

SCIENTIFIC INVESTIGATIONS

Effect of wearables on sleep in healthy individuals: a randomized crossover trial and validation study

Sarah Berryhill, BS¹; Christopher J. Morton¹; Adam Dean, BA¹; Adam Berryhill, BS¹; Natalie Provencio-Dean, MS¹; Salma I. Patel, MD, MPH^{1,2}; Lauren Estep, MD^{1,2}; Daniel Combs, MD^{1,3}; Saif Mashaqi, MD^{1,2}; Lynn B. Gerald, PhD, MSPH⁴; Jerry A. Krishnan, MD, PhD⁵; Sairam Parthasarathy, MD^{1,2}

¹University of Arizona Health Sciences Center for Sleep and Circadian Sciences, University of Arizona, Tucson, Arizona; ²Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Department of Medicine, University of Arizona, Tucson, Arizona; ³Department of Pediatrics, University of Arizona, Tucson, Arizona; ⁴Asthma and Airway Diseases Research Center, University of Arizona, Tucson, Arizona; ⁵Breathe Chicago Center, Division of Pulmonary, Critical Care, Sleep, & Allergy, University of Illinois, Chicago, Illinois

Study Objectives: The purpose of this study was to determine whether a wearable sleep-tracker improves perceived sleep quality in healthy participants and to test whether wearables reliably measure sleep quantity and quality compared with polysomnography.

Methods: This study included a single-center randomized crossover trial of community-based participants without medical conditions or sleep disorders. A wearable device (WHOOP, Inc.) was used that provided feedback regarding sleep information to the participant for 1 week and maintained sleep logs versus 1 week of maintained sleep logs alone. Self-reported daily sleep behaviors were documented in sleep logs. Polysomnography was performed on 1 night when wearing the wearable. The Patient-Reported Outcomes Measurement Information System sleep disturbance sleep scale was measured at baseline, day 7 and day 14 of study participation.

Results: In 32 participants (21 women; 23.8 ± 5 years), wearables improved nighttime sleep quality (Patient-Reported Outcomes Measurement Information System sleep disturbance: B = -1.69; 95% confidence interval, -3.11 to -0.27; P = .021) after adjusting for age, sex, baseline, and order effect. There was a small increase in self-reported daytime naps when wearing the device (B = 3.2; SE, 1.4; P = .023), but total daily sleep remained unchanged (P = .43). The wearable had low bias (13.8 minutes) and precision (17.8 minutes) errors for measuring sleep duration and measured dream sleep and slow wave sleep accurately (intraclass coefficient, 0.74 ± 0.28 and 0.85 ± 0.15, respectively). Bias and precision error for heart rate (bias, -0.17%; precision, 1.5%) and respiratory rate (bias, 1.8%; precision, 6.7%) were very low compared with that measured by electrocardiogram and inductance plethysmography during polysomnography.

Conclusions: In healthy people, wearables can improve sleep quality and accurately measure sleep and cardiorespiratory variables.

Clinical Trial Registration: Registry: [ClinicalTrials.gov](https://clinicaltrials.gov); Name: Assessment of Sleep by WHOOP in Ambulatory Subjects; Identifier: NCT03692195.

Keywords: sleep, sleep loss, sleep quality, sleep tracker, wearable

Citation: Berryhill S, Morton CJ, Dean A, et al. Effect of wearables on sleep in health individuals: a randomized crossover trial and validation study. *J Clin Sleep Med*. 2020;16(5):775–783.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Wearables that measure sleep are increasingly popular, but whether such devices modify sleep behaviors is unclear. Current consensus is that limited validation of wearables against gold standard measurements is a major limitation for large-scale use of wearables in sleep research.

Study Impact: In healthy people, wearables can improve sleep quality and modify sleep behaviors while accurately measuring sleep variables, respiratory, and heart rate. Accurate cloud-based remote monitoring of sleep and cardiorespiratory variables is feasible and could facilitate sleep and cardiorespiratory research.

INTRODUCTION

Wearable technology for sleep assessment is rapidly becoming one of the most popular consumer health products. Wearable technologies are used for the purpose of self-guided physical activity, sleep monitoring, sleep management, and behavioral change.^{1,2} Wearable technologies including sensors placed directly on the body (wrist, chest, hip) or sensors embedded in clothing or accessories (eg, bracelet, watch, pendant) are becoming widely available to measure both sleep and physical activity.^{1,3–7} Such technology adoption by the public is further amplified by greater public awareness of the health-promoting

aspects of sleep and potential adverse consequences of poor sleep quality and reduced sleep quantity.⁸ However, there is a lack of understanding as to whether wearing such devices changes sleep quality or quantity. Moreover, current consensus suggests that limited validation of wearables against gold standard measurements is a major limitation for large-scale use of wearables in sleep research.^{9,10}

Some of the validation studies have shown high variability in commercially available wearables compared with validated accelerometers that have been used to measure sleep.¹¹ Other validation studies have compared the performance of wearables against the gold standard polysomnography (PSG) and found

them to demonstrate acceptable sensitivity but poor specificity for measuring sleep.^{12,13} More importantly, besides measurement of sleep or physical activity, these wearables are often coupled with an application on a smartphone or electronic tablet that incorporates elements of behavior change techniques that offer guidance and support for increasing and sustaining greater physical activity and health-focused tracking. Whether such devices affect modification of sleep behavior in healthy individuals, and therefore sleep quality and sleep quantity, and their downstream effects on health promotion and health outcomes is unclear. Such information is vital considering the epidemic of sleep deprivation and the adverse health and safety consequences to populations.^{14–18} There is potential for wearables to change behavior that improves sleep quality and quantity and favorably influence population health. To our knowledge, there have not been any clinical trials that have rigorously studied the behavioral effect of wearables on sleep quality and sleep quantity in healthy individuals.

The overarching aims of the proposed study were to study the effect of a wrist-worn wearable device on sleep perception and perform a methodologic study to validate the accuracy of the wrist-worn wearable device to measure sleep quality and sleep quantity compared with the gold standard PSG in healthy volunteers without self-reported sleep disorders or chronic medical conditions.

METHODS

Participants

Healthy participants without recent hospitalizations were recruited through flyers, social media, and advertisements from the community. Selection criteria were as follows. Inclusion criteria included the following: age ≥ 18 years and ≤ 45 years; ability to provide informed consent; and willingness to undergo PSG study and wear the device (Strap 2.0; WHOOP, Inc., Boston, Massachusetts) that measures sleep. Exclusion criteria included the following: presence of an untreated sleep disorder that requires diagnostic testing and treatment (insomnia, obstructive sleep apnea, narcolepsy, restless leg syndrome, rapid eye movement sleep behavior disorder, or circadian rhythm sleep disorders); the sleep disorder may be discovered during screening with the sleep questionnaire (screen failure) or following the performance of the PSG, in which case the participant's data would not be included in the analysis; apnea-hypopnea index ≥ 15 events/h as per guidelines¹⁹; active substance abuse or alcoholism; pregnancy or lactation; consuming sedative medications; participant-reported chronic medical conditions such as hypertension, diabetes mellitus, cardiac disorders, arthritis, or other chronic medical conditions; and body mass index > 26 kg/m². The study was approved by the institutional review board of the University of Arizona (IRB#1808871454), and all participants underwent the informed consent process with written consent obtained before study participation started.

Study design

This was a single-site randomized crossover study with an embedded validation night at the midpoint. Participants were randomized to (1) using the wearable device for 7 days and

maintaining sleep-wakefulness logs for the same 7 days (treatment condition) or (2) maintaining the sleep logs only (control condition). After PSG, participants were crossed over to the other arm. In the original version of the protocol, an actigraph (accelerometer) was to be worn during the control condition. However, before the first participant enrollment, the protocol was modified to require maintenance of sleep logs only (without the need to wear actigraphs) because of concerns that wearing the actigraph may modify sleep behaviors.

Participants were asked to complete the Patient-Reported Outcomes Measurement Information System (PROMIS) sleep disturbance short form questionnaire at baseline and 7 and 14 days (primary outcome) to determine the effect of wearing the device on nighttime sleep disturbance (sleep quality). The PROMIS sleep disturbance 8-item short form questionnaire has been correlated strongly with the longer forms and has greater measurement precision than the Pittsburgh Sleep Quality Index, despite having fewer total items, and has been recommended for use in research and clinical settings.^{20,21} The PROMIS sleep disturbance short form assesses the pure domain of sleep disturbance in individuals age 18 years and older. Each item asks the participant to rate the severity of their sleep disturbance during the prior 7 days. Each item on the measure was rated on a 5-point scale (1 = never; 2 = rarely; 3 = sometimes; 4 = often; 5 = always), with a range in score from 8 to 40, with higher scores indicating greater severity of sleep disturbance. The raw scores on the 8 items were then summed to obtain a total raw score, and the corresponding *t*-score based on population norms was derived.

During the entire 14 days, participants maintained sleep logs as a measure of self-report sleep-wakefulness behaviors.²² Secondary endpoints were: (a) sleep duration by sleep logs; (b) sleep duration by wearable; (c) sleep fragmentation (wake after sleep onset); and (d) heart rate variability measured by the wearable.

For the validation part of the study, the following measurements derived from the wearable device were compared against the gold standard PSG-derived measures on a single night performed at the midpoint of study participation: sleep quantity (sleep duration); sleep fragmentation; proportion of night spent during light sleep (stages N1 and N2), slow wave sleep (N3), and rapid eye movement sleep (R sleep); (d) sleep-wakefulness state determination by collapsing all stages of sleep (N1, N2, N3, and R) and wakefulness and comparing such measures against the gold standard PSG derived sleep-wakefulness state; heart rate variability measured by the wearable device versus that derived from the electrocardiogram collected as part of the PSG; and heart rate and respiratory rate measured by wearable device versus that measured as part of the PSG with respiratory inductance plethysmography.

Study procedures

Wearable device

The wearable was programmed and worn on the nondominant arm. Participants were randomly assigned to receive the WHOOP Strap 2.0 (WHOOP Inc. Boston, MA) in the first week or second week of participation, with the PSG performed at the

midpoint of the 14-day participation. The device was recovered the morning after day 14. The device transmitted data daily to the participant's smartphone or wireless internet-enabled tablet, and from there the data were transmitted to a cloud platform. Data were downloaded from the WHOOP cloud platform for analysis. During the 14 days, participants were instructed to wear the device on either the first or second week as per the randomization schedule. The application on their smartphone or smart tablet would give them information regarding their sleep and physical activity performance on a daily basis, with instructions that stated that their sleep was adequate or inadequate. The device measures heart rate information using reflectance photoplethysmography and motion with a three-axis accelerometer and processes these signals using algorithms that generate sleep and activity data. The physical activity data are measured as "physical strain," which is a measurement on a 21-point scale that, unlike counting steps, is a personalized account of exertion and fitness based on duration of time one spends in their personal maximal heart rate zones. During setup, the device requires age, sex, and anthropometric measurements to calculate the maximum heart rate zones.

PSG sleep study

Participants underwent a video-assisted polysomnography for an 8-hour period that included electroencephalography (C4-A1, C3-A2, F4-A1, F3-A2, O1-A2, and O2-A1), left and right electrooculograms, submental electromyogram, electrocardiogram, chest and abdominal movement by inductance plethysmography (Ambulatory Monitoring Inc., Ardsley, New York), leg movements by bilateral anterior tibialis electromyograms, nasal pressure and thermistor recordings for airflow, and finger pulse-oximetry (Sandman, Ontario, California or Grass Systems, Inc.; Natus, Inc.). An infrared camera was used to collect continuous video recording that was recorded synchronous to the polysomnogram. Continuous video-synchronized audio recordings were made by an acoustic (calibrated) microphone to measure snoring, and the signal was recorded digitally as both a video-synchronized audio signal and as a channel in the digitized PSG. The motion of the rib cage and abdomen was measured noninvasively using respiratory inductive plethysmography. Electroencephalography, right and left oculography, submental electromyography, and electrocardiography were amplified, filtered (Sandman, Ontario, California; or Alice5 system; Philips-Respironics, Inc., Murrysville, Pennsylvania) and recorded along with other signals and stored in corresponding data acquisition systems. In each participant, the PSG was performed during an 8-hour period at the midpoint of study participation. All PSGs were performed between 2200 and 0600 hours, with a time window of ± 2 hours to suit individual participant needs for delayed or early bedtime. Participants needed an additional 2 hours for getting ready for the study, instrumentation, and filling out pre- and postsleep questionnaires.

Time synchronization of wearable and PSG

Before the start of recordings, the participants were asked to assist with performing bio-calibrations to ensure that the recordings were working well. The entire bio-calibration took 3–4 minutes. The wearable and PSG recordings were synchronized

by setting the clocks on the PSG monitor to the same as that on the mobile devices that synchronized with the wearable device.

Scoring

Two observers (each with 20 years of experience in analyzing PSGs) scored each PSG while blinded to the other observer's scores. The entire 8-hour PSG recordings that were artifact free were analyzed using the American Academy of Sleep Medicine guidelines.^{19,23} Scoring of respiratory events in the PSG was performed according to American Academy of Sleep Medicine scoring rules for obstructive apneas and hypopneas to yield an apnea-hypopnea index value.²³ An apnea-hypopnea index value > 15 events/h disqualified the participant from continuing in the study; however, there were no participants who failed the screening for that reason.

Data analysis

To determine the effect of the wearable on perception of sleep quality, we compared changes in baseline and 7- and 14-day PROMIS sleep disturbance scores in a generalized linear mixed model (SPSS v25.0; IBM, Armonk, New York) with (7-day) time periods of wearing or not wearing the device as the determinant variable and PROMIS sleep disturbances scores as the outcome variable adjusted for age, sex, order, and baseline PROMIS sleep disturbances scores. To determine the accuracy of the wearable to measure sleep variables, we performed comparisons of the device and PSG for measuring sleep duration and various sleep stages on the night of day 7 of study participation. We performed interclass correlation coefficient (ICC) determination for sleep stages and compared them to limits of agreement standards for interrater reliability measurements.²⁴ All analyses were performed in a blinded manner to prevent bias. $P < .05$ was significant. All data are provided as mean and standard deviations unless otherwise specified.

Randomization schedule

A randomization schedule stratified by sex was generated by the statistician using SAS PROC Plan (SAS, Cary, North Carolina) and loaded into the REDCap database. The sequence was concealed in that coordinators were unaware of the next assignment until a participant consented and randomization was performed within the REDCap database by the research coordinator.

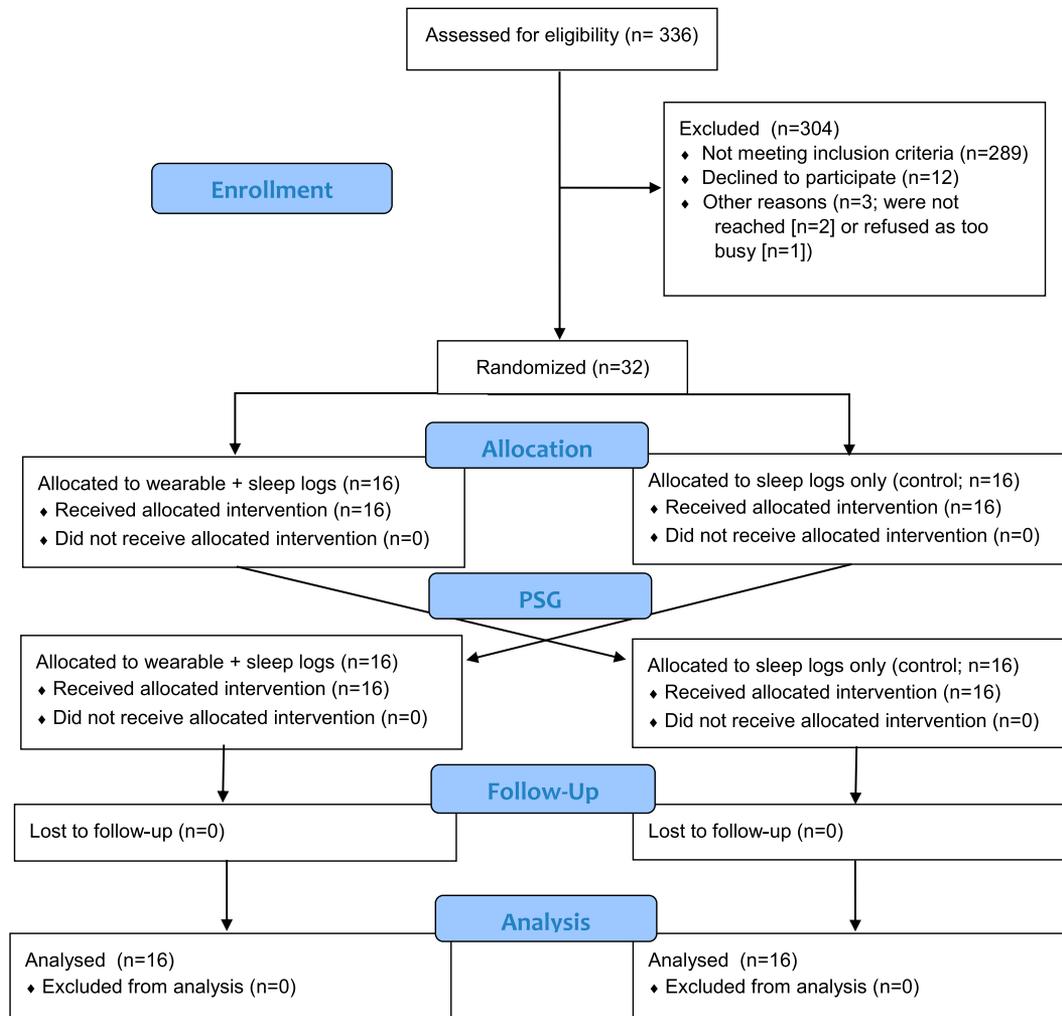
Sample size justification

To our knowledge, there were no prior studies measuring the behavioral modification of sleep habits by wearables in healthy individuals, and we therefore performed this as a preliminary study with intent to recruit 35 eligible participants.

RESULTS

In all, there were 35 eligible individuals enrolled; however, data are presented only on 32 individuals considering that 3 individuals were unable to undergo any of the study-related activities because they were too busy ($n = 1$) or

Figure 1—CONSORT diagram of flow of participants through the randomized crossover clinical trial.



PSG = polysomnography.

Table 1—Regression models for self-reported sleep disturbance.

PROMIS Sleep Disturbance (n = 32)	Contrast Estimate	SE	t	Adjusted P Value	Lower 95% CI	Upper 95% CI
Watch*	-1.69	0.71	2.38	.021‡	-3.11	-0.27
Women†	-0.15	0.74	-0.20	.84‡	-1.63	1.33
Order	-0.15	0.71	-0.21	.83	-1.58	1.27
Age	0					

*Compared with control condition in generalized linear mixed models adjusted for age, sex, and order effect. †Compared with men. ‡P < .05. CI = confidence interval, PROMIS = Patient-Reported Outcomes Measurement Information System, SE = standard error, t = t score.

were not able to be reached after giving informed consent (n = 2; **Figure 1**).²⁵

Sleep quality and quantity

PROMIS sleep disturbance short form score decreased, signifying improved nighttime sleep quality, during the week of wearing the device (intervention condition) compared with not wearing the device after adjusting for age, sex, baseline values, and order effect (**Table 1**). The unadjusted PROMIS

sleep disturbance short form score was 51.8 ± 3.0 when not wearing the device versus 50.1 ± 2.8 when wearing the device (P = .017). The change in PROMIS sleep disturbance scores was -1.43 ± 2.95 during the week when wearing the wearable versus 0.26 ± 2.63 when not wearing the device, with an adjusted difference between conditions of 1.69 ± 0.71 (SE). The adjusted reduction in PROMIS sleep disturbance score when wearing the device was greater than 0.5 SD, suggesting that such a change is meaningful. There were no sex differences or order effects noted

Table 2—Unadjusted self-reported sleep information derived from sleep logs.

Condition (n = 32)	Nighttime Sleep (min)		Naps (min)		Wake After Sleep Onset (min)		Total Sleep (min)	
	Men	Women	Men	Women	Men	Women	Men	Women
Watch	391 ± 82	435 ± 88*	19 ± 53	17 ± 48	16 ± 22	8 ± 14*	410 ± 92	452 ± 106*
	421 ± 88†		17 ± 49		11 ± 17		438 ± 103	
Control	399 ± 84	448 ± 87*	9 ± 32	11 ± 33	13 ± 27	8 ± 16*	408 ± 82	459 ± 94*
	433 ± 88		10 ± 33		9 ± 20		443 ± 93	

*In regression models, women had less wake after sleep-onset periods than men and greater nighttime and total sleep durations after adjusting for age, order, and effect of wearing the watch (see [Table 3](#)). †Compared with control condition ($P = .07$; see [Table 3](#)).

Table 3—Regression models for self-reported sleep information.

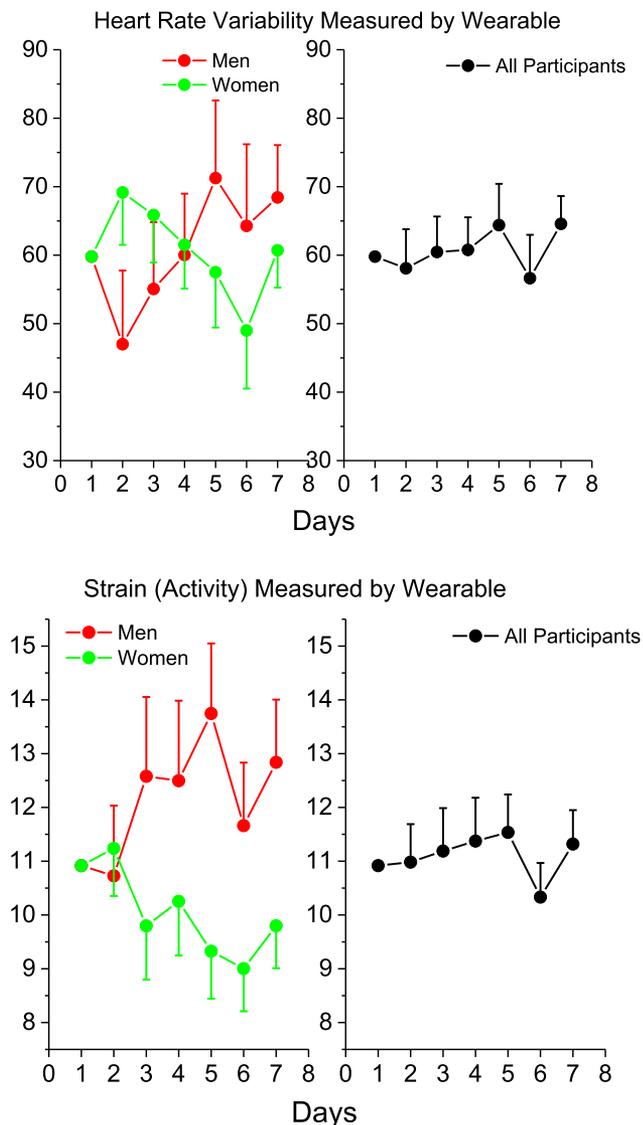
	Contrast Estimate	SE	t Value	Adjusted P Value	Lower 95% CI	Upper 95% CI
Nighttime sleep (n = 32)						
Watch*	-13.13	7.28	1.80	.07‡	-1.18	27.44
Women†	51.96	12.27	3.40	.001§	21.94	81.97
Order	-14.32	13.83	-1.04	.30	-41.5	12.87
Age	1.87	0.75	2.51¶	.012§	0.85	4.09
Total sleep in 24 hours (n = 32)						
Watch*	6.77	8.51	0.796	.43	-9.96	23.51
Women†	50.63	15.08	3.36	.001§	20.98	80.28
Order	-10.47	13.81	-0.76	.45	-37.62	16.68
Age	1.55	0.76	2.03¶	.042§	0.59	4.07
Naps (n = 32)						
Watch*	3.21	1.41	-2.29	.023§	0.45	5.98
Women†	0	0.001	0	1.0	-0.001	0.001
Order	3.21	1.41	2.29	.023§	0.45	5.98
Age	0	0	0	1.0		
WASO (n = 32)						
Watch*	1.07	0.99	-1.07	.28	-3.03	0.89
Women†	-4.71	2.30	-2.05	.041§	-9.23	-0.19
Order	2.13	2.08	1.02	.31	-1.98	6.23
Age	0.05	0.02	2.78¶	.006§	0.023	0.09

*Compared with control condition in generalized linear mixed models adjusted for age, sex, and order effect. †Compared with men. ‡ $P < .10$. § $P < .05$. ¶Z value. CI, confidence interval, SE = standard error, $t = t$ value, WASO = wake after sleep onset.

for changes in PROMIS sleep disturbance scores. Unadjusted values for nighttime sleep quantity, wake after sleep-onset duration (a measure of sleep fragmentation), time spent napping, and total sleep over a 24-hour period are provided in [Table 2](#). After adjusting for age, sex, baseline values, and order effect, the nighttime sleep duration tended to be lower during the week of wearing the device compared with the week of not wearing the device ($P = .07$; [Table 3](#)). Time spent awake after sleep onset was not different during the two conditions ([Table 3](#)). However, the duration spent napping was slightly but statistically greater when wearing the device in contrast to when not wearing the device ([Table 3](#)). In our study, women participants reported greater time spent asleep at night and over a 24-hour period of time than male

participants. Also, they tended to have lower wake after sleep-onset time periods.

Strain (a measure of physical activity) increased for the entire group over the course of the week when wearing the device ($P = .01$; [Figure 2](#)), but there was a noticeable sex difference between greater increase in men and a tendency for reduced strain in women ([Figure 2](#)). Such measurements were unavailable for the week when the device was not worn. Heart rate variability increased over the course of the week when the device was worn, with no noticeable sex difference ([Figure 2](#)). Total sleep duration measured objectively by the wearable device over a 24-hour period was unchanged over the course of the week, with no differences between the sexes ($P = .4$).

Figure 2—Heart rate variability and strain.

Heart rate variability (top) and strain (a measure of physical activity derived from heart rate; bottom) are shown for women (green symbols) and men (red symbols). During the 7 days of wearing the device, heart rate variability measured by the device increased over time (top right; black symbols; $P = .01$), and there were no sex differences, whereas physical activity increased in men but not in women (bottom).

Device validation

The bias and precision errors for measuring nighttime sleep duration (sleep quantity) and sleep fragmentation (sleep quality) by the wearable compared with the PSG were low (Table 4). Similarly, the accuracy of wearable-based measurements of heart rate, respiratory rate, and heart rate variability were excellent compared with the gold-standard PSG (Table 4). The ICC for scoring the various sleep stages between the two experienced blinded scorers was excellent (ICC, 0.91 ± 0.05). In contrast, the ICC between the wearable and the consensus scores of the blinded expert scorers was good (0.67 ± 0.15 [SD]; $n = 32$). However, the ICC was excellent for dream sleep

(0.85 ± 0.15) and good for slow wave sleep (0.74 ± 0.28). The ICC for light nondream sleep was fair (0.63 ± 0.15).

DISCUSSION

Sleep quality and quantity

Our study found that a wrist-worn wearable device resulted in improved perception of sleep quality (reduced sleep disturbance) in healthy volunteers. The mean reduction in PROMIS sleep disturbance score was 1.69, which was greater than the 0.5 SD of the change in PROMIS sleep disturbance score, suggesting that such a change is of moderate effect size and meaningful.^{26,27} By design, we targeted healthy participants, considering that these participants are representative of the general population who are increasingly adopting such wearables to monitor sleep. Therefore, such healthy participants are less likely to have significant deviation of their PROMIS scores from 50 (specifically, $51.8 + 3.0$). This is because the PROMIS t -scores are based on a score of 50, representing the mean of the general population reference sample plus clinical sample. Despite our sample having PROMIS scores close to the population mean, our study found that a wrist-worn wearable device resulted in improved perception of sleep quality (reduced sleep disturbance) in healthy volunteers and a mean reduction in the PROMIS sleep disturbance score of 1.69. Such a difference is greater than the half (0.5) the standard deviation of the PROMIS sleep disturbance score of 2.9, suggesting that such a change is of moderate effect size and is a meaningful change.

The underlying mechanism for the observed improvement in sleep quality (or reduction in sleep disturbance) is uncertain. Conceivably, an improvement in sleep quality may have been caused by the reduction in nighttime sleep duration, which by restricting nighttime sleep duration, improved self-reported sleep quality. Such findings are well described in individuals with insomnia, with sleep restriction comprising an important component of cognitive behavioral therapy for insomnia.²⁸ Such reductions in sleep duration may have been facilitated by the mobile phone application that is synchronized with the wearable device and provides instructions to the participant the next morning to advise them to increase or decrease their sleep duration based on their recent sleep-wakefulness behaviors. Although a placebo effect of the wearable on sleep quality is feasible, the improvement of sleep quality accompanied by sleep restriction supports a biological effect.

An alternative explanation for the observed improvement in sleep quality may be that, when wearing the device, the participants may have increased the degree of physical activity, which, in turn, could have improved sleep quality. Prior studies have shown that activity trackers combined with texting information may increase physical activity compared with monitoring activity alone.²⁹ Although in our study there was no such messaging occurring, the participants got daily feedback from the smartphone application regarding their previous day and week of physical activity and that may have promoted them to become more active when wearing the device. Such increased physical activity, in turn, has been shown to improve nighttime sleep quality.³⁰ However, others have shown that wearables that

Table 4—Validation of wearable against the gold standard polysomnography for sleep and physiologic measurements.

Physiologic Measure (n = 32)	Measurement Modality		Bias Error (Absolute Values)	Precision Error (Absolute Values)
	PSG	Wearable		
Sleep duration	5.3 ± 1.1 h	5.53 ± 1.0 h	4.16% (13.8 min)	5.38% (17.8 min)
REM sleep duration	0.95 ± 0.37 h (17.9% of sleep)	0.94 ± 0.43 h (16.9% of sleep)	0.01% (0.6 min)	6.69% (4.4 min)
NREM sleep duration	4.35 ± 1.3 hours (82.1% of sleep)	4.59 ± 1.2 hours (83.1% of sleep)	5.5% (14.4 min)	7.8% (22 min)
Sleep fragmentation (events/h)	1.2 ± 2.0	1.6 ± 0.9	0.48	2.4
Heart rate (beats/min)	66.0 ± 9.9	65.9 ± 9.8	-0.17% (-0.15)	1.6% (1.0)
Respiratory rate (breaths/min)	15.6 ± 1.7	15.7 ± 1.7	1.8% (0.1)	6.7% (1.0)
Heart rate variability (ms)	57.6 ± 24.6	61.7 ± 30.2	8.4% (4.8)	9.7% (6.3)

Values are mean ± SD. NREM = non-rapid eye movement, PSG = polysomnography, REM = rapid eye movement.

track physical activity do not change sedentary behavior.³¹ Nevertheless, in our study, as a group, physical activity (measured as strain by the device) based on increments in heart rate increased over the course of the 7 days of wearing the device. The application algorithms and notifications may have played a role in such observed increase in physical activity in our study. Such findings are in keeping with the observed improvements in heart rate variability over the 7-day period of wearing the device (Figure 2) and the known positive association between increased physical activity and heart rate variability.³² There were strong sex differences with greater increments in physical activity in men compared with women over the 7 days of wearing the device. If such physical activity was indeed contributing to the observed improvement in self-reported sleep quality, then sex differences in self-reported sleep quality should have probably been observed as well. Such sex differences in sleep quality were not evident in our analyses (Table 2).

The observed sex differences in greater sleep duration in young women compared with men was intriguing. A review of the literature suggests that, unlike older women who have reduced sleep duration compared with men, younger women report greater sleep duration than young men. In a recent study involving 17,355 participants, Kuula et al³³ reported that sleep duration (measured by accelerometers) was greater in young women than young men. A possible explanation is the earlier onset and offset of puberty in women compared with men and the observed earlier midpoint of sleep in women in relation to men. Regardless of the underlying mechanisms, the observed sex differences in sleep duration are externally valid.

Limitations

Our study has several limitations. The long-term effects of wearables on sleep were not studied. Moreover, although our study is generalizable to community-dwelling healthy young individuals, it is not applicable to individuals with chronic medical conditions or sleep disorders who were specifically excluded.

Validation

Our methodologic validation study revealed that the wearable device was accurate in measuring sleep quantity compared with the gold standard PSG in healthy volunteers. The observed differences were small, with low bias and precision errors. Such findings can enable population-health management considering that the data are cloud based and accessible centrally to health coaches who can then tailor and implement interventions aimed at abrogating sleep loss and poor sleep quality. Both sleep loss and poor sleep quality have been associated with risk for obesity, diabetes mellitus, adverse cardiovascular consequences, and even death.³⁴⁻⁴³

The accuracy of wearable devices in measuring sleep has a lot of variability in the published literature. Although some have shown unacceptably high variability in the accuracy of wearables compared with validated accelerometers that have been used to measure sleep,¹¹ others have demonstrated results comparable to ours.¹³ These validation studies are compounded by rapid changes in technology from the same manufacturer and updates to their software that prevent validation that is immediately generalizable to the real world.⁴⁴ Considering the explosion of wearables and mobile health (mHealth) applications aimed at improving sleep and detecting sleep disorders, recent position statements from professional societies call for rigorous testing of such devices and software against current gold standards and emphasize that, although powerful, these tools are not substitutes for medical evaluation.⁹

Dream and slow wave sleep are important sleep stages that are measured by PSG with good to excellent interobserver reliability. However, the accepted ICC is still between 0.68 and 0.82, and the ICC is even lower for nondream sleep (nonrapid eye movement [N1 and N2] or light sleep).^{23,45} To account for such inter- and intrarater reliability issues with the PSG gold standard, we considered only sleep stages that were unequivocally scored as a particular sleep stage by 2 different expert observers in a manner similar to other validation studies.⁴⁶

CONCLUSIONS

Despite improvements in wearables for measuring sleep stages and sleep fragmentation, such technologies are not accepted as

tools for measuring sleep duration in clinical practices where sleep logs and PSGs are still in use. The regulatory oversight of wearables and mHealth has many uncertainties, and therefore, more validation and scientific assessments of the effects of such devices on behaviors are needed.⁹ Our study is highly responsive to such a need and demonstrates that wearables can improve sleep quality and modify sleep behaviors while accurately measuring sleep in healthy individuals.

ABBREVIATIONS

ICC, interclass correlation coefficient
 PROMIS, Patient-Reported Outcomes Measurement Information System
 PSG, polysomnography

REFERENCES

- Ko PR, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. *J Clin Sleep Med*. 2015;11(12):1455–1461.
- DiClemente CC, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. *Am J Health Behav*. 2001;25(3):217–227.
- Miller JD, Najafi B, Armstrong DG. Current standards and advances in diabetic ulcer prevention and elderly fall prevention using wearable technology. *Curr Geriatr Rep*. 2015;4(3):249–256.
- Aminian K, Najafi B. Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications. *Comput Animat Virtual Worlds*. 2004;15(2):79–94.
- Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. *Sensors (Basel)*. 2016;16(5):646.
- Razjouyan J, Lee H, Parthasarathy S, Mohler J, Sharafkhaneh A, Najafi B. Improving sleep quality assessment using wearable sensors by including information from postural/sleep position changes and body acceleration: a comparison of chest-worn sensors, wrist actigraphy, and polysomnography. *J Clin Sleep Med*. 2017;13(11):1301–1310.
- Razjouyan J, Grewal GS, Rishel C, Parthasarathy S, Mohler J, Najafi B. Activity monitoring and heart rate variability as indicators of fall risk: proof-of-concept for application of wearable sensors in the acute care setting. *J Gerontol Nurs*. 2017;43(7):53–62.
- Consensus Conference P, Watson NF, Badr MS, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *J Clin Sleep Med*. 2015;11(6):591–592.
- Khosla S, Deak MC, Gault D, et al. Consumer sleep technology: an American Academy of Sleep Medicine Position Statement. *J Clin Sleep Med*. 2018;14(5):877–880.
- Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*. 2019; 43(2):zsz254.
- Kubala AG, Barone Gibbs B, Buysse DJ, Patel SR, Hall MH, Kline CE. Field-based measurement of sleep: agreement between six commercial activity monitors and a validated accelerometer (published online ahead of print 27 Aug 2019). *Behav Sleep Med*.
- Moreno-Pino F, Porras-Segovia A, Lopez-Esteban P, Artes A, Baca-Garcia E. Validation of Fitbit Charge 2 and Fitbit Alta HR against polysomnography for assessing sleep in adults with obstructive sleep apnea. *J Clin Sleep Med*. 2019;15(11):1645–1653.
- Lee XK, Chee N, Ong JL, et al. Validation of a consumer sleep wearable device with actigraphy and polysomnography in adolescents across sleep opportunity manipulations. *J Clin Sleep Med*. 2019;15(9):1337–1346.
- Zhang X, Holt JB, Lu H, et al. Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *Am J Epidemiol*. 2014;179(8):1025–1033.
- Wheaton AG, Olsen EO, Miller GF. Sleep duration and injury-related risk behaviors among high school students—United States, 2007–2013. *MMWR Morb Mortal Wkly Rep*. 2016;65(13):337–341.
- Lockley SW, Barger LK, Ayas NT, et al. Effects of health care provider work hours and sleep deprivation on safety and performance. *Jt Comm J Qual Patient Saf*. 2007;33(11 suppl):7–18.
- Mountain SA, Quon BS, Dodek P, Sharpe R, Ayas NT. The impact of housestaff fatigue on occupational and patient safety. *Lung*. 2007;185(4):203–209.
- Sigurdson K, Ayas NT. The public health and safety consequences of sleep disorders. *Can J Physiol Pharmacol*. 2007;85(1):179–183.
- Berry RB, Brooks R, Gamaldo C, et al. AASM Scoring Manual updates for 2017 (version 2.4). *J Clin Sleep Med*. 2017;13(5):665–666.
- Yu L, Buysse DJ, Germain A, et al. Development of short forms from the PROMIS sleep disturbance and sleep-related impairment item banks. *Behav Sleep Med*. 2011;10(1):6–24.
- Arnedt JT. PROMIS of improved tools for assessing sleep and wake function: commentary on “Development of short forms from the PROMIS sleep disturbance and sleep-related impairment item banks” *Behav Sleep Med*. 2011;10(1):25–27.
- Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*. 2012;35(2):287–302.
- Silber MH, Ancoli-Israel S, Bonnet MH, et al. The visual scoring of sleep in adults. *J Clin Sleep Med*. 2007;3(2):121–131.
- Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scoring reliability program: sleep stage scoring. *J Clin Sleep Med*. 2013;9(1):81–87.
- Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med*. 2010;152(11):726–732.
- Health Measures. Transforming how health is measured. <http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/meaningful-change>. Accessed October 13, 2019.
- Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care*. 2003;41(5):582–592.
- Falloon K, Elley CR, Fernando A 3rd, Lee AC, Arroll B. Simplified sleep restriction for insomnia in general practice: a randomised controlled trial. *Br J Gen Pract*. 2015;65(637):e508–e515.
- Martin SS, Feldman DI, Blumenthal RS, et al. mActive: a randomized clinical trial of an automated mHealth intervention for physical activity promotion. *J Am Heart Assoc*. 2015;4(11):e002239.
- Hartescu I, Morgan K, Stevinson CD. Increased physical activity improves sleep and mood outcomes in inactive people with insomnia: a randomized controlled trial. *J Sleep Res*. 2015;24(5):526–534.
- Sloan RA, Kim Y, Sahasranaman A, Muller-Riemenschneider F, Biddle SJH, Finkelstein EA. The influence of a consumer-wearable activity tracker on sedentary time and prolonged sedentary bouts: secondary analysis of a randomized controlled trial. *BMC Res Notes*. 2018;11(1):189.
- Soares-Miranda L, Sattelmair J, Chaves P, et al. Physical activity and heart rate variability in older adults: the Cardiovascular Health Study. *Circulation*. 2014;129(21):2100–2110.
- Kuula L, Graisar M, Martinmaki K, et al. Using big data to explore worldwide trends in objective sleep in the transition to adulthood. *Sleep Med*. 2019;62:69–76.
- Althuis MD, Fredman L, Langenberg PW, Magaziner J. The relationship between insomnia and mortality among community-dwelling older women. *J Am Geriatr Soc*. 1998;46(10):1270–1273.
- Hublin C, Partinen M, Koskenvuo M, Kaprio J. Heritability and mortality risk of insomnia-related symptoms: a genetic epidemiologic study in a population-based twin cohort. *Sleep*. 2011;34(7):957–964.

36. Kripke DF. Is insomnia associated with mortality? *Sleep*. 2011;34(5):555, author reply 557-558.
37. Kripke DF, Garfinkel L, Wingard DL, Klauber MR, Marler MR. Mortality associated with sleep duration and insomnia. *Arch Gen Psychiatry*. 2002;59(2):131-136.
38. Li Y, Zhang X, Winkelman JW, et al. The association between insomnia symptoms and mortality: a prospective study of US men. *Circulation*. 2014;129(7):737-746.
39. Nakazaki C, Noda A, Koike Y, Yamada S, Murohara T, Ozaki N. Association of insomnia and short sleep duration with atherosclerosis risk in the elderly. *Am J Hypertens*. 2012;25(11):1149-1155.
40. Parthasarathy S, Vasquez MM, Halonen M, et al. Persistent insomnia is associated with mortality risk. *Am J Med*. 2015;128(3):268-275.
41. Vgontzas AN, Liao D, Pejovic S, et al. Insomnia with short sleep duration and mortality: the Penn State cohort. *Sleep*. 2010;33(9):1159-1164.
42. De Bernardi Rodrigues AM, da Silva CC, Vasques AC, et al. Association of sleep deprivation with reduction in insulin sensitivity as assessed by the hyperglycemic clamp technique in adolescents. *JAMA Pediatr*. 2016;170(5):487.
43. Combs D, Goodwin JL, Quan SF, Morgan WJ, Parthasarathy S. Longitudinal differences in sleep duration in Hispanic and Caucasian children. *Sleep Med*. 2016;18:61-66.
44. Garber AM. Modernizing device regulation. *N Engl J Med*. 2010;362(13):1161-1163.
45. Iber C, Ancoli-Israel S, Chesson AL Jr, Quan SF; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. 1st ed. Westchester, IL: American Academy of Sleep Medicine; 2007.
46. Ambrogio C, Koebnick J, Quan SF, Ranieri M, Parthasarathy S. Assessment of sleep in ventilator-supported critically ill patients. *Sleep*. 2008;31(11):1559-1568.

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication October 31, 2019

Submitted in final revised form January 27, 2020

Accepted for publication January 28, 2020

Address correspondence to: Sairam Parthasarathy, MD, 1501 N. Campbell Avenue, UAHS Room 2342D, Tucson, AZ 86724; Tel: (520) 626-8309; Fax: (520) 626-1876, Email: spartha1@email.arizona.edu

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. Work for this study was performed at University of Arizona. This study was funded by a grant to the University of Arizona from WHOOP Inc., Boston, Massachusetts. The authors report no conflicts of interest.