

## REPRODUCIBILITY OF ACTIGRAPHY FOR SLEEP IN A US HISPANIC/LATINO POPULATION

# Reproducibility of a Standardized Actigraphy Scoring Algorithm for Sleep in a US Hispanic/Latino Population

Sanjay R. Patel, MD, MS<sup>1,2</sup>; Jia Weng, PhD<sup>1</sup>; Michael Rueschman, MPH<sup>1</sup>; Katherine A. Dudley, MD<sup>1</sup>; Jose S. Loredó, MD, MPH<sup>3</sup>; Yasmin Mossavar-Rahmani, PhD<sup>4</sup>; Maricelle Ramirez, BA<sup>2</sup>; Alberto R. Ramos, MD<sup>5</sup>; Kathryn Reid, PhD<sup>6</sup>; Ashley N. Seiger, BS<sup>1</sup>; Daniela Sotres-Alvarez, DrPH<sup>7</sup>; Phyllis C. Zee, MD, PhD<sup>6</sup>; Rui Wang, PhD<sup>1</sup>

<sup>1</sup>Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA; <sup>2</sup>Beth Israel Deaconess Medical Center, Boston, MA; <sup>3</sup>University of California San Diego, San Diego, CA; <sup>4</sup>Albert Einstein College of Medicine, Bronx, NY; <sup>5</sup>University of Miami, Miami, FL; <sup>6</sup>Northwestern University, Chicago, IL; <sup>7</sup>Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC

**Study Objectives:** While actigraphy is considered objective, the process of setting rest intervals to calculate sleep variables is subjective. We sought to evaluate the reproducibility of actigraphy-derived measures of sleep using a standardized algorithm for setting rest intervals.

**Design:** Observational study.

**Setting:** Community-based.

**Participants:** A random sample of 50 adults aged 18–64 years free of severe sleep apnea participating in the Sueño sleep ancillary study to the Hispanic Community Health Study/Study of Latinos.

**Interventions:** N/A.

**Measurements and Results:** Participants underwent 7 days of continuous wrist actigraphy and completed daily sleep diaries. Studies were scored twice by each of two scorers. Rest intervals were set using a standardized hierarchical approach based on event marker, diary, light, and activity data. Sleep/wake status was then determined for each 30-sec epoch using a validated algorithm, and this was used to generate 11 variables: mean nightly sleep duration, nap duration, 24-h sleep duration, sleep latency, sleep maintenance efficiency, sleep fragmentation index, sleep onset time, sleep offset time, sleep midpoint time, standard deviation of sleep duration, and standard deviation of sleep midpoint. Intra-scorer intraclass correlation coefficients (ICCs) were high, ranging from 0.911 to 0.995 across all 11 variables. Similarly, inter-scorer ICCs were high, also ranging from 0.911 to 0.995, and mean inter-scorer differences were small. Bland-Altman plots did not reveal any systematic disagreement in scoring.

**Conclusions:** With use of a standardized algorithm to set rest intervals, scoring of actigraphy for the purpose of generating a wide array of sleep variables is highly reproducible.

**Keywords:** actigraphy, reproducibility, Hispanic, sleep

**Citation:** Patel SR, Weng J, Rueschman M, Dudley KA, Loredó JS, Mossavar-Rahmani Y, Ramirez M, Ramos AR, Reid K, Seiger AN, Sotres-Alvarez D, Zee PC, Wang R. Reproducibility of a standardized actigraphy scoring algorithm for sleep in a US Hispanic/Latino population. *SLEEP* 2015;38(9):1497–1503.

## INTRODUCTION

Epidemiological studies have identified strong relationships between self-reported measures of poor sleep, both in terms of quantity and quality, and a host of adverse health consequences including obesity, hypertension, infections, diabetes, heart disease, and mortality.<sup>1–6</sup> However, these studies have been criticized for their reliance on self-report for measuring sleep given that the correlation between self-report and objective measurement is poor.<sup>7</sup> While polysomnography is the gold standard for assessing sleep, it is expensive and impractical for measuring sleep over multiple nights in large populations. As such, wrist actigraphy is increasingly favored as a noninvasive method for assessing sleep-wake activity.

Wrist actigraphy measures motion of the wrist; low levels of activity are inferred to indicate sleep. Validated actigraphy

algorithms have been developed for identifying sleep on an epoch-by-epoch basis during time spent in bed.<sup>8–10</sup> However, application of these algorithms across the 24-hour day may result in overscoring periods of quiet wakefulness as sleep. In order to limit this overestimation, sleep diaries have been traditionally used to identify rest intervals when the subject is trying to sleep, and the algorithm for scoring sleep is only applied to these selected intervals.<sup>11</sup> In actual practice, sleep diaries are often inaccurate or incomplete. No standardized approach has been developed to handle this problem and it is typically left to “expert” scorers to use their best judgment.<sup>12</sup> While the magnitude of error introduced as a result of this subjective process of setting intervals is unclear, the lack of published scoring rules makes it difficult to train technicians to perform this task or develop expertise. Furthermore, current actigraph devices have the technology to allow subjects to press an event marker in real time and/or collect light information and newer devices have the added benefit of off-wrist detection. This additional information can potentially be used to improve the accuracy of defining rest intervals. However, no systematic method to combine information from multiple inputs has been published to date. In addition, few studies have assessed the reproducibility of actigraphy-derived sleep measures across scorers.<sup>12</sup>

Submitted for publication November, 2014

Submitted in final revised form February, 2015

Accepted for publication March, 2015

Address correspondence to: Sanjay R. Patel, MD, MS, Brigham and Women's Hospital, 221 Longwood Avenue, Suite 225-C, Boston, MA 02115; Tel: (857) 307-0347; Fax: (617) 278-6946; Email: spatel@partners.org

The objective of this study was to evaluate, both within and between scorers, the reproducibility of a wide range of sleep measures obtained with the use of an explicitly defined method of setting rest intervals combining information across a number of inputs using actigraphy data. The study was conducted as part of the Sueño sleep ancillary study evaluating the impact of sleep on health outcomes in a cohort of middle-aged US Hispanics/Latinos recruited from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL).

## METHODS

### Study Cohort

The HCHS/SOL is a community-based cohort study of 16,415 self-identified Hispanic/Latino adults 18–74 years old recruited from randomly selected households at 4 U.S. field centers (Bronx, NY; Chicago, IL; Miami, FL; and San Diego, CA) between 2008 and 2011. The multi-stage sampling design and cohort selection procedures have been previously described.<sup>13,14</sup> As part of the HCHS/SOL baseline exam, participants underwent unattended home sleep apnea monitoring using the ARES Unicorder 5.2 (B-Alert, Carlsbad, CA).<sup>15</sup> Use of this device and the process for centralized scoring in HCHS/SOL have been previously described.<sup>16</sup> Respiratory events were defined as  $\geq 50\%$  reduction in airflow lasting  $\geq 10$  sec associated with  $\geq 3\%$  oxyhemoglobin desaturation. The apnea-hypopnea index (AHI) was calculated as the number of respiratory events per estimated sleep hour.

The Sueño sleep ancillary study recruited 2,252 individuals from December 2010 to December 2013 who were within 30 months of their baseline HCHS/SOL examination, aged 18–64 years, and without narcolepsy, severe obstructive sleep apnea (AHI  $\geq 50$ /h), or using nocturnal positive airway pressure therapy in order to identify both sociodemographic predictors of sleep patterns and the health consequences of those patterns in a Hispanic/Latino population. Individuals with insomnia, restless legs syndrome, shiftwork, or other sleep disorders were not excluded from participation. This study was approved by the institutional review boards at all 4 sites, and all participants provided written informed consent.

### Sueño Study Protocol

As part of the ancillary study, all Sueño participants completed a questionnaire in either English or Spanish based on language preference. The questionnaire collected information on Hispanic/Latino background, level of education, and employment status. Sleep-related symptoms were assessed using items from the Sleep Heart Health Study Sleep Habits Questionnaire,<sup>17</sup> the Epworth Sleepiness Scale,<sup>18</sup> and the Insomnia Severity Index.<sup>19</sup> Depressive symptoms were evaluated using the 10-item version of the Center for Epidemiological Studies–Depression (CES-D10) questionnaire.<sup>20,21</sup> Self-reported weekday and weekend sleep durations were computed as the difference between habitual wake time and bedtime. Self-reported average sleep duration was computed as the weighted average of weekday and weekend sleep durations. Height and weight were measured, from which body mass index was calculated.

All participants were asked to wear an Actiwatch Spectrum (Philips Respironics, Murrysville, PA) device on their

non-dominant wrist for 7 days and not remove the device for the duration of the study. Participants were instructed in the use of the event marker on the actigraph and asked to press the marker button upon getting into or out of bed. They also completed a sleep diary upon awakening each morning in which they noted time to bed, time out of bed, and any naps taken the previous day.

### Data Processing

Actigraphs were programmed to collect activity and light data in 30-sec epochs. Upon return of the actigraph and sleep diary, data were transmitted electronically to the central reading center at Brigham and Women's Hospital for scoring. Each study was evaluated by a scorer who used a standardized approach as described below to set rest intervals (periods when the subject was trying to sleep) based on 4 inputs: event markers, sleep diary, white light intensity, and activity in order of importance, respectively. For each day, a main rest interval was identified as the primary period for sleep based on diaries and usual sleep habits across the week. All other rest intervals for that day were considered naps. For each rest interval, the most likely time in bed was identified for each input (event marker, sleep diary, light intensity, and activity signal) in isolation. For light and activity, time in bed was identified by a sudden large drop in signal intensity. Light intensity needed to fall below 1 lux (or the lowest intensity observed for that day) for  $\geq 5$  epochs, and activity level needed to drop to zero for  $\geq 5$  epochs. If an input was not available (e.g., incomplete diary), the timing for that input was left as missing. Then concordance within a 15-min interval was assessed across inputs. If  $\geq 2$  inputs were within 15 min, the highest ranked input among those in concordance was used to define time in bed. If 2 pairs of inputs were in concordance for different times (e.g., both event marker and light suggested time in bed was at 22:00 while diary and activity both suggested 23:30), the highest ranked input (i.e., event marker) was used to define time in bed. If no pair of inputs was in concordance within 15 min, the process was repeated assessing concordance within 30 min. If no pair of inputs was in agreement across 30 min, then all inputs were deemed unreliable and activity was used to define time in bed. The same process was used for time out of bed. Of note, a gradual increase in light intensity in the early morning hours was inferred to represent sunrise and was not used for scoring light signals.

A similar approach was used to score naps. However, rest intervals for naps were only created if either a diary or event marker indicated an attempt by the participant to sleep. Otherwise, periods of low activity and low light intensity were interpreted as quiet wakefulness (e.g., watching a movie in relative darkness).

A day of recording was defined from noon to noon. Each day of recording was evaluated for quality. Any day with  $> 4$  h of missing data or  $> 2$  min of missing data during sleep in a main rest interval was considered invalid. Data could be missing due to off-wrist detection or a technical failure of the device. In the entire Sueño study, 208 out of 15,719 days (1.3%) were discarded due to missing data. Only studies with  $\geq 5$  valid days were considered adequate for analysis.

Once rest intervals were defined, the Actiware 5.59 algorithm using 5 immobile minutes to define sleep onset, 0

immobile minutes for sleep offset, and a wake threshold activity count of 40 was applied to generate sleep/wake status for each epoch. This algorithm has been previously validated on an epoch-by-epoch basis against polysomnography.<sup>8,22,23</sup>

### Outcomes of Interest

Actigraphic variables of interest were: mean nightly sleep duration, mean napping duration, mean 24-h sleep duration, mean sleep latency, mean sleep maintenance efficiency, mean sleep fragmentation index, mean sleep onset time, mean sleep offset time, mean sleep midpoint time, standard deviation (SD) of sleep duration, and SD of sleep midpoint. Mean nightly sleep duration was calculated as the total amount of time (in minutes) scored as sleep within each main rest interval averaged over the number of valid days of recording. Of note, for shiftworkers or others with unusual sleep habits, nightly sleep duration may actually occur during daytime hours. Mean napping duration was calculated as the total amount of time (in minutes) scored as sleep within all nap intervals each day averaged over the number of valid days of recording. Mean 24-h sleep duration was defined as the sum of nightly sleep duration and napping duration. Mean sleep latency is the total amount of time (in min) from the beginning of each rest interval until the first epoch scored as sleep averaged over all main rest intervals. Mean sleep onset time is the clock time (HH:MM) of the first epoch scored as sleep in each main rest interval averaged across all main rest intervals containing sleep. Mean sleep offset time is the clock time (HH:MM) of the last epoch scored as sleep in each main rest interval averaged across all main rest intervals containing sleep. Sleep midpoint was calculated as the midpoint between sleep onset and sleep offset. Mean sleep maintenance efficiency was calculated as the proportion of time from sleep onset to sleep offset in each main rest interval that was scored as sleep averaged across all valid days of recording expressed as a percentage. The sleep fragmentation index for each main rest interval is also expressed as a percentage and calculated as the sum of the proportion of all epochs from sleep onset to sleep offset that were mobile (i.e., the activity count was  $\geq 2$ ) and the proportion of all immobile bouts from sleep onset to sleep offset  $\leq 1$  min in duration (i.e., consecutive epochs where the activity count was  $< 2$  that were only  $\leq 2$  epochs in length).<sup>24</sup> The mean sleep fragmentation index averages the value for each main rest interval across all main rest intervals containing sleep. SDs of sleep duration and sleep midpoint were calculated as the SD of nightly sleep duration and sleep midpoint, respectively, over all main rest intervals in the recording and expressed in minutes.

### Reproducibility Subsample

From the first 790 valid studies completed, a random sample of 50 studies was selected for the reproducibility assessment. For these studies, 4 copies of both the actigraphy data and sleep diary were generated and labeled with new IDs. These studies were scored twice by each of 2 scorers blinded to the identity of the study. The 2 evaluations by each scorer occurred a median of 15 weeks apart (range 2–78 weeks). The studies were all then processed using the standardized approach described above to generate the 11 actigraphic variables considered in this analysis.

### Statistical Analysis

Characteristics of the reproducibility subsample and the remainder of the Sueño cohort were compared to assess for generalizability using t-tests for continuous variables and  $\chi^2$  tests for categorical variables. Differences between scorers were assessed using paired t-tests averaging within scorer data. Agreement between and within scorers was assessed with intraclass correlation coefficients (ICCs) which were computed using a random effects model that simultaneously models both intra-scorer and inter-scorer variability as random effects.<sup>25,26</sup> Age, sex, Hispanic background, and employment status were each assessed as predictors of both the inter-scorer difference and the inter-scorer variance for each of the 11 actigraphic sleep measures using linear regression. Because of the multiple comparisons (4 assessments  $\times$  2 outcomes  $\times$  11 sleep measures) being considered,  $P < 0.01$  was used *a priori* as the threshold for statistical significance in these analyses;  $P < 0.05$  was used to indicate statistical significance in all other analyses. Bland and Altman plots were created to assess systematic differences between scorers and whether any differences between scorers varied across the range of each measure.<sup>27</sup> All P values reflected 2-tailed testing. All analyses were performed using SAS v. 9.2 (SAS Institute, Cary NC).

In interpreting the magnitude of variability identified due to intra- or inter-scorer differences, we considered the minimum clinically important difference for nightly sleep duration, napping duration, 24 hour sleep duration, sleep latency, sleep onset time, sleep offset time, and sleep midpoint time to be 15 minutes and the corresponding difference for sleep maintenance efficiency and sleep fragmentation index to be 5%.

### RESULTS

A total of 2,252 subjects were enrolled in the Sueño ancillary to HCHS/SOL. Of these, 34 were excluded due to  $< 5$  days of valid actigraphy data, leaving 2,218 valid studies. Table 1 compares demographic characteristics between the 50 subjects whose data were selected for this reproducibility analysis and the remaining Sueño participants. There was a trend for those selected to be older with a difference in mean age of 3.1 years ( $P = 0.06$ ). Otherwise, no substantial differences were identified. Similarly, no differences were found in actigraphy characteristics between the reproducibility subsample and the remaining cohort (Table 2).

The mean values for each of the 11 actigraphic variables on each of the 4 assessments are displayed in Table 3. As can be seen from the mean inter-scorer differences, no substantial difference in scoring between the 2 scorers was identified. Both mean nightly and 24-h sleep duration differed between scorers by about 7 minutes. While this was statistically significant, the clinical relevance of a difference of this magnitude is questionable. Similarly, while the mean difference between scorers for sleep fragmentation index was statistically significant and the difference in sleep maintenance efficiency was of borderline significance, the absolute difference of  $< 1\%$  in each measure is not clinically important. Values for sleep latency differed by less than a minute. Among the sleep timing variables, the difference between scorers in both sleep onset and sleep offset was roughly 4 minutes but in opposite directions, resulting in virtually no difference in sleep midpoint time. For all 11

**Table 1**—Demographic characteristics of reproducibility study and remaining Sueño Participants, the Sueño Study 2010–2013.

	Reproducibility Study (n = 50)	Remaining Sueño Participants (n = 2,168)	P value
Age (years)	47.8 ± 10.5	44.7 ± 11.6	0.06
Male	34.0%	35.1%	0.87
Hispanic/Latino Background			0.16
Dominican	4.0%	12.5%	
Central American	22.0%	13.1%	
Cuban	12.0%	17.9%	
Mexican	26.0%	27.1%	
Puerto Rican	30.0%	20.4%	
South American	6.0%	8.2%	
More than one/other	0.0%	0.8%	
High School graduate	58.0%	67.7%	0.15
Employed	60.0%	58.0%	0.78
Body mass index (kg/m <sup>2</sup> )	29.3 ± 5.7	30.3 ± 6.7	0.32
Self-reported sleep duration (h)	8.1 ± 1.6	7.9 ± 1.4	0.49
Epworth sleepiness score	6.3 ± 5.7	5.7 ± 4.5	0.34
Apneaihypopnea index (events/h)	4.8 ± 7.3	4.8 ± 7.8	0.98
Insomnia severity index	7.5 ± 6.3	7.5 ± 6.6	0.98
CES-D10 score	8.9 ± 7.3	7.5 ± 6.2	0.12

Values displayed are mean ± standard deviation or percentage. CES-D10, Center for Epidemiological Studies–Depression 10 item scale.

**Table 2**—Actigraphic characteristics of reproducibility study and remaining Sueño participants, the Sueño Study 2010–2013.

	Reproducibility Study (n = 50)	Remaining Sueño Participants (n = 2,168)	P value
Number of valid days of recording	7.3 ± 1.2	7.1 ± 0.9	0.13
Nightly sleep duration (min)	416.3 ± 73.6	403.5 ± 63.0	0.16
Napping duration (min)	13.7 ± 26.8	11.4 ± 20.7	0.36
24 Hour sleep duration (min)	430.0 ± 69.1	414.9 ± 62.3	0.09
SD of sleep duration (min)	77.8 ± 42.9	72.5 ± 36.0	0.31
Sleep latency (min)	9.2 ± 6.9	10.8 ± 15.0	0.44
Sleep maintenance efficiency (%)	87.2 ± 4.5	85.8 ± 6.1	0.10
Sleep fragmentation index (%)	19.5 ± 5.7	21.3 ± 7.7	0.10
Sleep onset time (HH:MM ± min)	00:10 ± 98	00:01 ± 98	0.49
Sleep offset time (HH:MM ± min)	07:56 ± 101	07:38 ± 96	0.21
Sleep midpoint time (HH:MM ± min)	04:03 ± 91	03:49 ± 91	0.28
SD of sleep midpoint (min)	60 ± 43	55 ± 42	0.45

Values displayed are mean ± standard deviation. SD, standard deviation.

variables, both the intra-scorer and inter-scorer ICCs were extremely high, ranging from 0.911 to 0.995.

The inter-scorer difference and inter-scorer variance for the 11 actigraphic variables did not differ by age, sex, employment status, or Hispanic background ( $P > 0.01$  for all comparisons). In addition, Bland and Altman plots did not reveal substantial discrepancy in these actigraphic variables between scorers or any trend in inter-scorer differences as the average increases (Figure 1).

## DISCUSSION

Actigraphy is increasingly being utilized in large cohort studies to better understand the impact of sleep patterns on health. Unfortunately, few studies have explicitly defined the

methods used to define rest intervals making it difficult for findings in one study to be validated in other cohorts. We have devised explicit instructions for the setting of rest intervals and demonstrate that this standardized algorithm can produce highly reproducible values for a range of actigraphy-derived measures of sleep, and that this reproducibility is consistent across various subpopulations using the same scoring algorithm and thresholds.

Past studies using actigraphy in large cohorts have relied on earlier devices with more limited sensory abilities. In the Study of Women's Health across the Nation (SWAN), rest intervals were defined based on sleep diaries alone.<sup>28</sup> In an ancillary study of the Coronary Artery Risk Development in Young Adults (CARDIA) cohort as well as the Rotterdam Study, rest intervals were defined based on event markers with use of

**Table 3**—Intra- and inter-scorer differences in actigraphy scoring, the Sueño Reproducibility Study (n = 50).

	Scorer 1		Scorer 2		Intra-scorer ICC	Inter-scorer Difference	P value*	Inter-scorer ICC
	Evaluation 1	Evaluation 2	Evaluation 1	Evaluation 2				
Nightly sleep duration (min)	416.3 ± 73.6	410.9 ± 75.6	420.1 ± 73.4	421.5 ± 71.2	0.982	-7.18 ± 12.57	0.0002	0.977
Napping duration (min)	13.7 ± 26.8	12.2 ± 24.0	13.0 ± 23.7	13.1 ± 27.0	0.973	-0.10 ± 2.06	0.75	0.973
24 hour sleep duration (min)	430.0 ± 69.1	423.2 ± 72.1	433.1 ± 70.1	434.6 ± 66.5	0.977	-7.28 ± 12.54	0.0002	0.972
SD of sleep duration (min)	77.8 ± 42.9	76.8 ± 43.6	76.8 ± 42.5	76.8 ± 45.0	0.957	0.48 ± 9.82	0.73	0.957
Sleep latency (min)	9.2 ± 6.9	8.8 ± 6.9	9.1 ± 7.6	8.9 ± 7.1	0.911	-0.02 ± 2.58	0.95	0.911
Sleep maintenance efficiency (%)	87.2 ± 4.5	87.4 ± 4.4	86.9 ± 4.9	87.1 ± 4.6	0.941	0.29 ± 1.07	0.06	0.940
Sleep fragmentation index (%)	19.5 ± 5.7	19.3 ± 5.7	19.7 ± 5.7	19.8 ± 5.6	0.977	-0.33 ± 0.82	0.006	0.976
Sleep onset time (HH:MM ± min)	00:10 ± 98	00:14 ± 101	00:07 ± 98	00:08 ± 99	0.995	4.65 ± 8.04	0.0002	0.994
Sleep offset time (HH:MM ± min)	07:56 ± 101	07:53 ± 98	07:58 ± 100	08:00 ± 99	0.989	-4.33 ± 12.46	0.02	0.988
Sleep midpoint time (HH:MM ± min)	04:03 ± 91	04:04 ± 92	04:03 ± 91	04:04 ± 92	0.995	0.16 ± 7.17	0.87	0.995
SD of sleep midpoint (min)	60.0 ± 42.9	59.3 ± 41.4	60.0 ± 42.3	60.5 ± 43.9	0.988	-0.67 ± 6.08	0.44	0.988

Values displayed are means ± standard deviation. \*P value testing null hypothesis that inter-scorer difference is zero based on paired t-test. SD, standard deviation; ICC, intraclass correlation coefficient.

sleep diaries only when event markers were missing.<sup>29,30</sup> The implementation of event markers in actigraphs has improved the ability to infer when a subject is trying to sleep. The subject is asked to press the marker when they get in or get out of bed. However, failure to remember to press the event marker is a frequent occurrence.<sup>31</sup> In none of these studies is it clear how rest intervals are defined when sleep diaries are not completed, an important issue given the fairly high rate of diary non-compliance.<sup>32</sup> In actigraphic assessments conducted in the Study of Osteoporotic Fractures (SOF) and the Osteoporotic Fractures in Men Study (MrOS) cohorts, sleep diaries were used to set rest intervals, but scorers were allowed to use activity to adjust these intervals when diaries were missing or appeared inaccurate.<sup>12,33</sup> A manual of procedure was created to standardize this subjective process across scorers and reliability data from SOF demonstrate that the protocol used for setting rest intervals produced reproducible data between study scorers.<sup>12</sup> However, details about the protocol for replacing missing or incongruous diary data with activity data have not been published, limiting the ability of others to replicate the scoring approach used. This can be problematic when attempting to compare actigraphy results across studies.

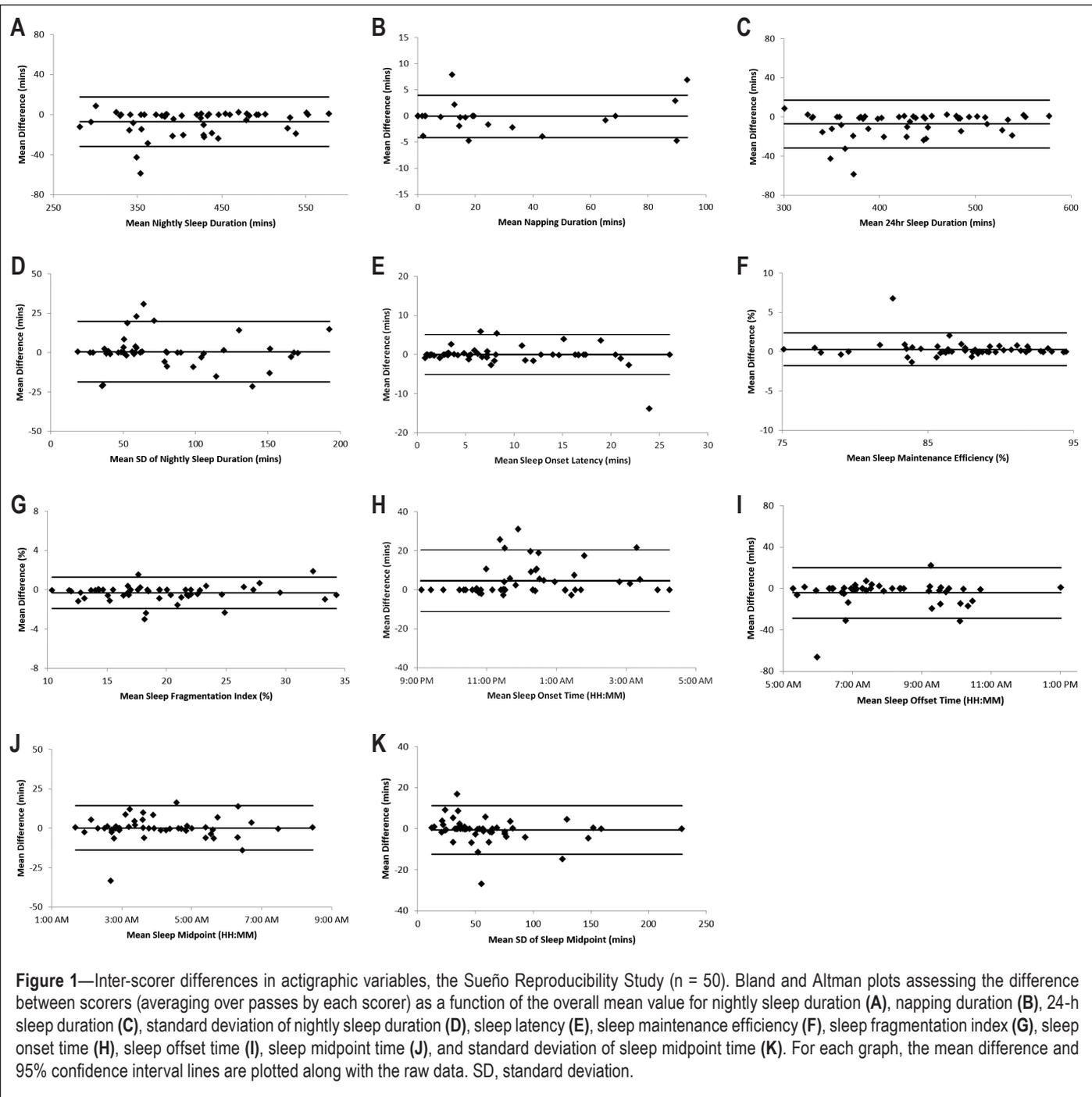
In our study, we utilized a newer device, the Actiwatch Spectrum, which provides illuminance data in addition to event marker and activity data. While photometric performance of the Spectrum device has been questioned, validation studies do find a linear absolute response.<sup>34,35</sup> This supports the use of illuminance data to identify abrupt changes in light intensity as would be expected with turning off and turning on bedroom lights at bedtime and wake time. To our knowledge, light data have not been previously incorporated into actigraphy scoring algorithms in a standardized fashion for defining the rest interval. In addition, we utilized an easy implementable strategy of assessing consistency across input signals. If multiple inputs were within 15 minutes of either, they were deemed consistent and guided rest interval setting while inputs that were not temporally close were deemed inconsistent and devalued. Finally, the Spectrum device has a novel off-wrist detection feature where the lack of a drop in electrical capacitance of

the device case due to the expected capacitive coupling to the conductive properties of the wrist surface is detected. Utilizing this feature allows for censoring of periods when the device is not being used and as a result, likely improves accuracy of identifying true periods of subject inactivity. However, actual performance of the off-wrist detection algorithm has not yet been independently verified.

All 11 actigraphic variables considered in this analysis were found to be highly reproducible with no significant differences in reproducibility across important subgroups. Both nightly sleep duration and 24-h sleep duration had very similar reproducibility and mean difference. The tight relationship relates to the highly reproducible scoring of naps using our conservative strategy of only scoring naps if marked by either event marker or sleep diary. Our high ICCs (above 0.95) for nightly sleep duration and nap duration are comparable to those reported in the SOF cohort. In contrast, our ICCs for sleep latency (0.91) and sleep maintenance efficiency (0.94) suggest greater reliability than those reported in SOF (0.88 and 0.84, respectively), suggesting our more rigid rules may preferentially improve reproducibility for these measures.<sup>12</sup> In terms of the diurnal phase measures (sleep onset, sleep offset, and sleep midpoint), sleep midpoint appeared to be the most robust to scoring variability. Given that sleep midpoint is also less influenced by sleep duration, our data support the use of sleep midpoint as a better marker of circadian phase than other measures commonly obtained from actigraphy. This is consistent with prior research based on self-report data.<sup>36</sup>

Several of the measures assessed have not been previously evaluated for reproducibility in a standardized fashion. However, they have been associated with relevant health outcomes making an understanding of the reproducibility of these measures important. Variability in sleep duration has been associated with subjective sleep quality and well-being,<sup>37</sup> while both the standard deviation of sleep duration and the sleep fragmentation index have been associated with obesity.<sup>30,38</sup>

Limitations of this work should be noted. We did not perform polysomnography, the gold standard of sleep assessment, so while our data speak to the reproducibility of our measures,



**Figure 1**—Inter-scorer differences in actigraphic variables, the Sueño Reproducibility Study (n = 50). Bland and Altman plots assessing the difference between scorers (averaging over passes by each scorer) as a function of the overall mean value for nightly sleep duration (A), napping duration (B), 24-h sleep duration (C), standard deviation of nightly sleep duration (D), sleep latency (E), sleep maintenance efficiency (F), sleep fragmentation index (G), sleep onset time (H), sleep offset time (I), sleep midpoint time (J), and standard deviation of sleep midpoint time (K). For each graph, the mean difference and 95% confidence interval lines are plotted along with the raw data. SD, standard deviation.

we cannot directly assess the accuracy of our scoring strategy. Further research is needed to assess the accuracy of measures derived from such a scoring protocol against electroencephalographic-based measurements of sleep. We also did not compare our results to alternative scoring strategies such as the standard practice of relying on best judgment of the scorer or using a strategy with a different hierarchy of scoring inputs. As such, we are unable to demonstrate directly whether our standardized strategy provides an improvement in accuracy or reliability over other methods. However, by providing a clear and detailed protocol for scoring, we allow others to replicate our scoring strategy and determine whether sleep patterns in other populations are similar or different from the cohort evaluated in this study.

It should be noted that this study was conducted in a middle aged Hispanic/Latino population screened to exclude severe sleep apnea and narcolepsy. The generalizability of our findings to younger or older populations who may have different levels of physical activity or different sleep patterns including napping behaviors as well as different rates of compliance with the use of event markers and sleep diaries is unclear. The importance of carefully setting rest intervals is magnified in populations where sleep habits are irregular such as our cohort, which was made up of young adults with a high prevalence of shift work and a high rate of napping.

In conclusion, our findings suggest the use of a standardized algorithm incorporating data from the multiple inputs available

in modern actigraphy limits intra- and inter-scorer variability, thereby providing reproducible sleep/wake summary data

## DISCLOSURE STATEMENT

This was not an industry supported study. This study was supported by NHLBI HL098297. In addition, the Hispanic Community Health Study/Study of Latinos was carried out as a collaborative study supported by contracts from NHLBI to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Center/Offices contribute to the HCHS/SOL through a transfer of funds to the NHLBI: National Center on Minority Health and Health Disparities, the National Institute of Deafness and Other Communications Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the Office of Dietary Supplements. No off-label or investigational uses were evaluated in this study. Dr. Zee has consulted for Philips Respironics, Aptalis, Jazz, Vanda, and Merck and has stock in Teva. Dr. Reid has received research support from Philips Consumer Lifestyles. The other authors have indicated no financial conflicts of interest.

## REFERENCES

1. Patel SR, Malhotra A, White DP, Gottlieb DJ, Hu FB. Association between reduced sleep and weight gain in women. *Am J Epidemiol* 2006;164:947–54.
2. Gangwisch JE, Heymsfield SB, Boden-Albala B, et al. Short sleep duration as a risk factor for hypertension: analyses of the first National Health and Nutrition Examination Survey. *Hypertension* 2006;47:833–9.
3. Patel SR, Malhotra A, Gao X, Hu FB, Neuman MI, Fawzi WW. A prospective study of sleep duration and pneumonia risk in women. *Sleep* 2012;35:97–101.
4. Yaggi HK, Araujo AB, McKinlay JB. Sleep duration as a risk factor for the development of type 2 diabetes. *Diabetes Care* 2006;29:657–61.
5. Ayas NT, White DP, Manson JE, et al. A prospective study of sleep duration and coronary heart disease in women. *Arch Intern Med* 2003;163:205–9.
6. Kripke DF, Garfinkel L, Wingard DL, Klauber MR, Marler MR. Mortality associated with sleep duration and insomnia. *Arch Gen Psychiatry* 2002;59:131–6.
7. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? *Epidemiology* 2008;19:838–45.
8. Kushida CA, Chang A, Gakary C, Guilleminault C, Carrillo O, Dement WC. Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Med* 2001;2:389–96.
9. Girardin JL, Kripke DF, Mason WJ, Elliot JA, Youngstedt SD. Sleep estimation from wrist movement quantified by different actigraphic modalities. *J Neurosci Methods* 2001;105:185–91.
10. Kripke DF, Hahn EK, Grizas AP, et al. Wrist actigraphic scoring for sleep laboratory patients: algorithm development. *J Sleep Res* 2010;19:612–9.
11. Littner M, Kushida CA, Anderson WM, et al. Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. *Sleep* 2003;26:337–41.
12. Blackwell T, Ancoli-Israel S, Gehrman PR, Schneider JL, Pedula KL, Stone KL. Actigraphy scoring reliability in the Study of Osteoporotic Fractures. *Sleep* 2005;28:1599–605.
13. Lavange LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* 2010;20:642–9.

14. Sorlie PD, Aviles-Santa LM, Wassertheil-Smoller S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol* 2010;20:629–41.
15. Westbrook PR, Levendowski DJ, Cvetinovic M, et al. Description and validation of the apnea risk evaluation system: a novel method to diagnose sleep apnea-hypopnea in the home. *Chest* 2005;128:2166–75.
16. Redline S, Sotres-Alvarez D, Loredo J, et al. Sleep-disordered breathing in Hispanic/Latino individuals of diverse backgrounds. *The Hispanic Community Health Study/Study of Latinos. Am J Respir Crit Care Med* 2014;189:335–44.
17. Lind BK, Goodwin JL, Hill JG, Ali T, Redline S, Quan SF. Recruitment of healthy adults into a study of overnight sleep monitoring in the home: experience of the Sleep Heart Health Study. *Sleep Breath* 2003;7:13–24.
18. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* 1991;14:540–5.
19. Bastien CH, Vallieres A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med* 2001;2:297–307.
20. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Measure* 1977;1:385–401.
21. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med* 1994;10:77–84.
22. Oakley NR. Validation with polysomnography of the Sleep-watch sleep/wake scoring algorithm used by the Actiwatch activity monitoring system. Technical Report to Mini Mitter Co., Inc., 1997.
23. Marino M, Li Y, Rueschman MN, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep* 2013;36:1747–55.
24. Kurina LM, Knutson KL, Hawkey LC, Cacioppo JT, Lauderdale DS, Ober C. Loneliness is associated with sleep fragmentation in a communal society. *Sleep* 2011;34:1519–26.
25. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996;1:30–46.
26. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res* 1998;7:301–17.
27. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
28. Appelhans BM, Janssen I, Cursio JF, et al. Sleep duration and weight change in midlife women: the SWAN sleep study. *Obesity (Silver Spring)* 2013;21:77–84.
29. Lauderdale DS, Knutson KL, Yan LL, et al. Objectively measured sleep characteristics among early-middle-aged adults: the CARDIA study. *Am J Epidemiol* 2006;164:5–16.
30. van den Berg JF, Knvistingh Neven A, Tulen JH, et al. Actigraphic sleep duration and fragmentation are related to obesity in the elderly: the Rotterdam Study. *Int J Obes (Lond)* 2008;32:1083–90.
31. Ustinov Y, Lichstein KL. Actigraphy reliability with normal sleepers. *Behav Sleep Med* 2013;11:313–20.
32. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. *Control Clin Trials* 2003;24:182–99.
33. Patel SR, Blackwell T, Redline S, et al. The association between sleep duration and obesity in older adults. *Int J Obes (Lond)* 2008;32:1825–34.
34. Price L, Khazova M, O'Hagan J. Performance assessment of commercial circadian personal exposure devices. *Light Res Technol* 2012;44:17–26.
35. Figueiro MG, Hamner R, Bierman A, Rea MS. Comparisons of three practical field devices used to measure personal light exposures and activity levels. *Light Res Technol* 2013;45:421–34.
36. Martin SK, Eastman CI. Sleep logs of young adults with self-selected sleep times predict the dim light melatonin onset. *Chronobiol Int* 2002;19:695–707.
37. Lemola S, Ledermann T, Friedman EM. Variability of sleep duration is related to subjective sleep quality and subjective well-being: an actigraphy study. *PLoS One* 2013;8:e71292.
38. Patel SR, Hayes AL, Blackwell T, et al. The association between sleep patterns and obesity in older adults. *Int J Obes (Lond)* 2014;38:1159–64.