	0.009	0.16
	Marital status	Single	-1.15*	0.51	0.03	-0.14
		separated/widow/divorce	0.53ns	0.49	0.28	0.07
	Smoking	Past	-1.39*	0.45	0.002	-0.19
		Current	-0.21ns	0.61	0.73	-0.02
REM (Percent)	Race	African American	-1.93*	0.91	0.03	-0.14
		Chinese	0.58ns	1.14	0.61	0.03

ns p > 0.05; *p < 0.05; **p < 0.001; ***p < 0.0001. Intercepts were not reported in the table. Beta(s), standardized Beta.
 Only significant factors (p < 0.05) were retained in parsimonious models using stepwise selection.

DISCUSSION

In evaluating the temporal night effect and patterns of variation arising from in-home PSG-derived measures with repeated nights of assessment, small but statistically significant differences in measures of sleep were observed when comparing Night 1 of data collection with Nights 2 and 3. Little or no difference was observed when comparing data from Nights 2 vs. 3. Mean total sleep time (TST) was increased 4% to 7% from 365 minutes in Night 1 to 391 minutes and 380 minutes, respectively, in Nights 2 and 3. All sleep measures on Night 1 were

significantly different from nights 2 and 3, except for S2 (%), S1 (minutes), and Delta %.

We used multiple nights of evaluation to address whether the degree of intra-individual variation exceeded the between-person comparisons. To that end, we reported the between- and within-person variation and the ratio of these two measures. The within-person variation for most sleep measures was greater than the between-person variation, when comparing nights with different amounts of monitoring instrumentation. The exception included measures of Delta % and Delta minutes, indices of sleep architecture, for which substantial within- and between-woman variation were observed while comparing nights with differing levels of instrumentation. This may help to explain why there is less discernible night-to-night variation was detected in these measures when nights have different levels of monitoring instrumentation.

These data do not suggest a major advantage of having more than two repeated measures to characterize habitual sleep with a sample size of approximately 280 women unless the focus of the assessment is on sleep architecture. These data also suggest that clinical studies focused on sleep architecture, and specifically Delta measures, might consider that only a single night is needed to describe specific patient groups. This would then reflect that this group of sleep measures had greater between-person variability (as compared to within-person variation) in comparison to measures of sleep duration and continuity.

While PSG is widely regarded as the gold standard in assessing sleep, its administration, even as an in-home study, also generates an environment that may curtail "usual" sleep. The SWAN Sleep study implemented an in-home protocol to be able to characterize the participants' usual sleep and mitigate the "first night" effects (FNE) that include sleeping in an unfamiliar environment and anxiety about being observed.[28-32] FNEs have been reported to result in less total sleep time (TST), greater rapid eye movement (REM), lower sleep efficiency (SE), more intermittent wake time, and longer REM latency.[9] Le Bon[9] studied two consecutive PSG nights in 83 patients with chronic fatigue syndrome and observed significant differences in SPT, TST, SE, SE minus sleep onset latency, REM sleep, sleep onset latency, and RL, and concluded that no single sleep variable could summarize the FNE. In a study of 36 healthy adults,[30] Sforza indicated that the FNE on arousal response was affected by individual susceptibility and circadian and homeostatic influences.

While participants were evaluated in-home, instrumentation to monitor respiration, sleep disordered breathing and leg movements were included in the Night 1 protocol, elements that were not present in Nights 2 and 3. Though we identified statistically significant night-to-night variation, particularly of Night 1 vs Nights 2 and 3, the magnitude of the differences was smaller than expected. Depending upon the sleep measure of interest, the differences were as little as 2% but not more than 20%. Study night or the presence of FNE was confounded with the PSG assessment protocol (i.e., the Night 1 assessment included additional electrodes and monitors to assess sleep disordered breathing and limb movements, whereas fewer signals were collected during Nights 2 and 3). Different PSG montages were used on the screening night and subsequent study nights. It is possible that the montage itself contributed to the difference in sleep from the first to the second and third nights, thereby confounding the potential effects due to montage and those due to study order. For instance, the greater amount of instrumentation on the first night may have led to greater sleep disruption, above and beyond any night order effect. However, many other studies that have used identical montages across multiple nights have reported similar findings to our own, making this explanation somewhat less compelling. A different study design would be required to disaggregate the effect of wearing additional monitoring devices as compared to night effect.

It should be noted that sleep stage scoring in our study differed from that currently recommended by the AASM scoring rules. In particular, we used 20-second scoring epochs, and all stages were scored using a central EEG derivation. These methods may limit generalizability for current research and clinical practice. On the one hand, shorter scoring epochs may lead to less misclassification of sleep and wakefulness: since individual epochs may contain up to 50% of another sleep-wake stage, 20-second epochs contain a maximum or 10 seconds of misclassified data, whereas 30-second epochs contain a maximum of 15 seconds of misclassified data. On the other hand, use of a single EEG derivation may provide less precise staging, since some EEG characteristics are better defined in occipital (alpha rhythm) or frontal (delta activity) derivations.

Characteristics associated with having greater differences in the measures of Night 1 vs. Nights 2 or 3 included financial strain, being obese, and past smoking. These are consistent with a newly reported cross-sectional evaluation of a sample representative of the US population and based on interview rather than instrumented monitoring.[33]

As aforementioned, 83 women who had at least one polysomnogram recording were excluded in this particular analysis. In order to evaluate the possible bias of missing these women (22.4% of the cohort, 22.6% of those with at least one night PSG data), a simulation study was conducted using TST as the outcome, night as the independent variable, and BMI, difficulty in paying for basics, and smoking as covariates. The simulation process was performed on the combination of sample size and strategies using three nights' PSG data following temporal sequence (night 1 and night 2; night 1, night 2 and night 3; night 1, mean of night 2 and night 3). The simulation results showed that about 125 women gave 95% coverage probability to detect "night" effect for all 3 strategies. The coverage probability curve as a function of sample size and strategies increased monotonically leading to the coverage probabilities approaching to 100% after about n = 200. This hints that the bias was minimized by using 285 women, even though 83 women were lost due to non-scorable night(s) or improper temporal sequence.

This study has strengths and limitations. It was an in-home-based PSG study with a substantial sample size and multiple nights of assessment. The ability to disaggregate the sources and magnitude of the night-to-night variability in this study allowed us to reduce the number of study nights in a follow-up study, implemented three years after the baseline. This study was conducted in a sample of healthy middle-aged women to characterize the normal physiological and psychological events of the menopause in relation to sleep, but care must be exercised in extrapolating these findings to studies directed toward samples or study designs that are highly enriched for sleep pathology (i.e., a case-control study of insomnia). Study

night was confounded with the PSG assessment protocol, so the effect of this additional monitoring could not be quantitatively differentiated from the anxiety of the FNE of participating in a sleep study.

As noted above, sleep disordered breathing and leg movements were not monitored on Nights 2 and 3. Intra-individual night-to-night variability in sleep disordered breathing as well as periodic leg movements in sleep is an unresolved concern in clinically assessing these sleep disorders, even when accurately measured, recorded, and analyzed.[6] Because both are associated with sleep fragmentation and loss, to presume night-to-night consistency without monitoring for respiratory events and leg movements can result in residual confounding of measurements of PSG sleep macrostructure.[34] Thus, individual night-to-night differences can be a substantial source of confounding and a limitation in our analysis of variability, and should be considered in determining how many nights of PSG are needed to assess sleep and sleep disorders.[34] Other sources of variation also could not be disaggregated. While African American women were recruited by three sites participating in the Sleep Study, only one site recruited Chinese women, a condition imposed by the design of the parent study. However, this precluded us from disaggregating sources of variation associated specifically with the site and race/ethnicity. This study focused on middle-aged healthy women, so while the study group did not include men, information about sources of variation from a large community-based sample should be helpful in anticipating the sources of variation that might be considered in designing studies of men or the elderly.

In summary, when we evaluated the night-to-night variation, the temporal night effect and patterns of variation arising from in-home PSG data with three repeated nights of study, modest differences in information were obtained on 3 nights of in-home sleep measures, even when Night 1 included a sleep staging and sleep disorder montage with additional instrumentation, whereas Nights 2 and 3 included a sleep staging montage with less instrumentation. This led us to conclude that two nights of in-home sleep assessment with an appropriate sample size can provide robust parameter estimates of sleep duration, continuity, and architecture in community samples. A study that focuses on measures of sleep architecture in which higher between-woman variability was demonstrated relative to the amount of within-woman variation may require a smaller sample size. Personal characteristics associated with greater variability between first and second night measures included smoking, obesity, and financial strain. Understanding the sources of variation can help in planning and developing laboratory and community-based studies of sleep and allow investigators to select appropriate statistical analyses to optimize the identification of important sleep relationships while minimizing bias.

ABBREVIATIONS

AHI, apnea/hypopnea events per hour of sleep

ATC, anatomical therapeutic chemical

BA, Bland-Altman

BMI, body mass index

DELTA%, [(S3 + S4)/TST]*100%, or equivalently SWS% = slow wave sleep%: S3% + S4%

DELTA (min), # minutes in stages 3 and 4 sleep from sleep onset to GMT, or equivalently SWS(min) = slow wave sleep: S3 + S4

CES-D, Center for Epidemiologic Studies Depression Scale

EEG, electroencephalography/electroencephalogram

EOG, electro-oculograms

EKG, electrocardiogram

EMG, electromyograms

FNE, first night effect

GMT, good morning time, wake up

GNT, good night time, turn light off and ready to sleep

ICC, intraclass correlation coefficients

IQR, interquartile range

MSE, mean squared error

NREM (min), # minutes of NREM (non-REM): S1 + S2 + S3 + S4

NUMA, # awakenings after sleep onset lasting ≥ 11 seconds (> 50% of a 20-sec epoch)

PLM, periodic leg movement

PLMAI, periodic leg movement index per hour of sleep with arousals

PSG, polysomnography/polysomnogram

REM, rapid eye movement

REM%, (REM/TST)*100%

REM (min), # minutes in REM sleep

RL (min), REM latency in minutes

RLmA (min), REM latency minus Awake in minutes

RLS, restless legs syndrome

Si(min), # minutes in stage i sleep, i = 1, 2, 3, 4

Si(%), percent in stage i sleep, (Si/TST)*100%, i = 1, 2, 3, 4

SE (%), sleep efficiency, (TST/TRP)*100%

SL (min), sleep latency, time in minutes elapsed from GNT to sleep onset

SM (%), sleep maintenance, (TST/SPT)*100%

SPT (min), sleep period time; the total time in minutes from sleep onset to GMT

SWAN, Study of Women's Health Across the Nation

TRP (min), total recording period; total amount of time in minutes from GNT to GMT

TST (min), total sleep time; the total time asleep in minutes

WASO (min), total time awake in minutes after sleep onset

REFERENCES

1. AARC-APT (American Association of Respiratory Care-Association of Polysomnography Technologists) clinical practice guideline. *Respir Care* 1995;40:1336-43.
2. Bloch KE. Polysomnography: a systematic review. *Technol Health Care* 1997;5:285-305.
3. Madani M, Frank M, Lloyd R, Dimitrova DI, Madani F. Polysomnography versus home sleep study: overview and clinical application. *Atlas Oral Maxillofac Surg Clin North Am* 2007;15:101-9.
4. Normal SY. Polysomnography in children and adolescents. *Chest* 2005;127:1080.
5. Chesson AL Jr, Ferber RA, Fry JM, et al. The indications for polysomnography and related procedures. *Sleep* 1997;20:423-87.
6. Kushida CA, Littner MR, Morgenthaler T, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep* 2005;28:499-521.
7. Su S, Baroody FM, Kohrman M, Suskind D. A comparison of polysomnography and a portable home sleep study in the diagnosis of obstructive sleep apnea syndrome. *Otolaryngol Head Neck Surg* 2004;131:844-50.
8. Tamaki M, Nittono H, Hayashi M, Hori T. Spectral analysis of the first-night effect on the sleep-onset period. *Sleep Biol Rhythms* 2005;3:122-9.

9. Le Bon O, Minner P, Van Moorsel C, et al. First-night effect in the chronic fatigue syndrome. *Psychiatry Res* 2003;120:191-9.

10. Le Bon O, Staner L, Hoffmann G, et al. The first-night effect may last more than one night. *J Psychiatr Res* 2001;35:165-72.

11. Marzec ML, Selwa LM, Malow BA. Analysis of the first night effect and sleep parameters in medically refractory epilepsy patients. *Sleep Med* 2005;6:277-80.

12. Sforza E, Haba-Rubio J. Night-to-night variability in periodic leg movements in patients with restless legs syndrome. *Sleep Med* 2005;6:259-67.

13. 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.

14. Edinger JD, Glenn DM, Bastian LA, et al. Sleep in the laboratory and sleep at home II: comparisons of middle-aged insomnia sufferers and normal sleepers. *Sleep* 2001;24:761-70.

15. Sowers MF, Crawford S, Sternfeld B, et al. Design, survey sampling and recruitment methods of SWAN: A multi-center, multi-ethnic, community-based cohort study of women and the menopausal transition. In: Wren J, Lobo RA, Kelsey J, Marcus R, (eds) *Menopause: biology and pathobiology*. Academic Press, 2000: 175-88.

16. Hall MH, Matthews KA, Kravitz HM, et al. Race and financial strain are independent correlates of sleep in midlife women: the SWAN sleep study. *Sleep* 2009;32:73-82.

17. Sowers MF, Zheng H, Kravitz HM, et al. Sex steroid hormone profiles are related to sleep measures from polysomnography and the Pittsburgh Sleep Quality Index. *Sleep* 2008;31:1339-49.

18. Allen RP. The resurrections of periodic limb movements (PLM): leg activity monitoring and the restless legs syndrome (RLS). *Sleep Med* 2005;6:385-7.

19. Quan SF, Griswold ME, Iber C, et al. Short-term variability of respiration and sleep during unattended non-laboratory polysomnography: the Sleep Heart Health Study. *Sleep* 2002;25:843-9.

20. 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.

21. World Health Organization Scientific Group. Research on the menopause in the 1990s. Report of a WHO Scientific Group. *World Health Organ Tech Rep Ser* 1996;866:1-107.

22. World Health Organization. *Guidelines for ATC Classification*. Available at: http://www.whocc.no/atcddd. Accessed December 10, 2007.

23. Srivastava MS. *Methods of multivariate statistics*. New York: John Wiley & Sons, Inc., 2002.

24. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307-10.

25. Donner A. A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *Int Stat Rev* 1986;54:67-82.

26. American Academy of Sleep Medicine Task Force. Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. *Sleep* 1999;22:667-89.

27. The Atlas Task Force. Recording and scoring leg movements. Sleep 1993;16:749-59.

28. Rechtschaffen A, Verdone P. Amount of dreaming: effect of incentive, adaption to laboratory, and individual differences. *Percept Mot Skills* 1964;19:947-58.

29. Agnew HW Jr, Webb WB, Williams RL. The first night effect: an EEG study of sleep. *Psychophysiology* 1966;2:263-6.

30. Sforza E, Chapotot F, Pigeau R, Buguet A. Time of night and first night effects on arousal response in healthy adults. *Clin Neurophysiol* 2008;119:1590-9.

31. Suetsugi M, Mizuki Y, Yamamoto K, Uchida S, Watanabe Y. The effect of placebo administration on the first-night effect in healthy young volunteers. *Prog Neuro-psychopharmacol Biol Psychiatry* 2007;31:839-47.

32. Scholle S, Scholle HC, Kemper A, et al. First night effect in children and adolescents undergoing polysomnography for sleep-disordered breathing. *Clin Neurophysiol* 2003;114:2138-45.

33. Krueger PM, Friedman EM. Sleep duration in the United States: a cross-sectional population-based study. *Am J Epidemiol* 2009;169:1052-63.

34. Van Dongen HPA, Vitellaro KM, Dinges DF. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep* 2005;28:479-96.

ACKNOWLEDGMENTS

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT

This was not an industry supported study. Dr. Buysse has served as a paid consultant for the following companies: Actelion, Cephalon, Eisai, Eli Lilly, GlaxoSmithKline, Merck, Neurocrine, Neurogen, Pfizer, Philips, Sanofi-Aventis, Sepracor, Servier, Somnus Therapeutics, Takeda, and Transcept Pharmaceuticals, Inc. He has helped produce CME materials and has given paid CME lectures indirectly supported by industry sponsors. The other authors have indicated no financial conflicts of interest.