Is There a Clinical Role For Smartphone Sleep Apps? Comparison of Sleep Cycle Detection by a Smartphone Application to Polysomnography

Sushanth Bhat, MD[1]; Ambra Ferraris, MD[1]; Divya Gupta, MD[1]; Mona Mozafarian, MD[1]; Vincent A. DeBari, PhD[2];
Neola Gushway-Henry, MD[1]; Satish P. Gowda, MD[1]; Peter G. Polos, MD, PhD[1]; Mitchell Rubinstein, RPSGT[1]; Huzaifa Seidu, MD[1];
Sudhansu Chokroverty, MD[1]

[1]JFK Neuroscience Institute/Seton Hall University, Edison NJ;
[2]Seton Hall University School of Health and Medical Sciences, South Orange, NJ

Study Objectives: Several inexpensive, readily available smartphone apps that claim to monitor sleep are popular among patients. However, their accuracy is unknown, which limits their widespread clinical use. We therefore conducted this study to evaluate the validity of parameters reported by one such app, the Sleep Time app (Azumio, Inc., Palo Alto, CA, USA) for iPhones.
Methods: Twenty volunteers with no previously diagnosed sleep disorders underwent in-laboratory polysomnography (PSG) while simultaneously using the app. Parameters reported by the app were then compared to those obtained by PSG. In addition, an epoch-by-epoch analysis was performed by dividing the PSG and app graph into 15-min epochs.
Results: There was no correlation between PSG and app sleep efficiency (r = −0.127, p = 0.592), light sleep percentage (r = 0.024, p = 0.921), deep sleep percentage (r = 0.181, p = 0.444) or sleep latency (rs = 0.384, p = 0.094). The app slightly and nonsignificantly overestimated sleep efficiency by 0.12% (95% confidence interval [CI] −4.9 to 5.1%, p = 0.962), significantly underestimated light sleep by 27.9% (95% CI 19.4–36.4%, p < 0.0001), significantly overestimated deep sleep by 11.1% (CI 4.7–17.4%, p = 0.008) and significantly

overestimated sleep latency by 15.6 min (CI 9.7–21.6, p < 0.0001). Epochwise comparison showed low overall accuracy (45.9%) due to poor interstage discrimination, but high accuracy in sleep-wake detection (85.9%). The app had high sensitivity but poor specificity in detecting sleep (89.9% and 50%, respectively).
Conclusions: Our study shows that the absolute parameters and sleep staging reported by the Sleep Time app (Azumio, Inc.) for iPhones correlate poorly with PSG. Further studies comparing app sleep-wake detection to actigraphy may help elucidate its potential clinical utility.
Commentary: A commentary on this article appears in this issue on page 695.
Keywords: actigraphy, apps, iPhones and sleep, mobile phones and sleep, sleep apps, sleep cycle, smartphones and sleep
Citation: Bhat S, Ferraris A, Gupta D, Mozafarian M, DeBari VA, Gushway-Henry N, Gowda SP, Polos PG, Rubinstein M, Seidu H, Chokroverty S. Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography. *J Clin Sleep Med* 2015;11(7):709–715.

Smartphones are now ubiquitous. As their technological capabilities continue to improve, both consumers and healthcare personnel are constantly finding new and innovative uses for smartphone apps in the field of health and medicine.[1] Apps have been incorporated into the practice of medical specialties as diverse as diabetology and neurosurgery, and have found a role in activities ranging from interpretation of radiology imaging to smoking cessation counseling. Sleep related concerns are commonplace as well, and given the burgeoning popularity and easy availability of inexpensive apps that purport to monitor multiple physiological parameters, it is not surprising that several apps have been designed to evaluate sleep quality.[2] The inexorable permeation of smartphones into the field of sleep medicine has resulted in the development of promising apps that screen for obstructive sleep apnea (OSA)[3,4] and periodic limb movements in sleep (PLMS).[5] Most of them do not, however, record sleep stages. It is often helpful to sleep physicians and cognitive behavioral therapists who treat patients with

BRIEF SUMMARY
Current Knowledge/Study Rationale: There are several preexisting, widely available, inexpensive smartphone apps designed to monitor sleep, but it is unclear whether they have clinical utility. Our goal was to systematically compare the results obtained by using one such app, the Sleep Time app (Azumio, Inc.) to the gold standard, polysomnography (PSG).
Study Impact: Our study shows that the absolute parameters and sleep staging reported by the Sleep Time app (Azumio, Inc.) for iPhones correlate poorly with PSG. Further studies comparing app sleep-wake detection to actigraphy may help elucidate its potential clinical utility.

insomnia, circadian rhythm disorders, and hypersomnolence to have objective data about sleep patterns in order to make diagnoses and treatment recommendations. There are several apps that allow users to report and analyze their sleep quality and duration,[6] but such data are often intrinsically subjective. In-laboratory polysomnography (PSG) is the gold standard for

sleep cycle analysis, but is labor intensive, expensive, often inaccessible, and clearly not a modality that can be used on a nightly basis. Long-term monitoring of sleep requires more practical methods. Use of actigraphy, while a viable alternative,[7–9] is limited by the need for costly, specialized equipment and may be considered cumbersome by some patients. A reliable, affordable and user-friendly smartphone-based app that can monitor sleep-wake cycles would be an invaluable addition to the arsenal at the disposal of practitioners of sleep medicine. Several commercially available apps claim to do exactly this, and are widely used by patients with sleep complaints. However, to our knowledge, they have not been validated by comparison with PSG, which is necessary to consider them viable clinical tools.[10] We therefore conducted this study to evaluate the reliability of the data provided by one of the most popular and widely downloaded of such apps, the Sleep Time app (Azumio Inc., Palo Alto, CA, USA), which, in addition to providing users with a graph depicting wakefulness, light sleep, and deep sleep after an overnight recording, has also been designed to specifically awaken the user from light sleep in the half hour before the time its alarm is set. Our goal was to determine if the results obtained by using this app installed on an iPhone would be accurate enough to be incorporated into the clinical management of sleep disorders.

METHODS

Subjects

Twenty adult subjects (40% women, ages ranging from age 22 to 57 y), with no previously diagnosed sleep disorders were enrolled between June 2013 and July 2014. Subjects were medical residents, fellows, attendings, and sleep laboratory technicians or persons known to them, and the study was publicized through word of mouth. Written informed consent was obtained and the study was approved by the Institutional Review Board at JFK Medical Center, Edison, NJ, USA. In the period during which our study was conducted, the Sleep Time app was available in free and paid versions, with the free version offering the same functionality but a limitation in the number of nights' worth of data stored and a restriction on the available music for the alarm, according to the developers' website. Subjects were asked to download the free version of the app onto their iPhones through the App Store; those who owned an iPhone used their own personal devices for the study, whereas those who did not were provided one of three iPhones by the investigators. Although not recommended for this particular app, the developers of similar apps instruct users to allow several nights' recording for the app to become acclimated to their sleep habits; therefore all subjects were required to use the app for a minimum of 5 nights at home before participating in the study. None of the subjects had previously downloaded the app onto their iPhones, and all subjects were instructed not to allow others to use the app once they had begun their minimum 5-night home use until after they had completed the study. In the case of those who used the iPhones provided by the investigators, data were purged and the app reloaded after each subject completed the study, in preparation for the next subject. Thus, the version of the app used by an individual subject depended on the latest version that had been released

by the developer at the time they were enrolled in the study; the versions mentioned on the developers' website are 2.2 and 2.10, although a detailed version release history was not currently available on the site at the time of writing. Because some subjects owned personal iPhones, many individual devices (models 4S and 5), all running iOS5.0 or newer, were used in the study.

After enrollment, all subjects completed an anonymous questionnaire that queried them about the presence of common sleep related complaints, whether they had used the alarm feature at home, whether they found their sleep more refreshing while using the app, and whether they slept alone while using the app at home.

In-laboratory Comparison Study

After they had used the app for a minimum of 5 nights at home, all subjects underwent a standard full night in-laboratory PSG study, while simultaneously using the app. All subjects slept alone during the in-laboratory PSG. The subjects' iPhone and the clock on the computer running the PSG software (and therefore the time imprinted onto the PSG) were synchronized to the minute, and "lights out" was annotated on the PSG when the patient set the alarm on the app, signifying that it had begun acquiring data. Subjects set the alarm on the app as per their wishes, and were not disturbed by the technician during the night, except to correct a major problem preventing the acquisition of an interpretable recording. In the morning, subjects awoke either spontaneously or to the alarm. "Lights on" corresponded to the subject's final awakening and was annotated on the PSG.

All PSGs were performed using hardware (Comet-PLUS XL laboratory-based PSG) and software (TWin PSG Clinical Software) developed by GRASS Technology (Natus Neurology, Inc, Warwick, RI, USA). Standard measurements included four to eight channels for electroencephalography, as well as channels for electrooculography, submental and bilateral tibial electromyography, electrocardiography, nasal airflow measurement using nasal cannulae connected to a pressure transducer (SleepSense, SLP Inc., St. Charles, IL, USA), oral airflow using an oronasal thermistor (Braebon Medical Corp., Kanata, ON, Canada), effort measurement using chest and abdominal respiratory impedance plethysmography belts (SleepSense, SLP Inc.), and arterial oxygen saturation using a pulse oximeter (SleepSense, SLP Inc.). Sleep stages on the PSG were scored in 30-sec epochs. Scoring began at "lights out". A traditional sleep latency was determined by the first epoch of any stage of sleep. In addition, in view of recent literature that suggests that patients' own perception of their sleep latency, which is more relevant to them than a sleep latency determined by rigid PSG criteria, tends to be better reflected by time taken for the onset of more consolidated sleep,[11] we calculated a sustained sleep latency for each subject (time taken in min after "lights out" for the patient to achieve and maintain sleep for a minimum of 10 consecutive min). Total recording time (TRT) was defined as the time between "lights out" and "lights on" in min. Total sleep time (TST) was defined as the total time scored as sleep on the PSG in min. Wake after sleep onset was calculated by TRT minus sleep latency and TST. Sleep efficiency was calculated as the ratio of TST to TRT, expressed as a percentage. All PSGs were reviewed and scored by board certified sleep medicine specialists, using the updated standard

2012 American Academy of Sleep Medicine criteria to score sleep stages and respiratory and movement events.[12]

Statistical Analysis

The absolute parameters provided by the app were compared to those obtained from the PSG recording. Although the app reported percentages of light sleep and deep sleep, we obtained equivalent PSG values by adding the percentages of N1 and N2 sleep, and N3 and rapid eye movement (REM) sleep, respectively. Both the app and the PSG reported sleep efficiency. All the aforementioned data were tested for fit-to-normality using the D'Agostino-Pearson test and were found to not differ significantly from normal distributions, so were correlated using Pearson product moment correlation coefficient (r). Multiple linear regression analysis was used to account, separately, for apnea-hypopnea index (AHI) and periodic limb movement index (PLMI) on PSG as covariates in the determination of the relationship between PSG sleep efficiency and app accuracy. The app did not provide an absolute value for sleep latency, but this was visually estimated from its graphical display and compared to the sleep latency obtained by PSG. However, because sleep latencies were nonnormally distributed, Spearman rank correlation measure (r_s) was used for analysis.

The app did not divide its recording into specific epochs. Therefore, to make an epoch-by-epoch comparison, both the PSG hypnogram and the sleep stage graph obtained from the app (**Figure 1**) were divided into 15-min epochs. When more than one stage was present in a given 15-min epoch, either on the PSG or on the app graph, the stage comprising the majority of the epoch was assigned to that epoch. In this manner, every 15-min epoch was assigned a PSG stage and an app stage. The epoch was designated as accurate when the app epoch stage was congruent with the PSG epoch stage, which was taken as the gold standard. We wished to evaluate the accuracy of overall app sleep-wake detection as well as discrimination between individual stages within sleep. For epochwise sleep-wake detection, the app epoch was considered to be accurate if it detected sleep when the PSG was scored as sleep (regardless of the stage of sleep reported by either modality) or if it detected wake when the PSG was scored as wake. Based on this epoch-by-epoch analysis, overall sensitivity (PSG and app both sleep), specificity (PSG and app both wake), positive predictive value, negative predictive value and accuracy (percentage of app epochs correlating with PSG epochs) were calculated. To be considered accurate while comparing individual sleep stages, the app epoch would have to be "light sleep" when the PSG epoch was N1 or N2 and "deep sleep" when the PSG epoch was N3 or REM sleep. Similar analyses were then carried out separately for each stage.

The significance of proportions of subjects being awakened from light sleep by the app was determined using the standard difference (Z) test, per the Primer of Biostatistics version 5.0.[13]

All calculations were made using Prism software (GraphPad Corp., San Diego, CA, USA), SPSS v.20 (IBM Corp., Armonk, NY) on a Windows 7/personal computer platform.

RESULTS

Baseline subject demographics and PSG characteristics are presented in **Table 1**. Subject responses to the anonymous

Figure 1
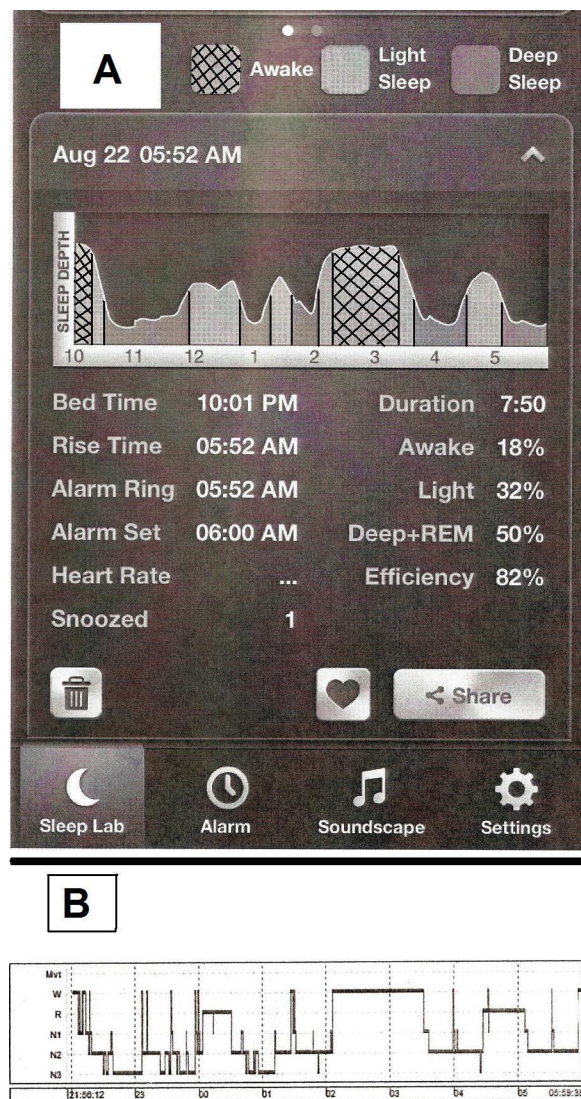
(A) A screenshot of the graphical display of a whole-night recording from a subject in the study by the Sleep Time app on an iPhone. **(B)** Hypnogram showing sleep stages recorded by simultaneous polysomnography.

questionnaire regarding sleep related complaints in general, and sleep habits and subjective sleep quality while using the app alarm for 5 nights at home are summarized in **Table 2**.

There was no correlation between PSG and app sleep efficiency (r = −0.127, p = 0.592), light sleep percentage (r = 0.024, p = 0.921) or deep sleep percentage (r = 0.181, p = 0.444). Similarly, there was no correlation between app sleep latency and the traditional PSG sleep latency (r_s = 0.384, p = 0.094) or sustained sleep latency (r_s = 0.214, p = 0.365). A comparison of the absolute parameters obtained by PSG and reported by the app is presented in **Table 3**. The app slightly and nonsignificantly overestimated sleep efficiency by 0.12% (95% confidence interval [CI] −4.9 to 5.1%, p = 0.962), significantly underestimated light sleep by 27.9% (95% CI 19.4–36.4%, p < 0.0001), significantly overestimated deep sleep by 11.1% (95% CI 4.7–17.4%,

p = 0.008). It significantly overestimated traditional sleep latency by 15.6 minutes (95% CI 9.7–21.6, p < 0.0001). In nine patients (45%), the sustained sleep latency was different from the traditional sleep latency. The app significantly overestimated sustained sleep latency by 10 min (95% CI −1.1 to 35.7, p = 0.015). The relationship between PSG sleep efficiency and app accuracy was not statistically significant (r = 0.196, p = 0.40757). However, when AHI and PLMI were considered as covariates in a multiple regression model, the relationship was detected to be statistically significant (for AHI as a covariate, r^2 = 0.445, p = 0.002; for PLMI as a covariate, r^2 = 0.427, p = 0.003).

A total of 527 15-min epochs were analyzed for the epoch-by-epoch comparison. The epochwise statistical performance of the app is presented in **Tables 4** and **5**, and app accuracy during wake and various stages of sleep is presented in **Table 6**. From a sleep-wake detection perspective, the app was highly accurate (85.9%) and sensitive (89.9%) for sleep-wake detection, but was poorly specific (50%), i.e., it tended to overidentify sleep. Analysis of intrasleep stage discrimination revealed that app sensitivity was better in deep sleep than light sleep (62.6% versus 33.9%), and it was most accurate during stage N3 sleep (71.2%), but the overall statistical performance of the app was unsatisfactory from a clinical standpoint, and it had an accuracy of just 45.9%.

As depicted in **Table 7**, a majority, 14 of 17 patients who were asleep at the time the alarm went off, were indeed in light sleep. However, although the app thus appropriately awoke patients out of light sleep 82.4% of the time, given that the mean percentage of light sleep in the sample was 60.7% (**Table 3**), there was no detectable difference between the awakenings out of light and deep sleep (p = 0.159).

DISCUSSION

Our findings suggest that, unfortunately, despite its high sensitivity for detecting sleep and apparent accuracy in sleep-wake detection, the Sleep Time app performs poorly when compared to the gold standard, PSG, in reporting absolute sleep parameters or staging sleep. The app's low specificity suggests that its accuracy would progressively deteriorate as the user's sleep efficiency worsens, making it least reliable in patients with true insomnia, where it could be potentially most helpful. This was demonstrated even in our sample; when OSA and PLMS are accounted for, sleep efficiency and app accuracy did have a statistically significant positive relationship. On the one hand, these data suggest that the app would be unreliable

Table 1—Baseline demographics and polysomnography characteristics of subjects in the study (n = 20).

Age (y)	39.5 ± 12.4
Sex	
Men	12 (60%)
Women	8 (40%)
Body mass index (kg/m^2)	24.7 ± 4.0
Apnea-hypopnea index (events/h)	4.6 ± 5.8
Periodic limb movement index (events/h)	4.9 ± 10.2
Arousal index (events/h)	6.9 ± 6.0
Sleep efficiency (%)	85.9 ± 8.4
Sleep latency (min)	11.6 ± 14.5
Total sleep time (min)	348.7 ± 37.1
Wake after sleep onset (min)	44.4 ± 31.7
Total recording time (min)	405 ± 46.8
N1 sleep (%)	9.1 ± 6.1
N2 sleep (%)	51.6 ± 10.8
N3 sleep (%)	23.1 ± 9.8
REM sleep (%)	16.5 ± 4.7

All values (except sex) expressed as mean ± standard deviation. REM, rapid eye movement.

Table 2—Responses of subjects to the questionnaire regarding presence of sleep complaints in general, and sleeping habits and subjective sleep quality while using the app (n = 20).

	Number of Subjects
Symptoms suggestive of OSA (snoring, choking/gasping in sleep, witnessed apneas)	7 (35%)
Excessive movements in sleep	5 (25%)
Restless legs syndrome-type symptoms (discomfort in legs or arms, worse in the evenings, improving with movement, or an urge to move the limbs)	2 (10%)
Insomnia	4 (20%)
Frequent daytime sleepiness	10 (50%)
Number of nights slept alone while using the app at home for 5 nights before in-laboratory study	
All 5 nights	7 (35%)
2 nights	1 (5%)
1 night	1 (5%)
No nights	11 (55%)
More refreshed while using the alarm feature on the app?	
Yes	3 (15%)
No	15 (75%)
Did not use alarm feature	2 (10%)

OSA, obstructive sleep apnea.

in patients with excessive movements in sleep, as may occur in OSA and PLMS, and on the other hand, it is clear that the high accuracy of the app in our sample with a relatively good sleep efficiency (86.6%) is not a redeeming feature in itself. Given our results, a true test of the app's reliability would be its ability to identify wakefulness while monitoring, for example, a subject lying in bed awake for most of the night, especially if they are relatively immobile. Future studies, therefore, may consider evaluating the performance of the app in patients with a history of insomnia who would be anticipated to have poorer sleep efficiencies.

We recognize that this app was developed for entertainment and educational purposes, and not as a substitute for a formal sleep medicine evaluation or standard sleep studies. We appreciate that the developers have been conscientious in cautioning users against treating the results of this app as diagnostic, or as a substitute for a formal sleep evaluation. Our aim with this study was not to single out this particular app or to comment on its relative merits or demerits, but rather to approach such sleep related apps from the perspective of clinicians, and to evaluate whether pre-existing, inexpensive tools of this kind could be useful in a clinical setting.

Admittedly, our study design carries a few limitations, the most obvious of which is the choice to use 15-min epochs while directly comparing the staging of the PSG and app graph. PSGs are staged in 30-sec epochs, but unfortunately, the app did not provide comparable 30-sec epochs for direct comparison, which would have been ideal. In fact, the app did not divide its recording into epochs at all, and graphically represented various stages of sleep as a continuum over the night (**Figure 1**). We were able to reliably divide the hour-long intervals on the app graph into quarters, representing 15-min epochs, but thought that any further subdivision would suffer in accuracy. We do recognize that several sleep stages are often present in a 15-min epoch, but this is frequently also true of 30-sec epochs. When more than a single stage of sleep is present in a 30-sec PSG epoch, the polysomnographer assigns to that epoch the stage that comprises its majority. We applied the same rule to the 15-min PSG and app graph epochs. Although it could be argued that this may have impacted the calculated accuracy of the app to a certain extent, we thought that overall this was the best approach to take given the limitations that arose from the manner in which the app presented data. In our experience, the app graph tended to show large uninterrupted blocks of a particular stage of sleep, as opposed to the PSG hypnogram that had frequent cycling of sleep stages during the course of the night. In this context, it is to be noted that even the staging of sleep on a PSG is very operator-dependent,

Table 3—Comparison of absolute parameters obtained by polysomnography and provided by the app for subjects in the study (n = 20).

	PSG	App
Sleep efficiency (%)	86.6 ± 8.0	86.5 ± 6.3
Light sleep (%)	60.7 ± 12.4	32.8 ± 13.7 [a]
Deep sleep (%)	39.6 ± 11.9	50.6 ± 8.9 [b]
Traditional sleep latency (min)	11.6 ± 14.5	27.3 ± 16.2 [a]
Sustained sleep latency (min)	17.3 ± 18.4	27.3 ± 16.2 [c]

All values expressed as mean ± standard deviation. PSG, polysomnography. [a] $p < 0.0001$. [b] $p = 0.008$. [c] $p = 0.015$.

Table 4—Sleep-wake detection by the app based on epoch-by-epoch comparison.

Total number of PSG sleep epochs	Total number of epochs correctly scored by the app as sleep (regardless of stage)	Sensitivity	Specificity	PPV	NPV
473	425	89.9% (CI 86.8–92.4)	50.0% (CI 36.1–63.9%)	94.0% (CI 91.4–96.0)	36.0% (CI 25.2–47.9)

CI, confidence interval; NPV, negative predicted value; PPV, positive predicted value; PSG, polysomnography.

Table 5—Statistical performance of intrasleep staging by the app based on epoch-by-epoch comparison.

Sleep stage	Total number of PSG epochs	Total number of epochs correctly staged by the app	Sensitivity	Specificity	PPV	NPV
Light sleep	283	96	33.9% (CI 28.4–39.8)	67.6% (CI 61.4–73.5)	54.9% (CI 47.2–62.4)	46.9% (CI 41.6–52.2)
Deep sleep	190	119	62.6% (CI 55.3–69.5)	53.1% (CI 47.6–58.5)	43.0% (CI 37.1–49.0)	71.6% (CI 65.6–77.1)

CI, confidence interval; NPV, negative predicted value; PPV, positive predicted value; PSG, polysomnography.

Table 6—Stage-wise accuracy of app staging based on epoch-by-epoch comparison.

	All epochs sleep-wake detection only	All epochs staging	Wake	N1	N2	N3	REM
Number of epochs	527	527	54	11	272	111	79
Accuracy	85.9%	45.9%	50.0%	54.5%	33.0%	71.2%	50.6%

N1, stage 1 sleep; N2, stage 2 sleep; N3, stage 3 sleep; REM, rapid eye movement sleep.

Table 7—App performance in awakening patients out of light sleep during the in-laboratory polysomnography.

	Number of Subjects
Awake at the time of app alarm	3/20 (15%)
Asleep and in light sleep at the time of app alarm	14/17 (82.4%)
Asleep and in deep sleep at the time of app alarm	3/17 (17.6%)

and several of the established rules are somewhat arbitrary and based on consensus guidelines.

The use of multiple iPhones, each of a different model and using different operating systems, as well as different versions of the app with each update released by the developer, may also have introduced a certain element of nonstandardization. However, this is a real-world problem in evaluating any app that is targeted at the general population and meant to be used on a device that is being constantly upgraded and re-released.

It is also important to keep in mind that we only studied the Sleep Time app for iPhones, and our findings should not be considered generalizable for all similar apps and on all smartphones. This app is also available for Android smartphones, and future research may focus on whether it proves more reliable on that platform. In addition, evaluating several similar such apps, perhaps simultaneously on the same device under the same conditions (or the same app simultaneously on different devices) may clarify whether the limitations found on our study are app-specific or device-specific.

The specifics of the algorithm used by the app to detect sleep stages has not been revealed and appear to be proprietary, and we did not receive a response when we contacted the developer through their website for more details. However, the app's website suggests that, like most similar apps, it uses the iPhone's inbuilt accelerometer to detect body movements when the device is placed on the subject's bed, and extrapolates this data into sleep staging, because body movements progressively decrease as a subject progresses from wakefulness to light sleep and deep sleep. The basic concept appears to be a fundamentally sound one, since such whole-body bed-based actigraphy has been found to be correlate impressively well with PSG and actigraphy in experimental studies.[14] However, clearly, the sensitivity of the smartphone's accelerometer is a major potential limiting factors in the accuracy of such apps, and the variation in mattress firmness (impacting transmission of movements to the iPhone), would also be a confounding factor. While these are, again, unavoidable obstacles when attempting to create a simple app on an mass-produced device used in different home environments, recent pilot studies have suggested that sleep-wake analysis based on movements detected by a smartphone accelerometer placed on the pillow correlates fairly well with watch-based actigraphy.[15] It is possible that physically attaching the smartphone to a subject's body (such as fastening it to a limb or the trunk) may improve the ability of the accelerometer to detect movements and improve app data validity, and this may be worth investigating in future studies. However, this is not the recommended means of gathering data per the developers, and may increase subject discomfort. Interestingly, investigators were recently able to demonstrate that placing wireless external sensors placed on a bed and integrating them into a mobile phone's platform is almost 80% accurate in distinguishing wakefulness, nonrapid eye movement and REM sleep in a home setting.[16] The use of such additional unobtrusive sensors, making no contact with the subject's body and therefore causing no discomfort, represents a potential means to improve the accuracy of such apps.

An interesting observation emerging from our data is the similarity in the epochwise performance of sleep-wake detection by the app to that reported in the literature for actigraphy.[7,9] Several studies have shown that, as with most devices that rely on movement-based algorithms, actigraphy has a high sensitivity but poor specificity in detecting sleep, and as a result its accuracy suffers as sleep efficiency decreases.[8] This raises the question as to whether using the app might represent an unobtrusive, cost-effective alternative to actigraphy with no loss of reliability. However, any further discussion needs to be tempered by the fact that such calculations for actigraphy are based on more traditional 30- or 60-sec epochs. In this context, the development of apps that are able to divide their sleep-wake data into epochs of durations that are more meaningful to researchers would greatly aid in analysis and app utility, and represents an avenue for software developers to focus future efforts. Nevertheless, our findings do suggest that testing the app against actigraphy may be worthwhile. Equally constructive would be a head-to-head comparison of the app and subjective measures such as sleep diaries and questionnaires. Such instruments, while widely used in clinical practice despite conflicting reports of their reliability,[17–20] require active patient involvement and motivation, and are often abandoned as being burdensome and inconvenient. Sleep apps such as this one that passively collect comparable data may represent a potential substitute for sleep diaries, especially in patients with poor motivation to maintain useful logs of their sleep.

In addition, future research may investigate whether apps that use nonmovement-based methods to stage sleep may produce better results. This includes PSG validation of apps using sleep detection algorithms based on the sound detection capabilities of smartphone microphones.[21] Intriguingly, several methods have been recently described to use existing smartphone technology (camera, optical pulse sensor, and photoplethysmography) to reliably monitor heart rate variability and respiratory rate.[22–26] The development of smartphone apps that use autonomic measures in such a manner might present an avenue for more clinically useful sleep staging.

Summarizing, our findings suggest that while the Sleep Time app is insufficiently precise to be alternative to PSG in sleep-wake cycle analysis, it merits further evaluation, especially in comparison to actigraphy or subjective measurements of sleep such as sleep diaries. Regardless of the results of any such investigations, however, the explosion of literature detailing fairly ingenious methods of using the existing capabilities of smartphones raises the hope that clinically acceptable app-based sleep cycle analysis outside the sleep laboratory is close to realization. It is clear that smartphones are now an indispensable part of life in the 21st century, and the development of inexpensive, reliable apps that provide accurate information pertinent to the practice of sleep medicine should be encouraged, especially in view of the trend toward

out-of-center testing for conditions such as sleep disordered breathing, where electroencephalography channels are often not used. Such apps would also be revolutionary in the evaluation of circadian rhythm disorders or in patients with hypersomnia or insomnia, where broad but dependable sleep-wake trends, rather than exact stage identification, suffice.

ABBREVIATIONS

AF, atrial fibrillation
AHI, apnea hypopnea index
BMI, body mass index
CAD, coronary artery disease
CHF, congestive heart failure
CI, confidence interval
COPD, chronic obstructive pulmonary disease
CPE, clinical pathway evaluation
DM, diabetes mellitus
EM, expectation-maximization
HTN, hypertension
IQR, interquartile range
MCC, Matthew's correlation coefficient
MI, myocardial infarction
NPV, negative predicted value
ODI, oxygen desaturation index
OSA, obstructive sleep apnea
PLMS, periodic movements in sleep
PPV, positive predicted value
PSG, polysomnography
REM, rapid eye movement
SD, standard deviation
TRT, total recording time
TST, total sleep time

REFERENCES

1. Batista MA, Gaglani SM. The future of smartphones in health care. Virtual Mentor 2013;15:947–50.
2. Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. Physiol Meas 2013;34:R29–46.
3. Al-Mardini M, Aloul F, Sagahyroon A, Al-Husseini L. Classifying obstructive sleep apnea using smartphones. J Biomed Inform 2014;52:251–9.
4. Nakano H, Hirayama K, Sadamitsu Y, et al. Monitoring sound to quantify snoring and sleep apnea severity using a smartphone: proof of concept. J Clin Sleep Med 2014;10:73–8.
5. Bhopi R, Nagy D, Erichsen D. Can a novel smartphone application detect periodic limb movements? Stud Health Technol Inform 2012;182:36–42.
6. Min YH, Lee JW, Shin YW, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014;16:e135.
7. 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.
8. Sivertsen B, Omvik S, Havik OE, et al. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. Sleep 2006;29:1353–8.
9. 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.
10. Grifantini K. How's my sleep? Personal sleep trackers are gaining in popularity, but their accuracy is still open to debate. IEEE Pulse 2014;5:14–8.
11. Wilson DL, Fung A, Walker SP, Barnes M. Subjective reports versus objective measurement of sleep latency and sleep duration in pregnancy. Behav Sleep Med 2013;11:207–21.
12. Berry RB, Brooks R, Gamaldo CE, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology, and technical specifications. Version 2.0.3. www.aasmnet.org. Darien, IL: American Academy of Sleep Medicine, 2014.
13. Glanz SA. Primer of biostatistics, fifth edition. New York, NY: McGraw-Hill, 2002.
14. Choi BH, Seo JW, Choi JM, et al. Non-constraining sleep/wake monitoring system using bed actigraphy. Med Biol Eng Comput 2007;45:107–14.
15. Natale V, Drejak M, Erbacci A, Tonetti L, Fabbri M, Marton M. Monitoring sleep with a smartphone accelerometer. Sleep Biol Rhythms 2012;10:287–92.
16. Gradl S, Leutheuser H, Kugler P, et al. Somnography using unobtrusive motion sensors and Android-based mobile phones. Conf Proc IEEE Eng Med Biol Soc 2013;1182–5.
17. Armitage R, Trivedi M, Hoffmann R, Rush AJ. Relationship between objective and subjective sleep measures in depressed patients and healthy controls. Depress Anxiety 1997;5:97–102.
18. Okifuji A, Hare BD. Nightly analyses of subjective and objective (actigraphy) measures of sleep in fibromyalgia syndrome: what accounts for the discrepancy? Clin J Pain 2011;27:289–96.
19. Gonzalez R, Tamminga C, Tohen M, Suppes T. Comparison of objective and subjective assessments of sleep time in subjects with bipolar disorder. J Affect Disord 2013;149:363–6.
20. Moore CM, Schmiege SJ, Matthews EE. Actigraphy and sleep diary measurements in breast cancer survivors: discrepancy in selected sleep parameters. Behav Sleep Med 2014;12:1–19.
21. Krejcar O, Jirka J, Janckulik D. Use of mobile phones as intelligent sensors for sound input analysis and sleep state detection. Sensors (Basel) 2011;11:6037–55.
22. Heathers JA. Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research. Int J Psychophysiol 2013;89:297–304.
23. Jonathan E, Leahy M. Investigating a smartphone imaging unit for photoplethysmography. Physiol Meas 2010;31:N79–83.
24. Nam Y, Lee J, Chon KH. Respiratory rate estimation from the built-in cameras of smartphones and tablets. Ann Biomed Eng 2014;42:885–98.
25. Matsumura K, Yamakoshi T. iPhysioMeter: a new approach for measuring heart rate and normalized pulse volume using only a smartphone. Behav Res Methods 2013;45:1272–8.
26. Kwon S, Kim H, Park KS. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. Conf Proc IEEE Eng Med Biol Soc 2012;2174–7.

ACKNOWLEDGMENTS

SUBMISSION & CORRESPONDENCE INFORMATION

DISCLOSURE STATEMENT