Short- and Long-Term Sleep Stability in Insomniacs and Healthy Controls

Jordan Gaines, BA[1]; Alexandros N. Vgontzas, MD[1]; Julio Fernandez-Mendoza, PhD[1]; Maria Basta, MD[1,2]; Slobodanka Pejovic, MD[1]; Fan He, MS[3]; Edward O. Bixler, PhD[1]

[1]*Sleep Research and Treatment Center, Department of Psychiatry, Pennsylvania State University College of Medicine, Hershey, PA;* [2]*Department of Psychiatry, School of Medicine, University of Crete, Heraklion, Crete, Greece;* [3]*Department of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA*

Study Objectives: Assess the short- and long-term stability of sleep duration in patients with insomnia and normal-sleeping controls.
Design: Observational short-term and prospective studies.
Setting: Sleep laboratory.
Participants: Patients with insomnia (n = 150) and controls (n = 151) were recruited from the local community or sleep disorders clinic. A subsample of 95 men from the Penn State Adult Cohort (PSAC) were followed up 2.6 y after their initial visit.
Measurements: Participants underwent a physical examination and 8-h polysomnography (PSG) recording for 3 consecutive nights (controls and insomniacs), or 2 single nights separated by several years (PSAC). Intraclass correlation coefficients (ICCs) assessed the stability of the variables total sleep time (TST), sleep onset latency (SOL), and wake after sleep onset (WASO). We also examined persistence of the first-night classification of "short" versus "normal" sleep duration on subsequent nights.
Results: Stability of TST, SOL, and WASO based on 1 night were slight to moderate in both patients with insomnia (ICC = 0.37–0.57) and controls (ICC = 0.39–0.59), and became substantial to almost perfect when based on the average of 3 nights (ICC = 0.64–0.81). We observed similar degrees of stability for TST and WASO in the longitudinal sample, with moderate stability based on a single night and substantial stability based on both nights. In examining the persistence of "short" and "normal" sleep duration, 71.4% (controls), 74.7% (patients with insomnia), and 72.6% (longitudinal sample) of participants retained their first-night classifications over subsequent nights.
Conclusions: Sleep duration variables, particularly total sleep time based on 3 consecutive nights in both patients with insomnia and controls or two single-night recordings separated by several years, are stable and reflect a person's habitual sleep. Furthermore, a single night in the laboratory may be useful for reliably classifying one's sleep duration.
Keywords: good sleepers, insomnia, polysomnography, sleep, stability
Citation: Gaines J, Vgontzas AN, Fernandez-Mendoza J, Basta M, Pejovic S, He F, Bixler EO. Short- and long-term sleep stability in insomniacs and healthy controls. *SLEEP* 2015;38(11):1727–1734.

INTRODUCTION

In the realm of sleep research, one of the most global and hotly debated questions has long been, "How many nights are enough?" Due to the time and expense associated with in-laboratory polysomnography (PSG), sleep researchers and clinicians often have to record participants for a single night only, especially in large epidemiological studies. The issue of night-to-night variability in some components of sleep, however, raises the question of how many nights in the laboratory will provide a sufficient representation of one's habitual sleep patterns.

The concept of the "first-night effect," or the well-recognized observation that participants do not sleep as well during their first night in the laboratory as they would on subsequent nights, has made multinight studies a common practice. Agnew and colleagues[1] first recognized this phenomenon in 1966 in 43 participants who spent 4 consecutive nights in the laboratory. Compared to the following nights, the first night's sleep was significantly more disrupted, with more total wake time and a higher proportion of stage 1 sleep, as well as significantly less rapid eye movement (REM) and delayed slow wave sleep.[1] Subsequent work by Merica and Gaillard[2] in the 1980s suggested that only stage 4 sleep produced stable results across consecutive study nights, and more recent work using quantitative electroencephalography (QEEG) methods has proposed that, compared to visual scoring, only certain components of QEEG-scored sleep exhibit stability across nights.[3–7] Statistical projections have even suggested that it may take between 1 and 3 w to achieve stability of certain sleep components.[8,9] Many of these studies, however, have been confounded by participants of older age, the use of *ad libitum* sleep schedules, or did not take obesity into account—factors that critically affect night-to-night stability of sleep.[10] However, Lorenzo and Barbanoj[11] reported that, in healthy volunteers across separate sessions of consecutive nights in the laboratory, only the "very first night" was significantly different, and only in REM-related variables. Their findings suggest that, once familiar with the PSG equipment, participants exhibit stability in sleep across consecutive nights.[11]

The issue of night-to-night variability in the sleep of individuals with insomnia has also raised questions as to how many nights are enough; however, research on the utility of a single night of PSG has been mixed. A large study of controls, patients with insomnia, and patients with movement and behavioral disorders observed a significant first-night effect in all four groups, although it was most pronounced in patients with insomnia.[12] Similarly, a 2-w actigraphy study of older adults

reported that those with insomnia complaints exhibited significantly more night-to-night variability than those without.[13] However, Vallières and colleagues[14] have demonstrated that even when patients with insomnia are clustered into three groups based on sleep diary variability, the groups do not differ on PSG variables across consecutive nights. Furthermore, Edinger et al.[15] reported that night-to-night variability across three nights in the laboratory do not differ between insomniacs and controls, and have suggested that a single night of PSG may be sufficient when the goal is to determine a diagnosis of insomnia.[16] Additionally, in the past several years, studies have begun to utilize a single night's sleep duration as a predictor of the cardiometabolic risks and neuropsychological deficits associated with chronic insomnia. Specifically, sleep duration extracted from 1 night has been used in large epidemiological samples to classify participants into "short" and "normal" sleepers.[17–22] The accuracy and stability of a single night in categorizing participants into these groups, however, has not been assessed.

Our aim was to evaluate sleep duration stability over the short term (consecutive nights) in relatively young, nonobese samples of both patients with insomnia and controls, as well as over the long term (across several years) in a general population sample of men. Our other major goal was to examine the persistence of subjects' classification as "short" or "normal" sleep duration from the first night to subsequent nights in the laboratory.

METHODS

Participants

The study sample consisted of 150 patients with insomnia and 151 normal-sleeping controls who spent multiple consecutive nights in the sleep laboratory ("short-term" cohort), as well as 95 men who visited the sleep laboratory twice, with several years between visits ("long-term" cohort). Patients with insomnia (52.0% male, mean age 36.6 ± 1.1y) were recruited from both the local community and the Sleep Disorders Clinic at the Penn State Milton S. Hershey Medical Center, complained of chronic difficulty (> 6 mo) in initiating or maintaining sleep,[23–25] and were not receiving treatment or medication for their insomnia at the time of the study. Controls (56.3% male, mean age 36.0 ± 1.2 y) did not have apnea, reported no sleep complaints, and were drawn from screening and baseline nights of studies on sleep restriction (n = 36),[26,27] total sleep deprivation (n = 41),[28] or served as controls in one of two studies examining the effect of continuous positive airway pressure on obstructive sleep apnea (n = 74).[29–32] In addition, a subsample of 95 men (mean age 51.1 ± 1.1 y) from the Penn State Adult Cohort, a random general population sample of 1,741 adults,[33] visited the sleep laboratory on two separate occasions. All studies were approved by the University Institutional Review Board (Pennsylvania State University College of Medicine) and all participants provided written informed consent.

Sleep Laboratory Protocol

During their visit in the laboratory, all participants underwent a physical examination, during which height and weight were recorded and body mass index (BMI) calculated (in kg/m^2). Sleep laboratory recordings were conducted in a sound-attenuated, light- and temperature-controlled room with a comfortable, bedroom-like atmosphere. Each subject was monitored continuously for 8 h (22:30–23:00 until 06:30–07:00) using 16-channel polygraph recordings of EEG, electrooculogram (EOG), and electromyogram (EMG). PSG, respiration (via thermocouple and thoracic strain gauges), and oximeter data were collected using Grass-Telefactor Gamma Sleep Recording software (Middleton, WI, USA). Patients with insomnia and controls spent 3 consecutive nights in the laboratory, and men from the general population cohort spent 2 single nights, with an average of 2.6 y between each visit. Visual sleep stage scoring was conducted by a registered polysomnography technologist based on Rechtschaffen and Kales criteria.[34] Apnea-hypopnea index (AHI, or the number of apneas and hypopneas summed per hour) was also ascertained; an apnea was defined as cessation of airflow for ≥ 10 sec and an out-of-phase strain gauge movement, whereas a hypopnea was defined as a 50% airflow reduction and associated decrease in SaO$_2$ of at least 4%.[33] Those with an AHI of ≥ 5 were excluded from the control and insomnia samples.

Statistical Analysis

For descriptive purposes, sociodemographic and sleep characteristics were compared using independent-samples t-tests and paired-samples t-tests (two-tailed) in patients with insomniacs/controls and the longitudinal cohort, respectively. Specifically, differences in the PSG variables total sleep time (TST), sleep efficiency (SE), sleep onset latency (SOL), wake after sleep onset (WASO), and percentage of stages 1, 2, slow wave sleep (SWS), and rapid eye movement (REM) sleep were assessed between groups (controls/insomniacs) or across time (longitudinal cohort). When appropriate, Pearson's r examined associations between sociodemographic variables and sleep outcomes.

Given their consistent association with cardiometabolic and insomnia-related outcomes in the sleep literature, we were then particularly interested in examining within-subjects short-term (in patients with insomnia and controls) and long-term (longitudinal cohort) stability of variables associated with sleep duration (TST, SOL, and WASO). Intraclass correlation coefficients (ICC), which incorporate both between- and within-subjects variance and are commonly used in test-retest reliability analyses, were computed separately for each outcome. Specifically, under the framework of analysis of variance (ANOVA), the total variance in the data can be attributed to two major systematic sources: between- and within-subjects variances. ICCs are interpreted as the proportion of total variance explained by between-subjects variance; in our study, ICCs reflect the proportion of variance in sleep characteristics explained by differences between participants. As the study participants are considered a random sample of the larger population, the between-subjects factor is treated as a random factor in the analysis. To retain the generalizability of the results, the nights included in the ICC analyses were also considered a random subset of all possible nights. Therefore, two-way random-effects ICCs were calculated to examine the short- and long-term stability of sleep characteristics. To explore the utility of 1 or more nights to attain stability, both "single

measures" (ICC[2,1]) and "average measures" (ICC[2,k]) were calculated based on a single night of data versus the average of multiple (k) nights of data, respectively.[35] Coefficients and their 95% confidence intervals (truncated when < 0) were interpreted using benchmark values established by Landis and Koch[36]: 0.00–0.20 = "poor stability," 0.21–0.40 = "slight stability," 0.41–0.60 = "moderate stability," 0.61–0.80 = "substantial stability," and 0.81–1.00 = "almost perfect stability." For simplicity, and in line with previous studies,[7,36] we denote ICC ≥ 0.60 as variables demonstrating "reliable" short- or long-term stability. By using these cutoffs and incorporating two forms of ICCs, we were then able to examine the utility of either a single night ("single measures") or multiple nights ("average measures") in attaining stable measures of sleep duration. Between- and within-subjects variances were also calculated from ANOVA output as the between- or within-people mean square minus the mean square error, divided by sample size.

To examine the stability of one's sleep duration classification, individuals were categorized as "short" or "normal" sleepers for each of their nights in the laboratory as defined by the median TST for night 1 as well as the median for the average TST of subsequent nights (< median TST = "short sleeper"; ≥ median TST = "normal sleeper"). Separate cutoffs were used for night 1 and subsequent nights in our three study samples to take into account the first-night effect, the heterogeneity of the groups, and to address the issue of regression to the mean. Persistence of the individuals' first-night classification on subsequent nights was assessed through contingency tables. All data were analyzed using the Statistical Package for the Social Sciences (SPSS) Version 20 (IBM Corp., Armonk, NY).

RESULTS

Short-Term Sleep Stability

Sociodemographic and PSG characteristics of controls and insomniacs are shown in Table 1. Though they did not differ in age or sex distribution, the control group had a higher BMI. Controls slept significantly more over 3 nights on average compared to patients with insomnia and had significantly higher SE, whereas patients with insomnia had a much longer SOL. Of note, controls experienced more WASO by 8.1 min, which

was significantly associated with increased BMI in controls (Pearson r = 0.45, P < 0.001), but not patients with insomnia (r = −0.06, P = 0.52). Additionally, compared to controls, patients with insomnia had less stage 1 and slow wave sleep (SWS), but more stage 2 and REM (Table 1).

Intraclass correlation coefficients for TST, SOL, and WASO in controls and patients with insomnia across 3 consecutive nights are shown in Table 2. Single-measure TST was moderately stable in controls (ICC = 0.50) and patients with insomnia (ICC = 0.43), whereas average-measure TST was substantially stable (ICC = 0.75 and 0.69, respectively). Similarly, single-measure SOL was slightly stable in controls (ICC = 0.39) and moderately stable in patients with insomnia (ICC = 0.57), but demonstrated substantial stability when average measures were applied (ICC = 0.66 and 0.80, respectively). Single-measure WASO demonstrated moderate stability in controls (ICC = 0.59) and slight stability in patients with insomnia (ICC = 0.37), though average-measure WASO

Table 1—Sociodemographic and polysomnographic sleep characteristics of controls and patients with insomnia.

	Control (n = 151)	Patients with Insomnia (n = 150)
% Male	56.3	52.0
Age (y)	36.0 (1.2)	36.6 (1.1)
Body mass index (kg/m²)	25.5 (0.4)	24.1 (0.4)*
TST (min)	416.7 (3.0)	392.4 (3.8)***
SE (%)	86.8 (0.6)	81.7 (0.8)***
SOL (min)	17.0 (1.1)	49.2 (3.1)***
WASO (min)	46.5 (2.7)	38.4 (2.5)*
Stage 1 (%)	8.7 (0.4)	6.4 (0.3)***
Stage 2 (%)	60.2 (0.8)	65.0 (0.6)***
SWS (%)	12.5 (0.7)	6.8 (0.6)***
REM (%)	18.6 (0.5)	21.8 (0.4)***

Mean values reported with standard error of the mean (SEM) in parentheses. Polysomnographic variables represent 3 consecutive nights in controls and patients with insomnia. *P < 0.05; ***P < 0.001. REM, rapid eye movement sleep; SE, sleep efficiency; SOL, sleep onset latency; SWS, slow wave sleep; TST, total sleep time; WASO, wake after sleep onset.

Table 2—Intraclass correlations for sleep outcomes in controls and patients with insomnia.

	ICC Nights 1–3 Single Measure	ICC Nights 1–3 Average Measure	Between-Subjects Variance	Within-Subjects Variance
Controls (n = 151)				
TST	0.50 (0.40–0.59)*	0.75 (0.67–0.81)*	990.28	18.63
SOL	0.39 (0.29–0.49)*	0.66 (0.55–0.74)*	118.27	0.65
WASO	0.59 (0.50–0.67)*	0.81 (0.76–0.86)*	922.17	9.80
Patients with Insomnia (n = 150)				
TST	0.43 (0.32–0.53)*	0.69 (0.59–0.77)*	1540.54	170.94
SOL	0.57 (0.48–0.65)*	0.80 (0.73–0.85)*	1198.69	46.47
WASO	0.37 (0.27–0.47)*	0.64 (0.52–0.73)*	594.74	38.10

ICC represents 3 consecutive nights in controls and patients with insomnia, with 95% confidence interval in parentheses. *P < 0.001. ICC, intraclass correlation coefficient; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset.

was substantially to almost perfectly stable (ICC = 0.81 and 0.64, respectively). Of note, within-subjects variances were consistently lower than between-subjects variances for all sleep outcomes examined (Table 2). When examining sleep stages (stages 1, 2, SWS and REM sleep), single-measure ICCs were substantially stable in controls and slightly to moderately stable in insomniacs; stability was improved to almost perfect (controls) and moderate to substantial (patients with insomnia) when examining average measures (Table S1, supplemental material). Interestingly, in terms of SWS, a single night yielded almost perfect stability in both normal-sleeping controls and patients with insomnia (both ICC = 0.89).

Long-Term Sleep Stability

Sociodemographic and PSG characteristics of the longitudinally studied general population sample of 95 men are shown in Table 3. The average time between sleep studies was 2.6 y, and neither BMI nor AHI changed significantly over time (all P > 0.05). There were also no differences in TST, SE, SOL, WASO, or any of the sleep architecture variables in this group after 2.6 y, except that stage 2 was significantly reduced by 1.9%.

Table 4 shows ICC for the sleep duration variables TST, SOL, and WASO across the 2 nights. Single-measure TST and WASO showed moderate long-term stability (ICC = 0.50

and 0.44, respectively); however, there was a poor correlation in SOL between the two time points (ICC = 0.18). Average-measure SOL was slightly stable across several years (ICC = 0.30), whereas TST and WASO showed substantial stability (ICC = 0.67 and 0.61, respectively). Again, within-subjects variances were consistently lower than between-subjects variances for all sleep outcomes examined (Table 4). In terms of sleep stages, single-measure ICCs were moderately stable, and became substantially stable when assessing average-measures (Table S2, supplemental material). SWS stability, however, demonstrated only slight to moderate stability; this is likely due to the loss of SWS that occurred between the initial visit and follow-up several years later (Table 3).

Persistence of Night 1 Classification on Subsequent Nights

Median sleep durations for night 1 as well as the average of subsequent nights in the sleep laboratory are presented for controls, patients with insomnia, and the longitudinal cohort of men in Table 5. In controls, when nights 2 and 3 were averaged and their median used as a cutoff, 71.4% of "short" and "normal" sleepers on night 1 retained these classifications over 2 subsequent nights. Similarly, 74.7% of patients with insomnia remained in the same category over 3 nights in the laboratory. In the longitudinal sample, 72.6% of the group retained their "short" or "normal" classifications over 2.6 y.

DISCUSSION

We examined the short-term stability of sleep duration in 150 patients with insomnia and 151 normal-sleeping controls. We also assessed long-term sleep stability over several years in a longitudinal general population sample of 95 middle-aged men. Overall, although the variables TST, SOL, and WASO were only slightly to moderately stable according to single-measures intraclass correlations, the average of 3 consecutive nights in the laboratory, or 2 nights separated by several years, produced moderately to very strongly stable (ICC ≥ 0.60) within-subject assessments of sleep duration. Furthermore, the majority of controls (71.4%) and patients with insomnia (74.7%) who were classified objectively as "short" or "normal" sleepers during their first night in the laboratory retained these classifications over consecutive nights. Similarly, 72.6% of the longitudinal sample remained "short" or "normal" sleepers across several years. In summary, our findings suggest that although a single night in the laboratory may not yield the most reliable, reproducible measures, 3 consecutive nights, or two single-night recordings separated by several years, are sufficient. Importantly, however, a single night may be useful for

Table 3—Sociodemographic characteristics of longitudinal sample (n = 95 men).

	Night 1	Night 2
Age (y)	51.1 (1.1)	53.7 (1.1)**
Body mass index (kg/m^2)	30.2 (0.6)	30.6 (0.7)
Apnea-hypopnea index (events/h)	9.7 (2.4)	5.9 (1.0)
TST (min)	356.4 (7.1)	354.7 (6.6)
SE (%)	74.2 (1.5)	73.9 (1.4)
SOL (min)	24.4 (3.1)	24.8 (2.9)
WASO (min)	98.5 (6.1)	99.0 (5.7)
Stage 1 (%)	15.7 (0.8)	16.7 (0.8)
Stage 2 (%)	68.1 (0.7)	66.2 (0.8)*
SWS (%)	0.62 (0.1)	0.45 (0.1)
REM (%)	15.6 (0.5)	16.6 (0.6)

Mean values reported with standard error of the mean (SEM) in parentheses. *P < 0.01; **P < 0.001. REM, rapid eye movement sleep; SE, sleep efficiency; SOL, sleep onset latency; SWS, slow wave sleep; TST, total sleep time; WASO, wake after sleep onset.

Table 4—Intraclass correlations for sleep outcomes in longitudinal sample.

	ICC Nights 1, 2 Single Measure	ICC Nights 1, 2 Average Measure	Between-Subjects Variance	Within-Subjects Variance
Male Cohort (n = 95)				
TST	0.50 (0.34–0.64)***	0.67 (0.51–0.78)***	1493.95	13.89
SOL	0.18 (0.00–0.37)*	0.30 (0.00–0.54)*	99.39	4.52
WASO	0.44 (0.26–0.59)***	0.61 (0.41–0.74)***	958.82	12.23

ICC represents 2 nights with 2.6 y between visits, with 95% confidence interval in parentheses. *P < 0.05; **P < 0.01; ***P < 0.001. ICC, intraclass correlation coefficient; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset.

reliably classifying one's sleep duration over the short-term and long-term.

Few studies employing visually scored sleep have examined stability across consecutive nights. In Merica and Gaillard's 1985 analysis of 147 healthy adults aged 16 to 71 y, only stage 4 sleep produced reliable results over several recordings.[2] Extending these findings in 50 older adults (ages 54 to 82 y), Larsen and colleagues[8] reported that, compared to spectral power scoring techniques, visually scored SWS (stages 3 and 4) had a significantly lower correlation over 2 consecutive nights. Using a Spearman-Brown formula to approximate projected SWS stability over more nights, Larsen et al. estimated that it would take 6 nights of visually scored SWS to attain the same reliability as 2 nights of computer-analyzed sleep.[8]

Similar to the current study, Wohlgemuth and colleagues[9] studied controls and patients with insomnia for 3 consecutive nights in the laboratory with PSG. Using statistical projections, the authors concluded that 1 w of visually scored PSG recordings is required to achieve adequate stability in the variables TST, SOL, WASO, SE, and time in bed (TIB) in normal sleepers; for their counterparts with insomnia, however, more than 2 w were needed to achieve SOL stability, and at least 3 w were necessary for stable WASO.[9] We should note, however, that this study allowed for habitual sleep periods, as evidenced by variability in TIB between participants. Without standardizing this variable across all participants, the assessment of short-term stability in sleep continuity may be compromised. Furthermore, more stringent criteria were placed in terms of what was considered "stable." Wohlgemuth et al.[9] calculated G coefficients, which can be interpreted analogously to ICC, and used a threshold of 0.80 to define "adequate stability." In line with others,[7,36] we had elected an ICC threshold of 0.60 as demonstrating "reliable" short- or long-term stability. Interestingly, when comparing the stability coefficients of Wohlgemuth et al.'s control participants and patients with insomnia to our own, many results—both for a single night and 3 nights—are quite comparable. Finally, the average age of participants in Wohlgemuth et al.'s study was close to 70 y. A large meta-analysis by Ohayon et al.[37] has demonstrated an exponential decrease in deep sleep and increase in WASO, as well as linear increase in stage 1 and decrease in REM sleep with aging. As such, it is possible that instability of sleep duration over consecutive nights could be confounded by fragmented sleep due to older age.

Israel and colleagues[7] recently examined short-term stability in 54 adults with insomnia and 22 normal-sleeping controls, employing both visual scoring and QEEG across 3 consecutive nights. Power spectral analysis yielded significantly higher ICCs than visually scored sleep, with the exception of SWS.[7] Visually scored TST, SOL, and WASO only ranged from ICC = 0.34–0.56 ("slight" to "moderate" stability) compared to QEEG-scored variables (0.55–0.88) in both groups. Like Wohlgemuth et al.[9], it is important to note that this study also did not standardize TIB. However, a 3-night, home-based, *ad libitum* PSG study by Coates and colleagues[38] in 12 patients with insomnia and 12 normal sleepers (age range = 20–60 y) achieved substantial stability of SOL and WASO, respectively, in both patients with insomnia (0.70 and 0.67) and controls (0.58 and 0.72).[34] Although these findings are in agreement with what

Table 5—Stability of sleep classification in controls, patients with insomnia, and the longitudinal sample.

Short-Term	Night 1	Nights 2, 3
Controls (n = 151)		
"Short" sleep duration (min)	< 425.0	< 430.3
"Normal" sleep duration (min)	≥ 425.0	≥ 430.3
Persistence of night 1 classification		71.4%
Patients with insomnia (n = 150)		
"Short" sleep duration (min.)	< 391.0	< 412.5
"Normal" sleep duration (min)	≥ 391.0	≥ 412.5
Persistence of night 1 classification		74.7%
Long-Term	**Night 1**	**Night 2**
Male cohort (n = 95)		
"Short" sleep duration (min)	< 367.9	< 365.0
"Normal" sleep duration (min)	≥ 367.9	≥ 365.0
Persistence of night 1 classification		72.6%

"Short" and "normal" sleep classification as determined by median sleep duration in controls, patients with insomnia, and a longitudinal general population sample.

we observed, the investigators omitted the first night from analysis. As such, extending these findings to a typical single-night research setting with fixed TIB may be troublesome.

Overall, the variable TST appeared to fare best in terms of both short-term and long-term stability, particularly when examining persistence of sleep duration classification; in turn, this may be consistent with the view that sleep duration is, to some degree, driven by biological factors.[39,40] Studies of monozygotic and dizygotic twins, for example, suggest that SE and WASO are highly heritable ($h^2 > 0.97$),[41] and that a significant portion of variance in stages 2, 4, and delta sleep appears to be genetically determined.[42] Although the difference in median TST between the first night and subsequent nights was larger in patients with insomnia than controls (21.5 min versus 5.3 min; Table 5), which is consistent with previous work,[12] both groups still demonstrated substantial stability (ICC = 0.69 and 0.75, respectively) in sleep duration when averaged across 3 nights. Interestingly, our findings are also in agreement with a study employing wrist actigraphy, which demonstrated little internight variability in TST and WASO across 7 consecutive nights.[43] The slightly greater stability of SOL compared to TST in volunteer research patients with insomnia may reflect the fact that a long self-reported sleep latency (≥ 45 min) was the primary quantitative selection criterion.[23] However, the relatively low long-term stability and wide confidence interval of SOL across several years in the longitudinal cohort suggests that this variable may be more influenced by an individual's "state" (i.e. how they felt that day, sleeping conditions during that particular night, and other fluctuating circumstances). Whereas such state-dependent variables (such as familiarity with the room, smells, general procedures, and the PSG technologist) are relatively controlled across consecutive nights, the increased time between visits in this longitudinal study may explain why certain variables, such as latency to fall asleep, are vulnerable to instability and the first-night effect. Therefore,

Agnew et al.'s[1] description of decreased SE, increased stage 1 sleep, and decreased REM as a result of the first-night effect would not be surprising to observe at both time points in this longitudinal sample. It is of interest to note that regardless of ICC, variance is consistently lower within- than between-subjects, suggesting relatively good overall stability of TST, SOL, and WASO across nights; this is consistent with previously published large studies suggesting that interindividual differences in sleep stability are fairly stable and robust.[4,44]

Another major finding in our study is that a single-night sleep recording is valid and clinically useful in classifying participants in terms of "short" and "normal" sleep duration (Table 5). In turn, this finding strengthens the validity of our approach in our previous reports that insomnia with short sleep duration, based on a single-night recording, is consistently associated with cardiometabolic morbidity and mortality.[17–21]

There are, however, several fundamental differences between the current study and previous studies on the topic. Firstly, our large sample size is a critical factor in our statistical approach. The intraclass correlation is a test-retest reliability method designed to quantify the degree to which related individuals (e.g., patients with insomnia) resemble each other in regard to a quantitative trait (e.g., TST).[45] ICC is calculated by dividing the between-subject variance of this trait by the sum of between-subject trait variance plus within-subject pooled variance. The value of the ICC is positively associated with number of observations (i.e., number of nights studied in the laboratory), as more trials reduce within-subject variance. However, ICC is negatively associated with sample size due to increasing between-subject variance.[46,47] The fact that our large sample size (95, 150, and 151 participants per group) continued to yield strong ICC values compared to similar studies with 50 or fewer participants per group[1,3,4,7–9,48,49] suggests that well-defined populations not only sleep similarly across nights, but also to one another. This conclusion is also in agreement with our data suggesting that, relative to others, one's sleep classification ("short" versus "normal") is reasonably persistent across nights (Table 5). In making these comparisons, it is especially important to note that our control and insomnia groups were of similar sample size and sociodemographic characteristics.

Additionally, focusing on visually scored sleep, as opposed to more sophisticated spectral analytic methods, allows us to extend our findings to the clinical setting and/or large research samples, where visual scoring is the norm. Our design is also unique in that it mimics typical multinight sleep study protocols, which aim to habituate participants to the laboratory setting. In many cases, the first night of sleep is omitted from analysis due to the first-night effect. Comparing both single- and average-measures stability allows us to observe the effect of this phenomenon. As multinight PSG studies are expensive, time-consuming, and often not even considered for use in large epidemiological studies, it is important to understand if and how one night differs from subsequent nights. Our multinight approach also addresses hypotheses from previous research which estimate that it takes 1 w or more to achieve adequate stability in variables related to sleep duration in adults.[8,9] These conclusions were not based on studies employing multiple nights of PSG, but were rather approximations based on statistical projections.

Finally, our comparison of sleep stability between patients with insomnia and controls does not have the confounding variables of older age or obesity that have been shown to affect day-to-day variation in sleep duration[37,50] or quality.[10] Controls and patients with insomnia did not differ in terms of sex distribution or mean age, nor included any subjects older than 70 y, and body mass index did not exceed the "obese" threshold in either group. Although our longitudinal general population sample was relatively older (mean age 51.1 ± 1.1 y), no differences in sleep duration over time could be attributed to changes in BMI or AHI (Table 4).

There are several limitations to the current study that may affect its generalizability to certain research samples. First, our control and insomniac groups included mostly non-Hispanic Caucasians. Although a number of studies have identified racial/ethnic disparities in sleep architecture,[51–53] no studies to date have explored the short- and/or long-term stability of sleep duration across ethnic groups. Additionally, although our longitudinal sample of 95 participants was derived from a representative general population sample, it consisted entirely of men because of the primary focus of our previous work. As studies examining the long-term stability of PSG-measured sleep are limited, future work should explore sex effects on stability of sleep over time, particularly in a middle-aged to older population such as ours. Examining the role of physiologic changes, such as gain or loss in BMI over time, would also be an interesting extension to this study. Finally, regression to the mean is an issue to keep in mind when interpreting the persistence of "short" and "normal" sleep duration, particularly as these groups were defined simply by a median split. Redefining the medians from night 1 to subsequent nights, however, may protect against this effect to some degree.

Despite these limitations, our multinight, in-laboratory design, which was not confounded by older age, obesity, or *ad libitum* sleep schedules, demonstrated relatively high short-term and long-term stability of visually scored sleep across 3 consecutive nights, or 2 nights assessed longitudinally. We conclude that a single night in the laboratory may provide reliable measures, particularly in the context of classifying one's sleep duration both in the short term and long term.

ACKNOWLEDGMENTS

DISCLOSURE STATEMENT

REFERENCES

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

2. Merica H, Gaillard JM. Statistical description and evaluation of the interrelationships of standard sleep variables for normal subjects. Sleep 1985;8:261–73.

3. Tan X, Campbell IG, Palagini L, Feinberg I. High internight reliability of computer-measured NREM delta, sigma, and beta: biological implications. Biol Psychiatry 2000;48:1010–9.

4. Tucker AM, Dinges DF, van Dongen HP. Trait interindividual differences in the sleep physiology of healthy young adults. J Sleep Res 2007;16:170–80.

5. Geiger A, Huber R, Kurth S, Ringli M, Jenni OG, Achermann P. The sleep EEG as a marker of intellectual ability in school age children. Sleep 2011;4:181–9.

6. Tarokh L, Carskadon MA, Achermann P. Trait-like characteristics of the sleep EEG across adolescent development. J Neurosci 2011;31:6371–8.

7. Israel B, Buysse DJ, Krafty RT, Begley A, Miewald J, Hall M. Short-term stability of sleep and heart rate variability in good sleepers and patients with insomnia: for some measures, one night is enough. Sleep 2012;35:1285–91.

8. Larsen LH, Moe KE, Vitiello MV, Prinz PN. A note on the night-to-night stability of stages 3 + 4 sleep in healthy older adults: a comparison of visual and spectral evaluations of stages 3 + 4 sleep. Sleep 1995;18:7–10.

9. Wohlgemuth WK, Edinger JD, Fins AI, Sullivan RJ. How many nights are enough? The short-term stability of sleep parameters in elderly insomniacs and normal sleepers. Psychophysiology 1999;36:233–44.

10. Zheng H, Sowers M, Buysse DJ, et al. Sources of variability in epidemiological studies of sleep using repeated nights of in-home polysomnography: SWAN Sleep Study. J Clin Sleep Med 2012;8:87–96.

11. Lorenzo JL, Barbanoj MJ. Variability of sleep parameters across multiple laboratory sessions in healthy young subjects: the "very first night effect." Psychophysiology 2002;39:409–13.

12. Newell J, Mairesse O, Verbanck P, Neu D. Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples. Psychiatry Res 2012;200:795–801.

13. Kay DB, Dzierzewski JM, Rowe M, McCrae CS. Greater night-to-night variability in sleep discrepancy among older adults with a sleep complaint compared to noncomplaining older adults. Behav Sleep Med 2013;11:76–90.

14. Vallières A, Ivers H, Beaulieu-Bonneau S, Morin CM. Predictability of sleep in patients with insomnia. Sleep 2011;34:609–17.

15. 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.

16. Edinger JD, Marsh GR, McCall WV, Erwin CW, Lininger AW. Sleep variability across consecutive nights of home monitoring in older mixed DIMS patients. Sleep 1991;14:13–7.

17. Vgontzas AN, Liao D, Pejovic S, Calhoun SL, Karataraki M, Bixler EO. Insomnia with objective short sleep duration is associated with type 2 diabetes: a population-based study. Diabetes Care 2009;32:1980–5.

18. Vgontzas AN, Liao D, Bixler EO, Chrousos GP, Vela-Bueno A. Insomnia with objective short sleep duration is associated with high risk for hypertension. Sleep 2009;32:491–7.

19. Vgontzas AN, Liao D, Pejovic S, Calhoun SL, et al. Insomnia with short sleep duration and mortality: the Penn State cohort. Sleep 2010;33:1159–64.

20. Fernandez-Mendoza J, Vgontzas AN, Liao D, et al. Insomnia with objective short sleep duration and incident hypertension: the Penn State cohort. Hypertension 2012;20:929–35.

21. Vgontzas AN, Fernandez-Mendoza J, Liao D, Bixler EO. Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. Sleep Med Rev 2013;17:241–54.

22. Fernandez-Mendoza J, Calhoun SL, Bixler EO, et al. Insomnia with objective short sleep duration is associated with deficits in neuropsychological performance: a general population study. Sleep 2010;33:459–65.

23. Kales A, Manfredi R, Vgontzas AN, Bixler EO, Vela-Bueno A, Fee EC. Rebound insomnia after only brief and intermittent use of rapidly eliminated benzodiazepines. Clin Pharmacol Ther 1991;49:468–76.

24. Vgontzas AN, Bixler EO, Kales A, Manfredi R, Tyson K. Validity and clinical utility of sleep laboratory criteria for insomnia. Int J Neurosci 1994;77:11–21.

25. Vgontzas AN, Kales A, Bixler EO, Manfredi RL, Vela-Bueno A. Usefulness of polysomnographic studies in the differential diagnosis of insomnia. Int J Neurosci 1995;82:47–60.

26. Vgontzas AN, Zoumakis E, Bixler EO, et al. Adverse effects of modest sleep restriction on sleepiness, performance, and inflammatory cytokines. J Clin Endocrinol Metab 2004;89:2119–26.

27. Pejovic S, Basta M, Vgontzas AN, et al. Effects of recovery sleep after one work week of mild sleep restriction on interleukin-6 and cortisol secretion and daytime sleepiness and performance. Am J Physiol Endocrinol Metab 2013;305:E890–6.

28. Pejovic S, Vgontzas AN, Basta M, et al. Leptin and hunger levels in young healthy adults after one night of sleep loss. J Sleep Res 2010;19:552–8.

29. Vgontzas AN, Pejovic S, Zoumakis E, et al. Hypothalamic-pituitary-adrenal axis activity in obese men with and without sleep apnea: effects of continuous positive airway pressure. J Clin Endocrinol Metab 2007;92:4199–207.

30. Vgontzas AN, Zoumakis E, Bixler EO, et al. Selective effects of CPAP on sleep apnoea-associated manifestations. Eur J Clin Invest 2008;38:585–95.

31. Kritikou I, Basta M, Tappouni R, et al. Sleep apnoea and visceral adiposity in middle-aged male and female subjects. Eur Respir J 2013;41:601–9.

32. Kritikou I, Basta M, Vgontzas AN, et al. Sleep apnoea, sleepiness, inflammation and insulin resistance in middle-aged males and females. Eur Respir J 2014;43:145–55.

33. Bixler EO, Vgontzas AN, Ten Have T, Tyson K, Kales A. Effects of age on sleep apnea in men: I. Prevalence and severity. Am J Respir Crit Care Med 1998;157:144–8.

34. Rechtschaffen A, Kales A. Manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. NIH Publication 204. Washington, DC: US Government Printing Office, Department of Health Education and Welfare, 1968.

35. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86:420–8.

36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74.

37. Ohayon MM, Carskadon MA, Guilleminault C, Vitiello MV. Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. Sleep 2004;27:1255–73.

38. Coates TJ, Killen JD, George J, Marchini E, Silverman S, Thorsen C. Estimating sleep parameters: a multitrait-multimethod analysis. J Consult Clin Psych 1982;50:345–52.

39. Klei L, Reitz P, Miller M, et al. Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 Hutterites. Chronobiol Int 2005;22:1041–54.

40. Barclay NL, Gregory AM. Quantitative genetic research on sleep: a review of normal sleep, sleep disturbances and associated emotional, behavioural, and health-related difficulties. Sleep Med Rev 2013;17:29–40.

41. Kuna ST, Maislin G, Pack FM, et al. Heritability of performance deficit accumulation during acute sleep deprivation in twins. Sleep 2012;35:1223–33.

42. Linkowski P. EEG sleep patterns in twins. J Sleep Res 1999;8:11–3.

43. Otte JL, Payne JK, Carpenter JS. Nighttime variability in wrist actigraphy. J Nurs Meas 2011;19:105–14.

44. Van Dongen HP, Vitellaro KM, Dinges DF. Individual differences in adult human sleep and wakefulness: leitmotif for a research agenda. Sleep 2005;28:479–96.

45. Koch G. Intraclass correlation coefficient. In: Kotz S, Johnson NL, eds. Encyclopedia of statistical sciences. New York: John Wiley & Sons, 1982:213–7.

46. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. Stat Med 2002;21:1131–5.

47. Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. Stat Methods Med Res 2004;13:251–71.

48. Tan X, Campbell IG, Feinberg I. Internight reliability and benchmark values for computer analyses of non-rapid eye movement (NREM) and REM EEG in normal young adult and elderly subjects. Clin Neurophysiol 2001;112:1540–52.

49. Buckelmüller J, Landolt HP, Stassen HH, Achermann P. Trait-like individual differences in the human sleep electroencephalogram. Neuroscience 2006;138:351–6.

50. Vgontzas AN, Bixler EO, Tan TL, Kantner D, Martin LF, Kales A. Obesity without sleep apnea is associated with daytime sleepiness. Arch Intern Med 1998;158:1333–7.

51. Rao U, Poland RE, Lutchmansingh P, Ott GE, McCracken JT, Lin KM. Relationship between ethnicity and sleep patterns in normal controls: implications for psychopathology and treatment. J Psychiatr Res 1999;33:419–26.

52. Profant J, Ancoli-Israel S, Dimsdale JE. Are there ethnic differences in sleep architecture? Am J Hum Biol 2002;14:321–6.

53. Redline S, Kirchner HL, Quan SF, Gottlieb DJ, Kapur V, Newman A. The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. Arch Intern Med 2004;164:406–18.

Table S1—Intraclass correlations for sleep stages in controls and patients with insomnia.

	ICC Nights 1–3 Single Measure	ICC Nights 1–3 Average Measure	Between-Subjects Variance	Within-Subjects Variance
Controls (n = 151)				
Stage 1	0.74 (0.67–0.79)*	0.89 (0.86–0.92)*	22.93	0.12
Stage 2	0.81 (0.76–0.85)*	0.93 (0.90–0.95)*	86.90	0.05
SWS	0.89 (0.86–0.92)*	0.96 (0.95–0.97)*	80.43	0.05
REM	0.70 (0.63–0.76)*	0.87 (0.83–0.91)*	34.90	0.68
Patients with Insomnia (n = 150)				
Stage 1	0.34 (0.24–0.45)*	0.61 (0.49–0.71)*	6.10	0.19
Stage 2	0.62 (0.51–0.70)*	0.83 (0.76–0.88)*	41.53	3.23
SWS	0.89 (0.86–0.92)*	0.96 (0.95–0.97)*	50.02	0.12
REM	0.47 (0.33–0.60)*	0.73 (0.60–0.820)*	17.63	4.00

ICC represents 3 consecutive nights in controls and patients with insomnia, with 95% confidence interval in parentheses. *$P < 0.001$. ICC, intraclass correlation coefficient; REM, rapid eye movement sleep; SWS, slow wave sleep.

Table S2—Intraclass correlations for sleep stages in longitudinal sample.

	ICC Nights 1, 2 Single Measure	ICC Nights 1, 2 Average Measure	Between-Subjects Variance	Within-Subjects Variance
Male Cohort (n = 95)				
Stage 1	0.55 (0.39–0.67)**	0.71 (0.56–0.81)**	32.56	0.26
Stage 2	0.55 (0.39–0.68)**	0.71 (0.56–0.81)**	29.58	1.54
SWS	0.29 (0.09–0.46)*	0.45 (0.17–0.63)*	0.51	0.001
REM	0.51 (0.34–0.64)**	0.67 (0.51–0.78)**	17.22	0.34

ICC represents two nights with 2.6 years between visits, with 95% CI in parentheses. *$P < 0.01$; **$P < 0.001$. ICC, intraclass correlation coefficient; REM, rapid eye movement sleep; SWS, slow wave sleep.