Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography

Miguel Marino, PhD[1,2]; Yi Li, PhD[3]; Michael N. Rueschman, MPH[4]; J. W. Winkelman, MD, PhD[5,6]; J. M. Ellenbogen, MMSc, MD[7]; J. M. Solet, PhD[5,8]; Hilary Dulin, BS[4]; Lisa F. Berkman, PhD[9,10]; Orfeu M. Buxton PhD[4,5,11,12]

[1]Department of Biostatistics, Harvard School of Public Health, Boston MA; [2]Department of Family Medicine, Oregon Health and Science University, Portland, OR; [3]Department of Biostatistics, University of Michigan, Ann Arbor MI; [4]Department of Medicine Brigham and Women's Hospital, Boston MA; [5]Division of Sleep Medicine, Harvard Medical School, Boston MA; [6]Department of Psychiatry, Massachusetts General Hospital, Boston, MA; [7]Department of Neurology, Johns Hopkins University, Baltimore, MD; [8]Department of Medicine-Massachusetts General Hospital, Cambridge Health Alliance, Cambridge MA; [9]Harvard Center for Population and Development Studies, Cambridge MA; [10]Harvard School of Public Health, Boston MA; [11]Department of Social and Behavioral Sciences, Harvard School of Public Health, Cambridge, MA; [12]Department of Biobehavioral Health, Pennsylvania State University, University Park, PA

Objectives: We validated actigraphy for detecting sleep and wakefulness versus polysomnography (PSG).
Design: Actigraphy and polysomnography were simultaneously collected during sleep laboratory admissions. All studies involved 8.5 h time in bed, except for sleep restriction studies. Epochs (30-sec; n = 232,849) were characterized for sensitivity (actigraphy = sleep when PSG = sleep), specificity (actigraphy = wake when PSG = wake), and accuracy (total proportion correct); the amount of wakefulness after sleep onset (WASO) was also assessed. A generalized estimating equation (GEE) model included age, gender, insomnia diagnosis, and daytime/nighttime sleep timing factors.
Setting: Controlled sleep laboratory conditions.
Participants: Young and older adults, healthy or chronic primary insomniac (PI) patients, and daytime sleep of 23 night-workers (n = 77, age 35.0 ± 12.5, 30F, mean nights = 3.2).
Interventions: N/A.
Measurements and Results: Overall, sensitivity (0.965) and accuracy (0.863) were high, whereas specificity (0.329) was low; each was only slightly modified by gender, insomnia, day/night sleep timing (magnitude of change < 0.04). Increasing age slightly reduced specificity. Mean WASO/night was 49.1 min by PSG compared to 36.8 min/night by actigraphy (β = 0.81; CI = 0.42, 1.21), unbiased when WASO < 30 min/night, and overestimated when WASO > 30 min/night.
Conclusions: This validation quantifies strengths and weaknesses of actigraphy as a tool measuring sleep in clinical and population studies. Overall, the participant-specific accuracy is relatively high, and for most participants, above 80%. We validate this finding across multiple nights and a variety of adults across much of the young to midlife years, in both men and women, in those with and without insomnia, and in 77 participants. We conclude that actigraphy is overall a useful and valid means for estimating total sleep time and wakefulness after sleep onset in field and workplace studies, with some limitations in specificity.
Keywords: Actigraphy, polysomnography, WASO, sensitivity, specificity, accuracy
Citation: Marino M; Li Y; Rueschman MN; Winkelman JW; Ellenbogen JM; Solet JM; Dulin H; Berkman LF; Buxton OM. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *SLEEP* 2013;36(11):1747-1755.

INTRODUCTION

Polysomnography (PSG) is the current gold standard for measuring sleep. This technique employs numerous collections of surface electrodes, each measuring physiologic parameters of sleep, including brain dynamics of electroencephalography (EEG), eye movements, muscle activity, heart physiology, and respiratory function. To achieve all this, individuals typically spend the night in a sleep laboratory—a controlled setting under the continued supervision of a sleep technician. Time-series data are aggregated, processed, and visually examined or mathematically transformed in order to reveal insights about sleep-wake states and many aspects of physiology.

The predictable state of immobility, relative to wakefulness, is a characteristic feature of sleep. Taking advantage of

this distinctive feature of sleep, clinicians and researchers have attempted to measure the binary presence of sleep or waking states by measuring wrist movements. This approach supports large-scale, population-level sleep research by facilitating inexpensive, unobtrusive sleep measurement without disrupting sleep as PSG sometimes does, and enables measurement across a wide range of circumstances and locations. The resulting opportunity for high participation rates can enhance generalizability of results and also renders longitudinal and repeated measure designs more feasible. Wrist actigraphy, measurement of wrist movements to assess sleep or waking state, is accomplished through an accelerometer in a wrist worn device. However, limited validations exist relative to the gold standard of PSG. As population studies continue to be increasingly valued, for example by the Healthy People 2020 Framework (2010),[1] researchers need a range of validated techniques beyond PSG alone.

As compared with PSG, actigraphy is known to overestimate sleep and underestimate wake time.[2] This is presumed in part to be because PSG and actigraphy mark the beginning of sleep periods in different ways. In an actigraphy recording, immobility of the participant marks the beginning of the sleep period,

Table 1—Participant characteristics

Study	N	Age	Male %	White %	Black %	Other %	Acc	Sens	Spec
Acoustics pilot	2	29.5 ± 0.7	100.0	50.0	50.0	0.0	0.830	0.957	0.327
Acoustics III	10	22.2 ± 2.1	30.0	100.0	0.0	0.0	0.931	0.965	0.299
Tiagabine	9	56.4 ± 3.4	44.4	88.9	0.0	11.1	0.834	0.970	0.307
Sleep Restriction	16	27.2 ± 4.9	100.0	43.7	37.5	18.8	0.870	0.961	0.347
Insomnia	17	40.5 ± 8.2	58.5	94.1	0.0	5.9	0.832	0.947	0.337
Nightwork	23	35.2 ± 9.2	52.2	65.2	30.4	4.4	0.878	0.975	0.323
Total	77	35.0 ± 12.5	61.0	74.0	18.2	7.8	0.863	0.965	0.329

Acc, accuracy; Sens, sensitivity; Spec, specificity.

whereas with PSG, rather than simple immobility, stereotypical changes in brain electrical activity patterns mark the onset of sleep. These changes can often begin well after a period of wrist immobility, and thus actigraphy may overestimate sleep time, particularly in those with abundant wakefulness throughout the night, as can be seen in insomnia.[3]

Additional aspects of actigraphy measurement lack critical confirmation including standardization for actigraphy device settings, reporting of actigraphy analysis parameters, and a subjectively determined time of the analyzed rest period for analysis. The adequacy of current scoring algorithms for actigraphy, known to vary with different participant types, sensitivity (correctly assigned sleep epochs), and specificity (correctly assigned wake epochs) are not reported, further complicating efforts to better standardize use of actigraphy as a measure of sleep.[4,5]

A current actigraphy algorithm (Cole-Kripke),[4] and a new algorithm, the Scripps Clinic algorithm[6] have been validated for the Actiwatch and Spectrum devices in healthy individuals and those with sleep apnea. The possible effects of gender or medication use were not determined. Both the Cole-Kripke and the Scripps Clinic algorithms use a weighted moving average compared to a fixed score threshold to score a given epoch as sleep or wake. Both the Scripps Clinic and the manufacturer's algorithm were reported to have an 87% agreement rate of total overnight sleep time to gold standard PSG in these populations.[6] Sensitivity (ability to correctly identify sleep), specificity (ability to correctly identify wake), and bias were not calculated for these algorithms.[6] In past studies, the agreement and sensitivity between actigraphic and PSG measures of sleep have been satisfactory, but the specificity has been low.[7,8] Specificity has been much higher in studies of nocturnal sleep in children and adolescents, 54% to 77%,[9] and healthy adults with short sleep periods, 96%.[10]

In this study, we evaluate the performance of a specific wrist actigraphy device and algorithm by comparing its performance with PSG across a range of populations and conditions: older adults, healthy sleep restricted participants, healthy participants on study control nights with sham noise administration, insomniacs, and night-workers during daytime sleep. Analyses include the data from 77 participants who had simultaneous PSG and actigraphy recordings while sleeping in the sleep laboratory. The goal of this study was to assess correspondence between actigraphy and PSG using two metrics: epoch-by-epoch analysis and sleep parameter concordance analysis. In our epoch-by-epoch analysis, we focused on evaluating accuracy, sensitivity, and specificity between actigraphy and PSG, and key participant factors that affect those measurements. Furthermore, we compared estimates of wake after sleep onset (WASO) on a nightly basis as measured by actigraphy versus PSG in the sleep parameter concordance analysis.

METHODS

Participants

To evaluate the agreement of a common wrist actigraphy algorithm ("Cole-Kripke")[4] to PSG in this retrospective study, actigraphy data and PSG data were collected simultaneously through inpatient sleep laboratory visits during several studies (Table 1). A total of 90 patients were considered from the following studies: insomnia[11]; baseline sleep in healthy participants from a pilot study and published studies[12-14]; older adults [unpublished]; sleep restriction in healthy participants[15]; and daytime sleep in night-workers.[16] Data from participants were excluded on nights when they were administered a medication that could affect sleep. The second and third nights of sleep were also excluded for participants in the acoustics study when sleep-disrupting stimuli were presented, and thus only control/sham (non-noise) nights are included for this analysis.[12-14] Finally, epochs where PSG or actigraphy measurements were missing were removed from the analyses.

Actigraphy Measures

Actigraphy data were collected using 2 types of devices: the AW-64 (Minimitter, Inc, Bend, OR) and the Actiwatch Spectrum (Philips/Respironics, Murrysville, PA). Both of these devices measure wrist movement time series using the digital integration method.[17] AW-64 devices (Minimitter, Inc, Bend, OR) were configured to collect data in 30-s epochs. The data was downloaded using Actiware software (version 3.4, Philips/Respironics, Murrysville, PA) and imported to Actiware 5 (versions 5.57 and 5.59, Respironics) for analysis. Spectrum devices were configured to collect activity in 30-s epochs, during white, red, blue, and green light exposures, but light differences were not included in the present analyses. Spectrum devices can also detect whether the device is on or off wrist; here off-wrist periods were set to missing. Data from the Actiwatch spectrums was downloaded and analyzed in Respironics Actiware 5 (versions 5.57 and 5.59).

PSG Measures

During each sleep period, electrodes were applied to each participant's face and scalp prior to the sleep period. For both systems, data was collected through an electroencephalogram (C3, C4, O1, and O2 placement), an electrooculogram, an electromyogram, and an electrocardiogram to determine the participants' sleep/wake states. Two different systems were used to collect PSG data, the Vitaport-3 digital sleep recorders (TEMEC; Beckman Instrument Company, Schiller Park, IL) and the Comet XL system (Grass Technologies, West Warwick, RI). The Vitaport system was used to collect data for the older adult (unpublished), sleep restriction,[18] insomnia,[19] and acoustics pilot studies (unpublished), and the data were scored according to the methods outlined in Rechtschaffen and Kales.[20] The Comet XL system was used to collect data for the acoustics[14] and night work studies. For later studies, acoustics[14] and night work[16] were scored according to the methods outlined in Iber et al.[21] All studies were scored visually in 30-s epochs by registered polysomnographic technologists. The start and end of the analysis window, or "rest" period, was set based on the lights-off and lights-on times recorded at the laboratory. Both actigraphy and PSG computers were networked and times updated automatically.

Temporal Alignment

To evaluate epoch-by-epoch agreement, temporal synchronization between actigraphy and PSG is essential. It is standard practice at the beginning of each sleep session to synchronize PSG and actigraphy clocks to begin recording at the same time. Occasionally, temporal gaps occur between the 2 measures. To verify synchronization of actigraphy and PSG timing, we compared specific epoch intervals across the entire sleep period. This comparison focused on unique periods of PSG-scored wake or movement that were 1, 2, or 3 epochs long. After flagging this subset of epochs of interest in a given PSG session, we calculated the proportion of these epochs that corresponded to actigraphy epochs showing activity counts ≥ 15. Subsequently, to confirm time alignment between the PSG and actigraphy computers, this calculation was repeated after systematically shifting the actigraphy timeline 10 epochs (5 min) in each direction. For each PSG session, we then realigned the PSG and actigraphy data at the point where the proportion of matched epochs was highest. For sessions that did not show a clear peak in the proportion of matched epochs, the alignment was secured. Detection of a clear peak in the proportion of matched epochs was assessed using scan statistics assuming a discrete Poisson model.[22] Scan statistics were used to assess whether the peak in proportion of matched epochs could be accounted for by chance only. The procedure can be described as follows: a changing window across time is gradually scanned, noting the number of observed and expected observations in and outside that window. The center of the window location and radius of the window are then varied. Using a Poisson model, the window with the maximum likelihood is selected and a P-value is obtained through Monte Carlo hypothesis testing. A P-value < 0.05 suggests a clear peak in the proportion of matched epochs. This procedure was performed using the SaTScan statistical program (Information Management Services Inc, Boston MA).[22] We compare the performance

Table 2—Epochs of actigraphy and polysomnography (30-second)

Original Timing

		Actigraphy			
		"Sleep"	"Wake"		Accuracy = 0.848
PSG	Sleep	186,889	8,794	195,683	Sensitivity = 0.955
	Wake	26,572	10,594	37,166	Specificity = 0.285
		213,461	19,388	232,849	

Modified Timing (temporally aligned PSG and actigraphy files used for all analyses)

		Actigraphy			
		"Sleep"	"Wake"		Accuracy = 0.863
PSG	Sleep	188,746	6,937	195,683	Sensitivity = 0.965
	Wake	24,920	12,246	37,166	Specificity = 0.329
		213,666	19,183	232,849	

of this proposed temporal alignment to the original timing by comparing accuracy, sensitivity, and specificity values from the original timing data and the modified timing data.

Once timing was aligned according to the procedure described in Methods, the agreement of PSG with actigraphy was recalculated, in the same manner as used with the original timing.[12] This modified time compensated for any shift or misalignment in the computers controlling the PSG or actigraphy clocks. Accuracy between the classifications of sleep and wake improved from the original timing (0.848) to the modified timing (0.863). This modification produced an average of 6.8 more epochs, or 3.4 min being correctly classified as sleep per night, and an average of 6.0 more epochs or 3 min correctly classified as wake. The sensitivity increased from (0.955) to (0.965) and the specificity increased from (0.285) to (0.329) by using the modified timing. The confusion matrix for the modified timing is also presented in Table 2. A confusion matrix is a cross-tabular representation of 2 rows containing actual sleep and wake epochs classified by PSG and 2 columns containing predicted sleep and wake epochs classified by actigraphy. The differences in sensitivity measures between the original timing and modified timing were not significant according to the McNemar test (P = 0.1023). The modified timing did have a significantly higher specificity as compared with the original timing (P < 0.0001).

Statistical Analysis

To assess concordance between actigraphy and PSG, 2 types of analyses were performed: epoch-by-epoch analysis and sleep parameter analysis. For the epoch-by-epoch analysis, the primary measures of agreement between PSG and actigraphy were measures of accuracy, sensitivity, and specificity. Polysomnography was used as the reference standard for which performance of actigraphy could be compared. *Accuracy* is defined by a proportion: the total number of 30-s epochs of sleep (defined by PSG) that were correctly classified by actigraphy, divided by the total number of events classified (be they correct or incorrect). *Sensitivity* for sleep corresponds to the proportion of epochs PSG-scored as *sleep* epochs that are correctly classified as sleep by actigraphy. *Specificity* for sleep corresponds to
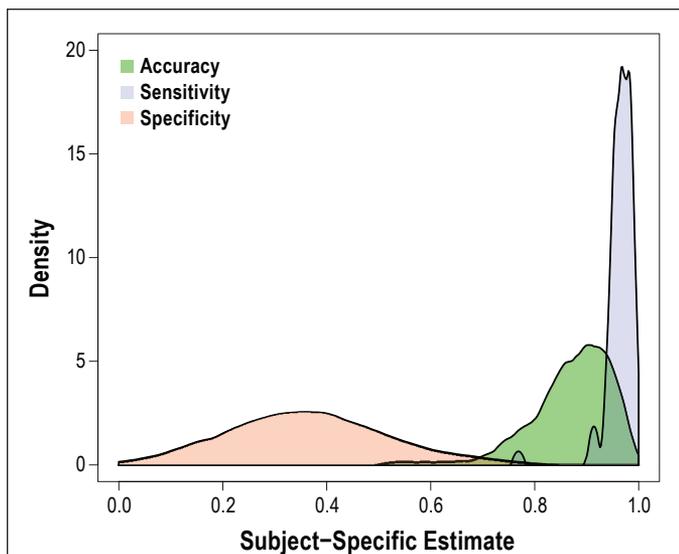
Figure 1—Kernel Density Estimates of the distribution of individual accuracy, sensitivity and specificity in the study. Sensitivity reflects the proportion of 30-sec epochs actigraphically defined correctly as sleep relative to the gold standard PSG. Specificity reflects the proportion of epochs correctly assigned as wake. Accuracy refers to the proportion of correctly characterized epochs relative to all epochs.

the proportion of epochs PSG-scored as *wake* epochs that are correctly classified as wake epochs by actigraphy. To assess how age, gender, time of sleep (sleep during the day or sleep during the night), and insomnia affected the accuracy, sensitivity, and specificity of actigraphy in our study, we performed a generalized estimating equation (GEE) analysis. The GEE model is a convenient and general approach for analysis of the clustering present in our study (where epochs are clustered within sleep periods, which are themselves clustered within individuals). The dichotomous nature of sleep/wake can be modeled with a GEE model assuming a Bernoulli variance function with logit link and an exchangeable correlation structure.[21]

For the sleep parameter concordance analysis, we calculated measurements of PSG and actigraphy wake time after sleep onset (WASO) for every night of every participant using the modified timing data set. To evaluate the relationship between PSG WASO and actigraphy WASO, we calculated the average WASO per night for each participant and we assessed correlation through a Spearman rank correlation measure. We also performed a linear regression of PSG WASO on actigraphy WASO to obtain a best-fit line. To determine whether there exist any differential effects of actigraphy as compared with PSG with relation to WASO between participants with or without chronic primary insomnia, we constructed linear regression analyses with PSG WASO as the outcome and actigraphy WASO, presence of chronic primary insomnia and the interaction between actigraphy WASO and insomnia state as covariates. A similar linear regression exploration with interaction terms was performed for age. A multivariable linear regression analysis with actigraphy WASO, insomnia, age, time of sleep, gender, and the two-way interactions of the predictors with actigraphy WASO was also performed.

Additionally, to assess agreement between PSG WASO and actigraphy WASO, we calculated the mean difference between

them and its corresponding standard deviation. The mean difference between PSG WASO and actigraphy WASO provides a measure of bias of the actigraphy device which we display using a Bland-Altman plot.[23] To build a Bland-Altman plot, we computed the mean WASO of PSG and actigraphy combined for each participant and plot it on the X-axis. We then plot that value against the difference in WASO measurement between actigraphy and PSG for each participant. Finally, we calculate the mean difference in WASO and its corresponding standard deviation. If actigraphy underestimates WASO and therefore underestimates time awake as compared with PSG, this will show up in the plot as a negative mean difference. Statistical analyses were conducted using Stata version 11.2 (StataCorp, College Station, TX) and R version 2.13.2 (open source).

RESULTS

Participants

The data in this study are a compilation of all available simultaneous PSG and actigraphy recordings from 6 different inpatient studies with a variety of subjects including: people with insomnia, healthy participants, older adults, healthy sleep-restricted participants, and night-workers. In total, 90 subjects participated in these studies. All available recordings were used in each study during sleep periods during which no study drug or disruptive noise was administered. Across all studies, 13 participants were excluded due to missing PSG (actigraphy data was available for every qualifying inpatient night), leaving 77 participants remaining in the analysis. Study sample characteristics are shown in Table 1. Age of participants at the time of sleep recording ranged from 20 to 61 years, with a mean of 35.0 years. The sample consisted of 61% males and 74% white participants. Chronic primary insomnia was present in 22% of the study participants.

Epoch-by-Epoch Results

We compared the concordance of actigraphy using Actiware 5 algorithms (versions 5.57 and 5.59, Respironics)[4] to the PSG, determining the accuracy, sensitivity, and specificity of the algorithms for detecting sleep and wake. The accuracy of actigraphy (0.863), the sensitivity (0.965), and specificity (0.329) is described in the confusion matrix in Table 2. Accuracy, sensitivity, and specificity were also calculated for each individual study, and no discernible differences were observed between the studies (results not shown).

To assess the variation in performance of actigraphy across individuals, accuracy, sensitivity, and specificity values were also calculated for each individual in the study. Figure 1 presents the kernel density estimates of the distribution of individual accuracy, sensitivity, and specificity values for the participants in the study using the modified timing. Overall, sensitivity of the actigraphy device is extremely high and does not vary substantially between individuals as evidenced by its narrow distribution. However, specificity was quite low and the distribution of specificity for individual participants is relatively flat, ranging from 0.0 to 0.7, which points to the substantial variability of actigraphy in properly defining sleep. The distribution of individual accuracy seems to be negatively affected by the low specificity, but overall, its distribution is moderately

Actigraphy Validation vs Polysomnography in Varied Groups—Marino et al

bell-shaped, and most observations lie above an accuracy of 0.80.

The covariate effects on the accuracy, sensitivity, and specificity of actigraphy on PSG were analyzed using a GEE model with logit link. The covariates selected for adjustment were age, gender, chronic primary insomnia, and time of sleep (daytime or nighttime). The multivariable GEE analysis simultaneously models PSG, all of the covariates and the two-way interaction term between PSG and the covariate.[24] To calculate accuracy, sensitivity, and specificity values for each covariate, adjusting for the other covariates, we performed an adjusted probabilities approach. The adjusted probabilities approach takes the differences in other predictors in the model into account by holding every other covariate constant in its mean value and using the estimated parameters from the GEE model to estimate accuracy, sensitivity, and specificity. This approach allows us to evaluate the performance of actigraphy across the covariate of interest while adjusting for the typical response of other factors in the model.

The results of the adjusted probabilities approach are described in Table 3 for gender, insomnia, and time of sleep. Females have slightly higher accuracy and sensitivity values but a discernibly lower specificity than males. There appears to be very little difference in performance of actigraphy on PSG related to day or nighttime sleep timing, with accuracy, sensitivity, and specificity being within 2% across day and night sleepers. Recordings from participants with chronic primary insomnia demonstrate lower sensitivity and accuracy than those of participants without insomnia, even after controlling for age, gender, and time of sleep. Univariable analyses were also performed and showed similar findings (results not shown). The overall mean sleep efficiency for insomnia patients in our sample was 83.1%; for subjects without insomnia, the mean sleep efficiency for the patients was 85.6%. For insomnia patients with sleep efficiency > 90%, the specificity was 0.355, while the specificity for insomnia patients with sleep efficiency < 85% was 0.302. Thus there is a slight effect of insomnia severity using this contrast. Overall specificity for females with insomnia is 0.351. For females aged < 40, the specificity is 0.418, whereas for females aged ≥ 40, the specificity is 0.250.

In multivariable GEE analyses of age, adjusted predicted probabilities were calculated and used to assess the effect of age on performance of actigraphy on PSG. To illustrate the age effect, we constructed lowess smoothing curves on the accuracy, sensitivity, and specificity values by age as determined by the GEE model (Figure 2). Figure 2 suggests that age has no meaningful impact on the sensitivity of actigraphy, but does have a negative, yet minimal, effect on the specificity, which slightly decreases the accuracy of actigraphy as age increases. We also performed a univariable analysis of age; the observations from the analysis are similar to what is described in Figure 2 (data not shown).

Table 3—Univariable and multivariable GEE models assessing the effect of gender, age, and time of sleep on accuracy, sensitivity, and specificity

	Accuracy		Sensitivity		Specificity	
	Estimate	P-value	Estimate	P-value	Estimate	P-value
Overall	0.865	n/a	0.965	n/a	0.333	n/a
Gender						
Male	0.854	0.428	0.962	0.237	0.336	0.752
Female	0.893		0.971		0.325	
Time of Sleep						
Day	0.866	0.105	0.972	0.030	0.325	0.802
Night	0.865		0.963		0.334	
Insomnia						
Yes	0.833	0.516	0.946	0.132	0.347	0.680
No	0.869		0.967		0.331	

These models are adjusted for age. The findings for age are presented in Figure 2. The P-values for the effect of age on accuracy, sensitivity and specificity were not significant.
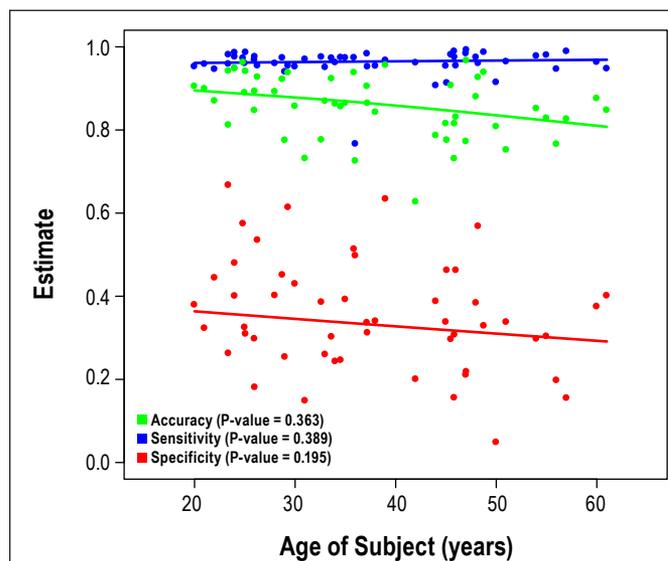

Figure 2—Age effect on accuracy, sensitivity, and specificity of actigraphy on PSG based on GEE multivariate regression adjusting for gender, time of sleep, and chronic primary insomnia.

Sleep Parameter Concordance Results

Using the actigraphy with the modified timing, we also calculated the amount of PSG and actigraphy WASO per sleep period for each participant. The mean amount of PSG WASO per sleep period was 50.6 min, compared to 38.0 min for actigraphy WASO. Figure 3A presents the scatter plot of PSG WASO versus actigraphy WASO and their corresponding distribution histograms. We performed a Spearman rank correlation of PSG WASO on actigraphy WASO to assess their correlation. We also present a simple linear regression of PSG WASO on actigraphy WASO and plot the best-fit line and its 95% confidence bands in Figure 3A. The regression coefficient ($\beta = 0.81$; 95% CI = 0.42, 1.21) and positive Spearman rank correlation ($r_s = 0.611$) suggests a positive and statistically significant correlation between actigraphy and PSG WASO (P < 0.0001).

We investigated the differential effects of age and insomnia on the association between actigraphy and PSG WASO through
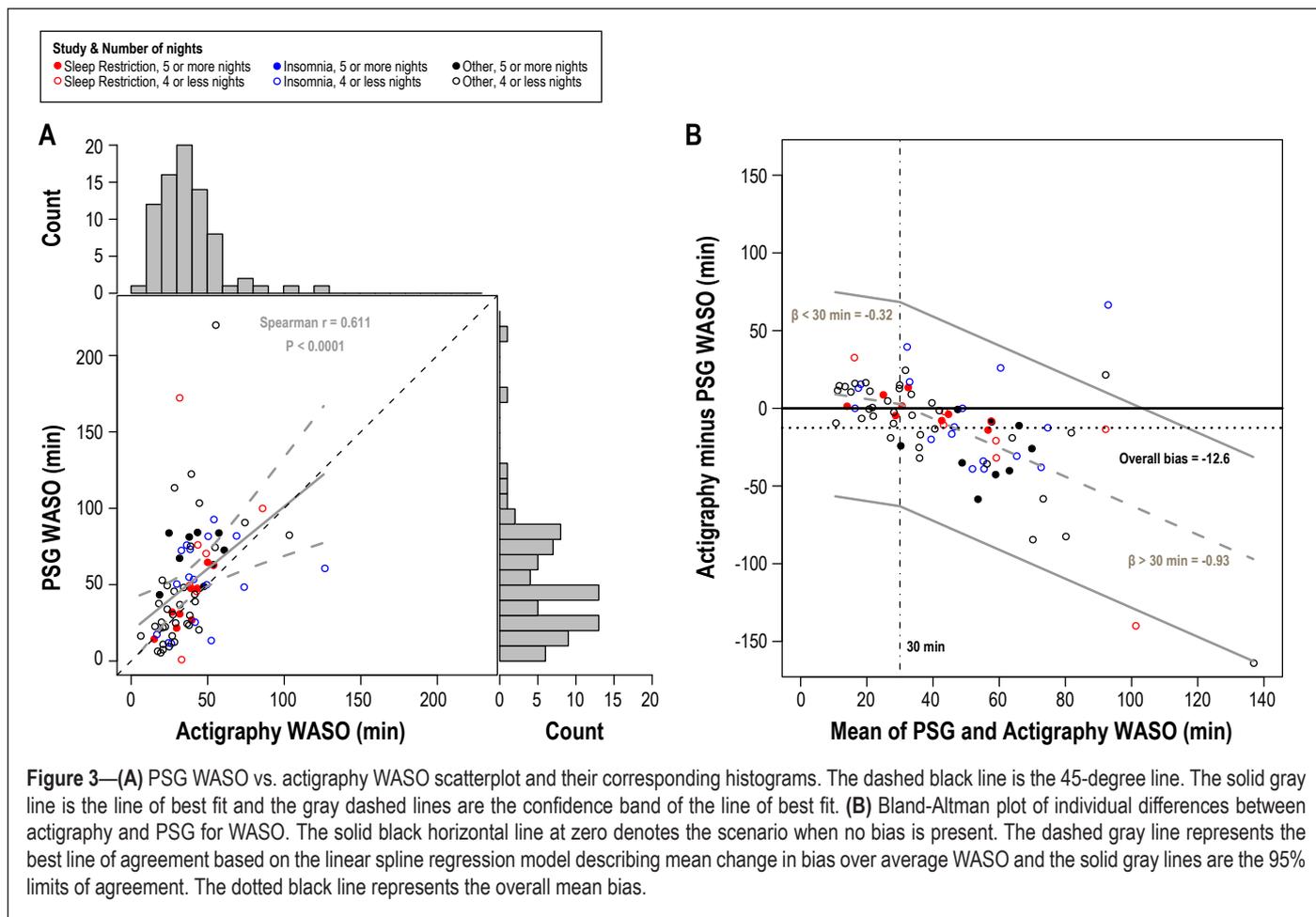
Figure 3—**(A)** PSG WASO vs. actigraphy WASO scatterplot and their corresponding histograms. The dashed black line is the 45-degree line. The solid gray line is the line of best fit and the gray dashed lines are the confidence band of the line of best fit. **(B)** Bland-Altman plot of individual differences between actigraphy and PSG for WASO. The solid black horizontal line at zero denotes the scenario when no bias is present. The dashed gray line represents the best line of agreement based on the linear spline regression model describing mean change in bias over average WASO and the solid gray lines are the 95% limits of agreement. The dotted black line represents the overall mean bias.

Table 4—Univariable and multivariable linear regression of PSG WASO on actigraphy WASO to determine the differential effect of actigraphy on PSG by chronic primary insomnia and age

| | Interaction × Actigraphy | | |
	Estimate	95% CI	P-value
Gender			
Male	-0.093	-1.010, 0.824	0.840
Female			
Time of Sleep			
Day			
Night	-0.422	-1.485, 0.641	0.431
Insomnia			
Yes	-0.489	-1.586, 0.608	0.376
No			
Age	-0.038	-0.081, 0.003	0.070

None of the covariates significantly modify the performance of Actigraphy WASO on PSG WASO.

univariable and multivariable regression analyses. To estimate the differential effects of gender, time of sleep, age, and insomnia, we constructed two-way interaction terms of the covariates by actigraphy WASO. Table 4 presents the multivariable regression results. Presence of insomnia did not materially change the association between actigraphy and PSG in relation to WASO. The interaction term between insomnia and

actigraphy (-0.71) suggests that subjects with insomnia have slightly lower correspondence between actigraphy and PSG WASO than subjects without insomnia. However, this result is not statistically significant (P = 0.0747). The age by actigraphy interaction term is also nonsignificant (coefficient = -0.03; P = 0.0911). The nonsignificance of the interaction terms for age, time of sleep, gender, and insomnia suggest that there is no statistical evidence for age or insomnia modifying the association between actigraphy and PSG WASO.

In Figure 3B, we construct the Bland-Altman plot to visually inspect the level of agreement between actigraphy and PSG WASO. Overall mean bias was estimated at minus 12.6 (SD 34.0 min), suggesting that actigraphy tended to underestimate minutes of WASO compared to PSG. Upon further inspection of Figure 3B, we noted a decreasing trend in the values of the Bland-Altman plot. By convention, the mean of actigraphy and PSG serves as our estimate of WASO. Therefore, a negative trend in the Bland-Altman plot suggests that as the participants' estimated true WASO values increase, the more the actigraphy device will underestimate that true WASO.

Because the agreement of actigraphy and PSG varies as mean WASO varies, we constructed a regression-based limits of agreement analysis to model the change in bias as a function of average WASO. The 2-part spline regression model allows the relationship between the bias and mean WASO to have different slopes over different values of mean WASO. The change-point of the linear spline was placed at 30 min of mean WASO. The Bland-Altman plot in Figure 3B shows the 2-part spline

regression line. The estimated intercept of the piecewise linear spline was 12.5 min, which defines the estimated bias when the mean WASO is zero. This estimate suggests that for mean WASO values close to zero, the actigraphy device overestimates PSG WASO. The estimate regression slope for the segment of mean WASO from 0 to 30 min is -0.33 min (P = 0.592) and for mean WASO > 30 min is -0.93 (P < 0.0001). Between 0-30 min, the rate of decline in bias is not as pronounced as it is for mean WASO > 30 min. The Bland-Altman plot highlights the observation that for mean WASO between 0-30 min, the actigraphy device has minimal positive bias in estimating WASO. The differences between actigraphy WASO and PSG WASO increase as WASO increases. For WASO > 30 min, actigraphy underestimates PSG by 0.93 min for every minute increase in WASO.

DISCUSSION

In this study we compare the technique for measuring the presence of sleep or wake by actigraphy, compared to the accepted gold standard of polysomnography. This study characterized the participant-specific variation of accuracy, sensitivity, and specificity values. Overall, the high accuracy (86%) suggests that wrist actigraphy is a reasonable technique for measuring sleep. The density plot (Figure 1) shows that sensitivity of actigraphy to correctly detect sleep does not vary much between participants and is above 90% for every participant. The largest difference between participants in terms of performance of the actigraphy device is through the specificity of actigraphy (mean 33%). The distribution of specificity is moderately flat and highly variable across individuals, suggesting difficulty with detecting wake patterns between individuals. Overall, the participant-specific accuracy is relatively high, and for most participants, above 80%. We validate this finding across multiple nights and a variety of patient populations. We conclude that actigraphy is overall a useful and valid means for estimating total sleep time and WASO in clinical, field and workplace studies, with some limitations in specificity.

This study validates and determines the empirical bounds for an existing and widely used actigraphy analysis algorithm in adults across much of the young to midlife years, in many participants, both men and women, in those with and without insomnia, and across one to several nights. This study also avoids the difficulties of some previous reports using a convenience sample from a sleep clinic in which participants with sleep disordered breathing, arousal-related, and potentially movement-related sleeping disorders are overrepresented. Unlike prior studies, the analyses spanned both sleep-wake states on an epoch-by-epoch basis, and assessed WASO over the entire sleep period using Bland-Altman methods for assessing bias of the WASO estimate as it varies by the amount of WASO recorded. The use of a novel temporal alignment strategy ensured that estimates were not compromised by relative computer timing differences that are independent of the validity of the actigraphy algorithm itself to for determining sleep and wake states. The use of longer sleep periods here than most previous studies (8.5 h time in bed being "the new 8 hours") yielded more time for wake during the analyzed recording interval and therefore an increased WASO.

This study compared actigraphy to PSG across several participant characteristics. In multivariable analyses, gender and chronic primary insomnia slightly modified the performance of actigraphy. As expected, the specificity of actigraphy versus PSG in insomnia patients with high sleep efficiency was higher than the specificity for insomnia patients with low sleep efficiency. As with other contrasts among groups with differing sleep efficiency, it is important to note that sleep periods were longer than typical, mostly 8.5 hours time in bed, with some healthy subjects 10 hours of time in bed, allowing more time for quiet wakefulness in bed, when actigraphy is weakest. As previously observed,[9,10] specificity is highest in healthy younger participants with nocturnal sleep, an effect that is further increased by shortening the sleep period to maximize sleep efficiency.

The effect of age on actigraphy sensitivity was linear but demonstrated a minimal effect on accuracy and specificity as age increased from young adult through midlife. As sleep efficiency declines with age, accuracy is slightly reduced in the context of a greater proportion of wake, which reveals the primary weakness of the actigraphy algorithm (specificity). In older adult women (mean age 69 years) with insomnia, actigraphy accuracy falls below 80% as PSG-measured sleep efficiency falls below 73%.[25]

In sleep parameter concordance analyses, we noted a moderately positive and statistically significant correlation between actigraphy and PSG WASO. The Bland-Altman plot suggests that actigraphy tended to overestimate PSG WASO by an average of about 5 minutes for WASO nights of 30 minutes or less. For mean WASO more than 30 minutes, actigraphy begins to underestimate PSG WASO. The underestimation of WASO time could be due to the limited ability of the wrist actigraphy device to correctly identify immobility as part of the wake state during the sleep period. This observation was not modified by age or insomnia in multivariate analyses. A similar result from a study of actigraphy versus PSG in insomnia patients using a Pearson correlation method (not Bland-Altman) observed an underestimation of WASO by actigraphy from about 50 minutes and greater, with no effect of age or sex of the participant.[26]

The results of the study highlight the potential of actigraphy to measure sleep in numerous research and clinical circumstances. Clearly, shifting realities with regard to third party reimbursements are driving use of home study equipment. For instance, actigraphy might serve as an important objective, repeated, accessible measure of sleep in the home for those patients with insomnia or disorders of circadian function. In both instances, actigraphy might serve as a vehicle to provide feedback to the patient and his or her care provider showing the overall pattern of sleep-wake behavior. This information might, in turn, facilitate treatment decisions related to medications or cognitive-behavioral therapy. (For a review, see Avidan 2006.[27])

Lauderdale et al. report that there is a significant discrepancy in self-reported and measured sleep duration.[31] Inadequate self-assessment of sleep duration has been associated with several negative health-related outcomes including obesity, diabetes, hypertension and mortality.[28-30] Actigraphy might provide a more reliable estimate of sleep/wake patterns for longitudinal studies than self-reported sleep. Were it to, actigraphy might serve as an objective measure for people undergoing treatments for insomnia, such as cognitive-behavioral therapy for insomnia.

Limitations of This Study

This study was conducted in the laboratory and not home settings. It is possible, however unlikely, that actigraphy behaves differently in a home setting. The 77 subjects studied represent a diverse population across a wide range of adult ages, but are not strictly a generalizable population. Specificity is a particular weakness of actigraphy when measured on an epoch-by-epoch basis, yet affords a reasonable empirical classification of WASO across the entire sleep period. Although we are encouraged by the high accuracy finding in our data (86.3%), we acknowledge that accuracy reveals optimal information when the event in question is as likely as not to occur. Accuracy is effectively a weighted linear function of sensitivity and specificity, such that both specificity and sensitivity play an important role in the overall accuracy of the device. However, we note that most of the study sleep period is occupied by sleep, thus the high accuracy of actigraphy is largely explained by high sensitivity. An alternative research strategy to actigraphy using interactive behavioral response monitoring in poor sleepers is available to detect wake during the sleep period with a higher specificity.[8] Other studies in subjects with high sleep efficiency during shorter, nighttime sleep periods (and thus with minimal wake epochs) have observed much higher specificity.[9,10] The estimates of the sensitivity, specificity, and accuracy of actigraphy cover only the sleep interval, not the entire 24-hour day. Future studies are needed to determine whether wrist actigraphy can be used to derive a 24-h algorithm defining sleep and wake patterns which would provide useful information about a person's voluntary (or involuntary) napping during the daytime. This information would better equip researchers and clinicians to develop insights about individuals' overall sleep behaviors and relate them to safety, which is especially critical in certain occupations.

CONCLUSION

Through the present analyses, we conclude that wrist actigraphy with current algorithms is of value for individual-level estimates of both sleep duration and wakefulness after sleep onset. To increase participation rates and generalizability of results in large population-level research and longitudinal design studies, there is a strong need for ecologically valid home sleep technology that imposes only a limited burden on research participants. Future studies should be undertaken to expand the validated use of actigraphy to additional populations; to full day recording; to home use; to examination of sleep as related to important health outcomes predictors such as physical activity; and to assess the impact on sleep of environmental factors such as noise, light, temperature, and media use.

DISCLOSURE STATEMENT

REFERENCES

1. US Department of Health and Human Services. Healthy People 2020 Objective Topic Areas and Page Numbers. 2010 2011/01/04/ [cited; 300-1]. Available from: http://healthypeople.gov/2020/topicsobjectives2020/default.aspx
2. Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W, Pollak CP. The role of actigraphy in the study of sleep and circadian rhythms. Sleep 2003;26:342-92.
3. Tryon WW. Issues of validity in actigraphic sleep assessment. Sleep 2004;27:158-65.
4. Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep 1992;15:461-9.
5. Jean-Louis G, Kripke DF, Cole RJ, Assmus JD, Langer RD. Sleep detection with an accelerometer actigraph: Comparisons with polysomnography. Phys Behav 2001;72:21-8.
6. Kripke DF, Hahn EK, Grizas AP, et al. Wrist actigraphic scoring for sleep laboratory patients: algorithm development. J Sleep Res 2010;19:612-9.
7. de Souza L, Benedito-Silva AA, Pires ML, Poyares D, Tufik S, Calil HM. Further validation of actigraphy for sleep studies. Sleep 2003;26:81-5.
8. Blood ML, Sack RL, Percy DC, Pen JC. A comparision of sleep detection by wrist actigraphy, behavioral response, and polysomnography. Sleep 1997;20:388-95.
9. Meltzer LJ, Walsh CM, Traylor J, Westin AM. Direct comparison of two new actigraphs and polysomnography in children and adolescents. Sleep 2012;35:159-66.
10. Basner M, Dinges DF, Mollicone D, et al. Mars 520-d mission simulation reveals protracted crew hypokinesis and alterations of sleep duration and timing. Proc Nat Acad Sci U S A 2013;110:2635-40.
11. Winkelman JW, Buxton OM, Jensen JE, et al. Reduced brain GABA in primary insomnia: preliminary data from 4T proton magnetic resonance spectroscopy (1H-MRS). Sleep 2008;31:1499-506.
12. Dang-Vu TT, McKinney SM, Buxton OM, Solet JM, Ellenbogen JM. Spontaneous brain rhythms predict sleep stability in the face of noise. Curr Biol 2010;20:R626-R7.
13. McKinney SM, Dang-Vu TT, Buxton OM, Solet JM, Ellenbogen JM. Covert waking brain activity reveals instantaneous sleep depth. PLoS One 2011;6:e17351.

14. Buxton OM, Ellenbogen JM, Wang W, Carballeira A, O'Connor SP, Cooper D, McKinney S, Solet JM. Sleep disruption due to hospital noises: A prospective evaluation. Ann Int Med 2012;157:170-9.

15. Buxton OM, Pavlova M, Reid EW, Wang W, Simonson DC, Adler GK. Sleep restriction for one week reduces insulin sensitivity in healthy men. Diabetes 2010;59:2126-3.

16. Buxton OM, Pavlova M, Wang W, Scheer FL, Klerman EB, O'Connor SP, Porter JH, McLaren DT, Cooper DT, Ellenbogen JM. Examining the effects of daytime eszopiclone administration on daytime sleep and nighttime wakefulness: a randomized, double-blind, placebo-controlled, crossover trial in shift workers. Sleep 2013;36(Abstract Supplement): A184.

17. Gorny SW, Spiro JR. Comparing different methodologies used in wrist actigraphy. Sleep Med 2001;2:135-43

18. Buxton OM, Cain SW, O'Connor SP, et al. Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption. Sci Trans Med 2012;4:129ra43.

19. Buxton OM, Pavlova M, O'Connor SP, Wang W, Winkelman J. Primary insomnia and glucose metabolism: changes in actigraphically derived wake after sleep onset (WASO) related to changes in glucose metabolism. Sleep 2010;33(Abstract Supplement):A196.

20. Rechtschaffen A, Kales A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects. Washington, DC: U.S. Government Printing Office, 1968.

21. Iber C, Ancoli-Israel S, Chesson AL Jr, Quan SF. The new sleep scoring manual- The evidence behind the rules. J Clin Sleep Med 2007;3:107.

22. Kulldorff M. SaTScan: Software for the spatial and space–time scan statistics, version 4.0 [computer program]. In: Information Management Services, 2003.

23. Bland JM, Altman DG. Measuring agreement in method comparison studies. Stat Methods Med Res 1999;8:135-60.

24. Sternberg MR, Hadgu A. A GEE approach to estimating sensitivity and specificity and coverage properties of the confidence intervals. Stat Med 2001;20:1529-39.

25. Taibi DM, Landis CA, Vitiello MV. Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. J Clin Sleep Med 2013;9:217-25.

26. Lichstein KL, Stone KC, Donaldson J, et al. Actigraphy validation with insomnia. Sleep 2006;29:232-9.

27. Avidan AY. Sleep and neurologic problems. Sleep Med Clin 2006;1:273-92.

28. Buxton OM, Marcelli E. Short and long sleep are positively associated with obesity, diabetes, hypertension, and cardiovascular disease among adults in the United States. Soc Sci Med 2010;71:1027-36.

29. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010;33:414-20.

30. Grandner MA, Hale L, Moore M, Patel NP. Mortality associated with short sleep duration: The evidence, the possible mechanisms, and the future. Sleep Med Rev 2010;14:191-203.

31. 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.