Objective Measurement of Sleep by Smartphone Application: Comparison with Actigraphy and Relation to Cognition, Mood, and Self-Reported Sleep

Taylor Maynard
Objective Measurement of Sleep by Smartphone Application: Comparison with Actigraphy and Relation to Cognition, Mood, and Self-Reported Sleep

TAYLOR MAYNARD

Sleep duration and quality have a considerable impact on countless aspects of individuals’ lives. Sleep deprivation can be extremely taxing on one’s overall health. This is in large measure due to impairment of cognitive functions, such as psychomotor and processing speed, executive attention, higher cognitive abilities, and working memory (Goel, Hengyi, Durmer, & Dinges, 2009). The National Heart, Lung, and Brain Institute (2012) identify myriad reasons as to why sleep is important to health, such as how it protects and improves physical health, mental health, safety, and overall quality of life.

A human spends one third of their life asleep; this equates to 25 years spent sleeping (Heppner, 2015). Wilckens, Erickson, and Wheeler (2012) state that deep sleep (also known as slow-wave sleep) can substantially benefit the functioning of the prefrontal cortex (PFC) and can consequently contribute to normal cognitive functions such as working memory, inhibition, and controlled memory processes. This idea supports the hypothesis that executive functioning associated with networks of the PFC would be the most sensitive to subjective variations in sleep (Wilkens et al., 2012). Given that PFC networks have been reported to be the most sensitive to individual variations in sleep, it is important to note how detrimental clinical sleep disorders can be for individuals suffering from them. Approximately 70 million Americans are affected by sleep disorders; 90 individual sleep disorders having been identified in the literature (Colten & Altevogt, 2006). Sleep disturbances that result in impairments during the daytime are frequently reported in adults (Stepanski, Rybarczyk, Lopez, & Stevens, 2003) and sleep deprivation has been shown to be detrimental to cognition (Goel et al., 2009) and even motor sequence acquisition in younger adults (Appleman, Albouy, Doyon, Conin-Golomb, & King, 2015). Bonnet (1986) reported that poor quality of sleep or lack of long periods of uninterrupted sleep produces adverse effects on cognitive performance and motor functioning during the daytime hours; in turn, this sleep deficit can significantly contribute to the occurrence of automobile accidents, lack of concentration, memory impairment, proper limbic system functioning, and the inability to complete daily tasks. Baglioni, Spiegelhalder, Lombardo, and Riemann (2010) claim that disturbances in sleep have yielded a number of detriments, including emotional reactivity and poor emotion regu-
lation that is problematic to one’s mental health. Monitoring sleep could easily become an integral part of treatment in the field of rehabilitation psychology. The standard methods used to evaluate sleep variables such as total sleep time (TST), sleep efficiency, and stages of sleep have blossomed in sync with technological advancement over the years, beginning with the evolution of polysomnography (PSG) throughout the 1960s and 1970s (Haba-Rubio & Krieger, 2012), actigraphy in 1988 for the United States military (Actigraph, 2016), and finally the incorporation of accelerometer technology in smartphones during recent years.

Polysomnography (PSG)

Today, PSG continues to be the gold standard for objective measurement of sleep. PSG consists of continuous monitoring of variables of neurophysiology and cardiorespiration. PSG sleep studies are typically conducted over the span of one night for the purpose of monitoring both disturbed and normal sleep (Bloch, 1997). Since 1997, it has become increasingly common for PSG sleep studies to require more than just one night of sleep monitoring (e.g., Jarrin et al., 2016). Despite how common overnight sleep studies are, some researchers are skeptical of whether external factors (for example, a laboratory setting) can impact the validity of study results. For instance, Iber and colleagues (2003) questioned the external validity of sleep studies and postulated that body position (while in bed), alcohol consumption, sleep stage distribution, and sleep duration may all play a significant role in subjects’ sleep patterns and quality. Iber et al. (2003) also state that within-laboratory or within-home sleep monitoring as well as the degree of supervision of monitoring could potentially affect physiologic parameters of sleep. From a broader perspective on PSG that falls outside of sleep apnea studies, it is imperative to look not just at the effects of location on data outcomes, but also on individual comfort levels that could consequently affect the data. Siversten et al. (2006) asserts that PSG can be invasive, expensive, and time-consuming for some. Employing wrist actigraphy technology can be a comfortable and inexpensive alternative to traditional PSG monitoring, depending on patients’ needs and preference.

Actigraphy

Actigraph technology can be set apart from PSG monitoring, as it allows the convenience and comfort of wearing the device as a watch-like bracelet on the non-dominant wrist or ankle, directly in contact with the body, and within one’s typical environment, thus, providing significant external validity. Actigraphy has been utilized to measure sleep for the purpose of research for over two decades (Mantua, Gravel, & Spen-

154 • The Undergraduate Review • 2017 Bridgewater State University
in actigraphs offer promising calculations of movement and exerted energy, and since the 1980’s, researchers have been employing actigraphy as a reliable means of quantifying physical activity (Dinesh & Freedson, 2012).

Despite supportive and validating research on actigraphy technology, Verbeek et al. (1994) proposed that actigraphy has been shown to overestimate TST in comparison to PSG. However, Martin and Hakim (2011) claim that in recent studies, actigraphy bracelets have been valuable in calculating the TST of an individual as well as the efficiency of sleep.

**Smartphone Applications**

Today, there is a variety of free smartphone apps that are used to monitor individuals’ health, diet, fitness, smoking habits, etc.; likewise, wearable technology such as the Fitbit (Fitbit, Inc., San Francisco, CA, USA) is becoming ubiquitous in Western society (Zambotti et al., 2016). Apple, Inc. proclaimed in a 2009 television commercial posted on YouTube by user CommercialKid et al. (2009) that “There’s an app for that,” regarding the creation of countless smartphone applications (many of which were free of charge) that yielded a plethora of uses (including healthcare) to people around the world (Bhat et al., 2015). Consequently, the creation of smartphone apps that monitor and evaluate the quality of one’s sleep and its duration have been marketed through the app stores for use on smartphones and other mobile devices, such as smartwatches and tablets. Batista and Gaglani (2013) suggest that the potential for a new frontline of clinical practice and research is being offered by the prevalent use of advanced smartphone technology. One such example of a smartphone application that could be utilized in both healthy and clinical populations is the Sleep Time smartphone app (Azmio, Inc., Palo Alto, CA, USA). Bhat and colleagues’ 2015 study compared sleep efficiency, sleep latency, and light sleep/deep sleep percentages recorded through the Sleep Time app to those recorded through PSG.

As correlations between sleep data collected through the app and through PSG were poor, further research could compare both sleep efficiency and TST recorded through the app to data recorded through actigraphy technology. Contingent upon future research findings, the Sleep Time app could be a cost-efficient and innovative alternative to the expensive and invasive use of overnight PSG monitoring or even the in-home use of clinical devices like actigraphy.

Certain clinical populations may prefer to use smartphone applications that record their sleep data, rather than to fill out a daily sleep diary when asked to by their doctor. Likewise, those who wish to have their sleep data in real time and at their disposal might prefer a smartphone app to actigraphy, which does not immediately display a previous night or week of sleep without first returning to a doctor’s office or research laboratory for data extraction. There are a variety of clinical populations that could benefit from technological sleep tracking methodologies, such as adults with insomnia, narcolepsy, sleep apnea, or other sleep disorders. Older adult populations with cognitive deficits or neurodegenerative diseases could also make use of smartphone applications that track their sleep, as poor sleep quality is often a co-occurring symptom of diseases such as Alzheimer’s and Parkinson’s disease (Vitiello, Poceta, & Prinz, 1991).
Effects of Sleep on Cognition and Mood

It has been suggested that both sleep quality and TST can be culprits of weaker waking neurobehavioral functioning, such as slowed working memory, reduced cognitive throughput, lapses of attention, and depressed mood in both middle-aged and older healthy adults (Banks & Dinges, 2007; Sutter, Zollig, Allemand, & Martin, 2012). Moreover, Van Dongen et al. (2003) and Wilckens et. al (2014) demonstrated that sleep continuity and TST impact executive functioning across the lifespan, while Yoo, Hu, Gujar, Jolesz, and Walker (2007) showed that sleep is integral for encoding memories and consolidation following learning. In addition, Baglioni et al. (2010) and Leotta, Carskadon, Acebo, Seifer, & Quinn (1997) conjectured that sleep deprivation and fragmented sleep are associated with negative emotional consequences, such as fluctuations in emotional reactivity.

Current Project

The current project compared data collected objectively by the Sleep Time app (Azumio, Inc., Palo Alto, CA, USA) to data recorded from actigraphy bracelets worn by participants recruited from individuals known to members of Boston University’s Vision and Cognition Laboratory. Sleep diaries from the National Sleep Foundation as well as self-report questionnaires were utilized to gather subjective data from participants. The primary variables of interest were TST and sleep efficiency; both the Sleep Time app and actigraphy bracelets monitored and recorded time asleep and calculated sleep efficiency as a percentage. Furthermore, differences between objective data generated from the Sleep Time app and actigraphy as well as subjective self-report sleep measures completed by participants were examined. Participant-completed sleep diaries were used in conjunction with the Sleep Time app and actigraphy to control for potential malfunction or non-compliance with either the actigraph or the app. Subjective sleep measures were used to account for when actigraphy overestimated TST in individuals who may have had trouble falling asleep at night but remained motionless in their beds (Martin et al., 2011). Thus, self-report measures were integral to lending the most accurate data possible to the project. Lastly, participant scores from two assessments of executive functioning and two self-report measures of mood were analyzed to examine a potential interplay of sleep quality and sleep duration with cognition and mood.

Hypotheses for the study include the following: 1) Self-report information recording TST as obtained by sleep diaries will differ from objective information on TST as collected by the Sleep Time app; 2) Self-reported sleep quality will correlate with performance on tasks of executive functioning and working memory in addition to self-reported scores of perceived depression or anxiety; 3) There will be a significant difference between means of TST and sleep efficiency as recorded by the Sleep Time app and actigraphy bracelets. Comparing means on the two variables of sleep could lend credence to the idea that the Sleep Time app can be used as a valid objective method of measuring sleep if no significant differences are calculated.

Method

Participants

The current project collected and examined both subjective and objective data from 29 healthy adult volunteers ranging from 20 to 67 years (\(M_{\text{age}} = 26.8\) years,
The project was publicized through convenience sampling (i.e. word-of-mouth and social media technology). Screening questionnaires were administered to interested adults to determine individual eligibility. If screened individuals were free of acute or chronic medical conditions and sleep disorders (e.g. obstructive sleep apnea, restless leg syndrome, clinical insomnia, etc.), and were not taking any prescription or over-the-counter sleep medications, they were queried and included in the study. Additionally, participants had access to the daily use of an iPhone with IOS 8.0 or later for the daily use of the Sleep Time app (a free app download in the Apple store). Written informed consent of each participant was obtained on the day of their first study visit. This study was approved by the Institutional Review Board (IRB) of Boston University.

Methods and Procedures
In addition to filling out a daily sleep diary from the National Sleep Foundation and using the Sleep Time app, 19 out of the 29 participants also agreed to wear an actigraphy bracelet for one week. All 29 participants returned for a second study visit; five self-report questionnaires and two cognitive tasks were administered. Among the five questionnaires was an exit questionnaire that reflected the app’s feasibility and likelihood of individual use in the future. Each of the other questionnaires and the two cognitive tasks are further described below.

Self-report measures of sleep.
Epworth Sleepiness Scale (ESS).
The ESS is a questionnaire that provides subjective information about one’s perceived daytime sleepiness across eight different situations of everyday life, such as reading, watching TV, or being in a car during traffic. Participants rate themselves on a scale of 0 (“would never doze”) to 3 (“high chance of dozing”) across 8 situations of everyday life. The ESS asks for responses that reflect the most recent way of living. Each test’s 8 responses (0-3) are totaled (dependent variable). Any total score over 10 signifies the presence of excessive daytime sleepiness, indicating the possibility of a clinical sleep disorder.

Self-report measures of mood.
Beck Depression Inventory II (BDI-II).
The BDI-II was administered upon each participant’s second study visit. This 21-question self-report mood metric of depression was employed to gather data on perceived symptoms of depression; each answer provided is scored on a scale of 0-3. The dependent variable of the BDI-II is the total score out of 21 questions. A score range of 14-19 suggests that the individual experiences mild depression, while a score of 29-63 suggests severe depression.

Beck Anxiety Inventory (BAI).
The BAI was also completed by participants during the second study visit. The BAI is a 21-question self-report mood metric of anxiety, utilized to gather data on subjective symptoms of anxiety; each answer provided is scored on a scale of 0-3, with a score of 0 indicating that the individual does not experience a specific symptom of anxiety at all, while a score of 3 indicates that individual is frequently bothered by a particular symptom. The dependent variable of the BAI is the total score out of 21. A score of 0-9 suggests minimal anxiety while a score of 30-63 suggests severe anxiety.
Pittsburgh Sleep Quality Index (PSQI).
The PSQI is a 19-item self-report questionnaire that collects subjective information about participants’ sleep quality over the course of the past month. The global score of the PSQI can fall in the range of 0-21, and is calculated through 7 component scores, with each question weighted on an interval scale of 0-3. The lower an individual’s score, the better their suggested sleep quality. The current study compared general sleep quality data as reported a single PSQI question (about overall sleep quality) to performance on cognitive assessments.

Measures of executive functioning.
Letter-number sequencing.
Letter-Number Sequencing is a cognitive measure from the Wechsler Adult Intelligence Scale IV (WAIS-IV) that assesses participants’ attention, concentration, short-term auditory memory, sequential processing, memory span, and mental manipulation. Individuals first listen intently to the assessor as they read aloud a random sequence of letters and numbers; the individual is instructed to repeat the sequence back to the assessor by first listing the given numbers in order, followed by the letters alphabetically. The Letter-Number Sequencing assessment consists of 10 test items with 3 trials per item. Each trial is scored with either 0 or 1 point; a total of 3 points can be earned per test item. Once a score of 0 is earned for an entire test item, the assessment is discontinued and all scores from the completed test items are added together for one sum score (dependent variable), ranging from 0-30. In addition, a longest letter-number sequence score (dependent variable) is calculated, given the total number of letters in numbers in one sequence that the participant was able to accurately recall; this score can range from 0-8.

Verbal fluency (FAS and category-animals).
The Verbal Fluency (FAS) assessment is one form of the Controlled Oral Word Association (COWA) Test. The Verbal Fluency (FAS) is a free-recall task which instructs individuals to name as many words as they can (excluding proper nouns) beginning with the letters F, A, and S (phonemic processing) within a one-minute time period, excluding any proper nouns or different forms of the same words, such as “big,” “bigger,” and “biggest.” The number of words recorded in 15-second increments of time is also noted in order to easily examine cognitive clustering. In addition, naming as many animals as possible (semantic processing) in a one-minute time frame was included in this assessment to further measure executive functioning. The dependent variable is the total number of acceptable words produced within each one-minute period of time.

Objective measures of sleep.
ActiGraph GT9X Link by Actigraph.
For a continuous period of one week, a subset of 19 out of 29 participants wore the Actigraph GT9X Link watch on their non-dominant wrist. Movement during wake and sleep time was recorded through accelerometer technology 24 hours a day. The dependent measures examined were averages of each participant’s 7 days of calculated TST (in minutes) and sleep efficiency (percentage).

Sleep Time (Azumio, inc.) smartphone application.
Sleep Time (Azumio Inc., Palo Alto, CA) was utilized by participants to monitor their sleep. The free version of this smartphone app is available in the Apple App
Store. The current model of this application is 5.28 (updated 05/12/2016). TST and sleep efficiency were monitored and recorded with this app; daily TST minutes and sleep efficiency percentages were averaged for each participant, based on the 7 days of use.

Subjective measures of sleep.
National Sleep Foundation sleep diary.
Provided for free by the National Sleep Foundation, sleep diaries were administered to all participants to fill out each morning and night for 7 days. This accounted for actigraph and Sleep Time app malfunctioning and noncompliance. Self-reported TST and sleep quality were analyzed in addition to certain aspects of participants’ daytime and nighttime routines. Each item on the sleep diary was assigned a numerical value (0-7, depending on the kind of question) for use in correlational analyses, and other participant responses were recorded verbatim in the database for qualitative use.

Ease of use and likelihood of future use.
Exit questionnaire.
Participants filled out an exit questionnaire during their second study visit that asked about subjective lifestyle factors such as the type of mattress they use and whether or not they sleep with pets in their beds. In order to gather data on participants’ subjective opinions of using the Sleep Time app to monitor TST and sleep efficiency in the future, participants were asked to rate the app’s ease of use as well as the likelihood that they would use the app again in the future. Correlational analyses were conducted to examine whether or not ease of use correlated with likelihood of future use as well as to see if age played a role in ease of use.

Data Analysis
TST and sleep efficiency data were extracted from the actigraphs and the Sleep Time app and each participant’s 7 days of data were averaged for each methodology. Paired samples t-tests were performed to determine significant differences between TST and sleep efficiency as recorded through actigraphy and the smartphone application. Additionally, Pearson correlation analyses were executed to identify potential correlations between self-reported sleep quality and performance on two cognitive tasks that assessed executive functioning and working memory. The sleep data were also correlated with certain variables from the administered self-report sleep and mood questionnaires.

Results
Subjective vs. Objective TST Data
It was initially hypothesized that self-report information recording TST as obtained by sleep diaries would differ from objective data on TST as collected by the Sleep Time app. A paired samples t-test indicated that there was not a significant difference between TST recorded by the app and TST recorded in the sleep diaries ($M = -3.96$, $SD = 42.1$), $t(10) = .31, p = .77$.

Effects on Cognition and Mood
The second hypothesis predicted that self-reported sleep quality would correlate with performance on tasks of executive functioning in addition to self-reported scores of perceived depression or anxiety. There were no correlations between self-reported sleep quality and performance on cognitive assessments, with the exception of a correlation between self-reported sleep quality as measured by the Epworth Sleepiness Scale and lower scores on the Verbal Fluency animals.
category, Pearson’s $r(27) = -.07, p < .05$. No significant correlations between scores on the BDI-II and BAI and TST or sleep efficiency were observed, though individuals’ scores on the BDI-II and the BAI were positively correlated at the 0.05 level in this population, Pearson’s $r(23)= .43$. In addition, age positively correlated with BDI-II scores at the 0.01 level, Pearson’s $r(23)= .54$; older participants had higher scores on the BDI-II.

**TST and Sleep Efficiency**

The third hypothesis stated that TST and sleep efficiency as recorded through the sleep time app would significantly differ from sleep data recorded through actigraphy. A paired-samples t-test showed that the smartphone app overestimated TST when compared to actigraphy, thus yielding a significant difference, $t(18) = 6.64, p < .001$. Moreover, a statistical trend indicates that the smartphone app also overestimated sleep efficiency when compared to actigraphy, $t(18) = 2.06, p = .055$.

**Self-reported App Use**

Participants’ self-reported ease of use of the smartphone app positively correlated with their likelihood of future use at the 0.01 level, Pearson’s $r(27)= .471$. No significant correlations between age, ease of use, and likelihood of future use were found.

**Discussion**

This study investigated the utility of a smartphone application in monitoring sleep in healthy adults in comparison to actigraphy technology. Potential correlations between sleep efficiency and performance on tasks of executive functioning and mood metrics were also examined.

The first hypothesis stated that self-report information recording TST as obtained by sleep diaries would be significantly different from objective TST data collected by the Sleep Time app. The results of a paired-samples t-test refuted this hypothesis, as there were no significant differences between subjective data recorded in sleep diaries and objective sleep data collected by the Sleep Time app. This finding suggests that the Sleep Time app, when compared to subjectively recorded TST, is an accurate predictor of TST in a healthy adult population.
should further investigate the effects of slow-wave sleep disruption on mood within clinical populations. Establishing the importance of sleep continuity in both healthy and clinical populations could offer new insights into the development of new sleep-tracking devices as well as improvement of current sleep-tracking devices such as smartphone applications and wrist watches.

The final hypothesis predicted that there would be a significant difference between the means of TST and sleep efficiency as recorded by the Sleep Time app and actigraphy bracelets. Supporting this hypothesis and consistent with the little literature that tests the utility of the Sleep Time app, significant differences were found between the two sleep-tracking methodologies, as calculated through paired-samples t-tests. This suggests that the Sleep Time app is not a valid measure of TST and sleep efficiency in a healthy adult population. Bhat and colleagues (2015) arrived at similar results when assessing the utility of the Sleep Time app in monitoring sleep efficiency and other sleep variables when compared to overnight PSG sleep monitoring. Despite the prior findings, the present study further tested the Sleep Time app by monitoring TST and sleep efficiency and comparing it to data collected through actigraphy.

Future directions in smartphone and wearable, sleep-tracking technology should aim to adopt standard metrics for validation of these devices in order to inform both the research world as well as companies interested in marketing their devices for research and clinical applications. Relatedly, Zambotti and colleagues (2016) stress that there is a pressing need for active communication between the sleep research community and companies that market wearable technology, so that marketing output of these devices can be optimized; this would endorse the implementation of commercial sleep-tracking devices in both clinical and laboratory settings as a cost-efficient option for researchers, patients, and doctors.

Currently, there is insufficient evidence to support that commercial measures of objective sleep such as apps and smartwatches possess the utility for applications in research and clinical settings (Kolla, Mansukhani, & Mansukhani, 2016). Given the consistent overestimation of TST and sleep efficiency recorded through the Sleep Time app in the current study, however, clinical sleep professionals could potentially utilize the app in gathering sleep data on those suffering from sleep disorders. The Sleep Time app may offer a more accurate metric in individuals whose sleep is extremely poor, as recording frequent sleep fragmentation throughout the night in an individual with a sleep disorder could yield a more accurate calculation of sleep time and efficiency.

Conclusion

While potentially appealing to some users, smartphone technology does not provide a sensitive enough metric of common sleep variables in a healthy adult population; future research is needed to clarify the Sleep Time app’s objective measurement of sleep as well as its utility within clinical populations and across other models of smartphones. Likewise, objective measurement of sleep through the use of wearable movement and sleep-tracking devices should be further validated for the purpose of employing cost-efficient and expedient means of monitoring sleep in both research and clinical settings.
References


Sleep Technology, 64, 19-30. doi: 10.1007/978-94-007-5470-6


Mantua, J., Gravel, N., & Spencer, R. M. C. (2016). Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors, 16, 646. doi: 10.3390/s16050646


Wake Research in the Netherlands, 5, 163-169.


About the Author

Taylor Maynard is a senior majoring in Psychology who will graduate in May 2017. Her research project, made possible through the Adrian Tinsley Program Summer Research Grant, was conducted throughout the spring, summer, and fall of 2016 under the mentorship of Dr. Sandy Neargarder (Psychology, Bridgewater State University) and a fifth-year graduate student, Erica Appleman, from Dr. Alice Cronin-Golomb’s Vision and Cognition Laboratory at Boston University. Taylor presented her research at Harvard University’s 2017 National Collegiate Research Conference (NCRC) in Boston, MA, as well as at the 2017 International Neuropsychological Society (INS) conference in New Orleans, LA. Taylor plans to attend an experimental psychology Ph.D. program in the fall of 2017 to pursue her career goal of becoming a researcher and professor of Psychology.