ACCURACY OF THREE ASSESSMENTS OF SLEEP TIMING, DURATION AND EFFICIENCY COMPARED TO A SINGLE-CHANNEL EEG DEVICE

Jessica R. Dietch, M.S.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2019

APPROVED:

Daniel J. Taylor, Major Professor
Camilo J. Ruggero, Committee Member
Kimberly Kelly, Committee Member
Vicki Campbell, Chair of the Department of Psychology
David Holdeman, Dean of the College of Liberal Arts and Social Sciences
Victor Prybutok, Dean of the Toulouse Graduate School

Dietch, Jessica R. *Accuracy of Three Assessments of Sleep Timing, Duration and Efficiency Compared to a Single-Channel EEG Device*. Doctor of Philosophy (Clinical Health Psychology), August 2019, 55 pp., 8 tables, 4 figures, references, 62 titles.

Poor sleep measured across many dimensions has been linked to adverse physical and mental health outcomes including cardiovascular disease, diabetes, cancer, increased mortality, depression, and anxiety. Current research typically relies upon brief, subjective, inadequately validated methods to assess limited dimensions of sleep, resulting in inaccurate measurements and possibly faulty conclusions. Specifically, research validating objective (e.g., actigraphy) and subjective (e.g., sleep diaries, retrospective surveys) measurement methods against the gold standard of polysomnography (PSG, an overnight sleep study) is primarily limited by a) a lack of reliability based on too short (e.g., 24 or 48 hours) of an assessment period to capture night-to-night variability, b) a lack of ecological validity (e.g., full PSG in a laboratory setting), and c) a lack of generalizability due to limited or special populations (e.g., individuals with insomnia). Barriers such as prohibitive cost, extensive setup time, and personnel training requirements diminish the ability of researchers to conduct measurement comparison studies using gold standard measures like traditional PSG. These barriers can be circumvented with the use of low-cost, minimally invasive single-channel EEG devices (e.g., Zmachine), but to date few studies have employed these devices. The current study evaluated the accuracy of retrospective surveys, sleep diaries, and actigraphy compared to a single-channel EEG device for assessment of sleep timing, duration, and efficiency in participants' homes over one week using a broad community sample (N = 80). Actigraphy generally demonstrated the best agreement with Zmachine across sleep variables, followed by diary and then survey. Circadian midpoint was the most consistent across measures, followed by sleep duration and then sleep efficiency. Implications and future directions are discussed.

ACKNOWLEDGEMENTS

First, thank you to my mentor and friend Daniel Taylor for never holding back or holding *me* back, and for showing (and telling) me how to think, write, and live like a scientist. If I can pass on to my mentees just a piece of what you gave me, I will be a tremendous success. Thank you to my other mentors, current and former, formal and informal, particularly Kristi Pruiksma, Ruby Christian-Brougham, Shari Kuchenbecker, and the late John Flowers. Thank you to my committee members, Camilo Ruggero and Kimberly Kelly, for your patience and support.

Thank you to my research assistants, in particular Kirti and Brett. I am amazed by how freely you dedicated your time, energy, and thoughtfulness to our project. Most importantly, you gave me the hope that the next generation of scientists will surpass our expectations. Thank you to my participants for sleeping in the name of science. Thank you to General Sleep Corporation, particularly Richard Kaplan, and the Foundation for Rehabilitation Psychology for your generosity and financial support. Thank you to my Clinical Health cohort, the last of our kind—we're basically almost doctors now. Thank you to my friends and colleagues who helped me process this tremendous journey and also reminded me to stop processing and live once in a while. Thank you to my Sleep and Health Research labmates who shared a passion for sleep and a tolerance for Dr. Taylor and whose guidance I intend on seeking for the rest of my career. Thank you to the women (and men) of North Texas Roller Derby. From you, I learned how to fail, how to keep getting off the floor, and how to empower and be empowered.

Thanks Mom, for believing I can do anything. Thanks Dad, for helping me discover the pleasure of finding things out. Finally, thank you to my partner Jordan Eusebio for everything you do, always. This victory is half yours.

Science is measurement – no knowledge without measurement.
—Patrick O'Brian, Master and Commander

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Poor sleep measured across many dimensions has been linked to adverse physical and mental health outcomes (Buysse, 2014; Czeisler, 2015). Current research in health disciplines typically relies upon brief, subjective, inadequately validated methods to assess limited dimensions of sleep, which results in inaccurate measurements that impede scientific progress and discovery (Grandner, Hale, Moore, & Patel, 2010). Specifically, research validating objective (e.g., actigraphy) and subjective (e.g., sleep diaries, retrospective surveys) measurement methods against the gold standard of polysomnography (PSG) is primarily limited by a) a validation period too short to capture night-to-night variability (e.g., 24 or 48 hours), b) a lack of ecological validity (e.g., full PSG in a laboratory setting), and c) limited or special populations (e.g., individuals with insomnia). However, barriers such as prohibitive cost and setup time diminish the ability of researchers to conduct measurement comparison studies using traditional PSG. These barriers can be circumvented with the use of low-cost, minimally invasive single-channel EEG devices (e.g., Zmachine), but to date few studies have employed these devices. The current study seeks to evaluate the accuracy of retrospective surveys, sleep diaries, and actigraphy compared to a single-channel EEG device in participants' homes over one week using a broad community sample.

Overview of Sleep

Sleep is an essential, behaviorally modifiable biological function typically associated with a sitting or lying posture, closed eyes, and reduced motor activity, perception, and consciousness (Carskadon & Dement, 2011). However, sleep is multidimensional and can be

assessed at many levels of analysis (Buysse, 2014). Further, sleep is not a static construct from night to night, but rather demonstrates considerable intra-individual variability.

Sleep Dimensions

"Poor sleep" has been associated with numerous poor physical and mental health outcomes. However, the definition of "poor sleep" has been poorly and incompletely defined and is often inadequately assessed in the literature (Buysse, 2014). Three primary domains of sleep are typically thought to impact health and safety: duration, quality, and timing (Czeisler, 2015). However, interactions within and across these dimensions make obtaining a clear picture of sleep health quite complicated.

Sleep duration, or total sleep time, refers to the amount of sleep obtained in a 24-hour period. Total sleep time is typically measured by PSG, actigraphy, sleep diary, or retrospective questionnaire. Abnormal total sleep time (e.g., <7 hours or >9 hours) is associated with poor health outcomes such as cardiovascular diseases, diabetes, and mortality (Alvarez & Ayas, 2004). Sleep disorders/complaints associated with short total sleep time include insufficient sleep (Hublin, Kaprio, Partinen, & Koskenvuo, 2001) and insomnia (Vgontzas, Fernandez-Mendoza, Liao, & Bixler, 2013), and sleep disorders associated with long total sleep time are typically secondary to a medical or psychological disorder (Patel, Malhotra, & Gottlieb, 2006). Total sleep time has been the most frequently studied aspect of sleep in the medical and psychological literatures (Cappuccio, D'Elia, Strazzullo, & Miller, 2010).

Sleep quality, typically defined as one's perception of the "restfulness" of sleep, is largely a subjective and complex phenomenon (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989). Sleep quality is most often assessed by retrospective questionnaire, but can also be

assessed via PSG, actigraphy, and some forms of sleep diaries. Notably, although objective measurements can correlate highly with subjective sleep quality, they cannot fully define it due to the subjective aspects of sleep quality (Buysse et al., 1989). However, subjective and objective assessments of sleep quality likely provide distinct, yet overlapping information about an individual's heath status. Parameters associated with sleep quality are sleep efficiency, movement during sleep, number and length of awakenings, "depth" of sleep (i.e., sleep stages), fragmentation, and self-rated restorativeness or restfulness. One parsimonious metric that captures much of the objective aspects of sleep quality is sleep efficiency, which is the percentage of time that the intended sleep period is filled with sleep. Sleep efficiency is calculated based on a set of sleep parameters, typically dividing total sleep time by time in bed to obtain a percentage between 0 – 100. Depending on what information is available, total sleep time is calculated by subtracting total wake time (sleep onset latency, wake after sleep onset, and terminal wakefulness) during the intended sleep period (i.e., time in bed; Spielman, Saskin, & Thorpy, 1987). For example, sleeping for 6 hours but spending 10 hours in bed trying to sleep would be 60% sleep efficiency. Poor sleep efficiency is associated with poor health outcomes including increased mortality, metabolic syndrome, diabetes, hypertension, coronary heart disease, and depression (see Buysse, 2014 for citations). Sleep disorders associated with poor sleep quality or efficiency include insomnia (Spielman et al., 1987), circadian rhythm disorders (Reid & Zee, 2009), sleep apnea (Kimoff, 1996), and periodic limb movement disorder (Wetter & Pollmächer, 1997).

Timing of sleep refers to the placement of sleep within a given day (i.e., 24-hour period). Sleep timing can be considered both a trait and a state phenomenon, meaning that individuals may not sleep during their preferred schedule. Trait sleep timing, or circadian preference, is

typically measured with self-report questionnaires (e.g., Composite Scale of Morningness; Smith, Reilly, & Midkiff, 1989) or physiological measures of circadian rhythm (e.g., dim light melatonin onset; Lewy & Sack, 1989). State sleep timing is typically measured by PSG, actigraphy, sleep diary, or retrospective questionnaire, using parameters such as sleep onset, sleep offset, and circadian midpoint. Abnormal sleep timing (i.e., majority of sleep occurs outside of 9pm – 4am window) is associated with poor health outcomes including increased mortality, coronary heart disease, metabolic syndrome, diabetes, and accidents (Costa, 1996). Sleep disorders associated with abnormal sleep timing include circadian rhythm disorder (Reid & Zee, 2009) and shift work disorder (Schwartz & Roth, 2006).

Similar to wake, there is considerable variability in sleep both between and within individuals (i.e., night-to-night; Van Dongen, Vitellaro, & Dinges, 2005). This is due to a multitude of biological, psychological, and social factors that can differentially affect sleep on an hourly, nightly, weekly, monthly, seasonally, yearly, and lifetime basis. Because of this variability, a single-night snapshot of an individual's sleep does not yield an accurate estimate of one's habitual average total sleep time (Van Dongen et al., 2005).

Most sleep research has focused on inter-individual variability—primarily differences between groups or individuals in total sleep time and related causes and consequences (Bei, Wiley, Trinder, & Manber, 2016; Van Dongen et al., 2005). This research seeks to understand why certain people may have different sleep patterns than others (e.g., Do people with lower socioeconomic status have worse sleep?), or how sleep may differentially affect outcomes (e.g., Do people with poor sleep quality have an increased risk of cardiovascular disease?). Examining this inter-individual variability provides a good understanding of the factors that may contribute to or result from different sleep patterns. However, studying differences between individuals

does not necessarily provide information about how individuals change over time. Further, considerably less research has focused on intra-individual variability in multi-dimensional components of sleep—differences in timing, duration, and efficiency—and their relationship to poor outcomes (Bei et al., 2016; Van Dongen et al., 2005).

Almost no research to date has examined intra-individual variability in multi-dimensional assessments of sleep—differences in sleep across contexts or time periods for a single individual (Bei et al., 2016; Van Dongen et al., 2005). Intra-individual variability research seeks to understand the antecedents and consequences of sleep patterns for an individual (e.g., Do depressive symptoms during a day lead to poor sleep that night? Does a week of fragmented REM sleep increase inflammation the following week?) Examining intra-individual variability provides a direct observation of changes in sleep over context and time, which allows for the identification of the most influential determinants of the processes and systems underlying sleep health. Further, assessing intra-individual variability provides information about temporal sequencing and limits the effect of confounding variables which allows for stronger causal inferences. Finally, some have suggested inter-night variability in sleep (e.g., consistency of bedtime and waketime) may be a useful variable to examine in itself (Vallieres, Ivers, Bastien, Beaulieu-Bonneau, & Morin, 2005).Therefore, examining both inter- and intra-individual variability is crucial for developing targeted interventions and improving outcomes.

In order to understand both inter- and intra-individual variability, sleep characteristics must be accurately measured across seven or more days (Van Someren, 2007; Wohlgemuth, Edinger, Fins, & Sullivan, 1999). Sleep characteristics measured across less than seven days do not yield acceptable levels of reliability (i.e., temporal stability). The reliability of the measurement also depends on measurement method, sleep disorders, setting, and sleep parameter

of interest. Actigraphy and PSG measures of total sleep time alone require 7-11 days and sleep diaries require 2 weeks to achieve adequate stability (Van Someren, 2007; Wohlgemuth et al., 1999). In general, normal sleepers in a home environment demonstrate the greatest temporal stability in total sleep time and individuals with insomnia sleeping in the home environment demonstrate the lowest temporal stability (Wohlgemuth et al., 1999). However, this finding is moderated by sleep parameter and measurement method. For example, at 7 days of measurement, sleep onset latency show relatively high temporal stability for normal sleepers and individuals with insomnia when measured with PSG across lab and home settings, but low temporal stability for normal sleepers at home when measured with sleep diaries (Wohlgemuth et al., 1999). In sum, accurate and reliable measurement of sleep parameters requires at least 7 days of measurement and may vary considerably based on 1) sleep component measured and 2) between- and within-subjects factors.

Sleep Measurement

Sleep parameters are measured via objective (e.g., actigraphy, polysomnography), subjective (e.g., retrospective survey, sleep diaries), or combined (e.g., actigraphy and sleep diary) methods. Each measurement method has associated benefits and limitations. Further, these different methods provide data varying in type and quality.

Polysomnography (PSG) is an objective, prospective measure of total sleep time that, at a minimum, uses electroencephalography (EEG) to assess brain states and thereby determine sleep/wake. PSG is considered the “gold standard” of objective sleep measurement because it provides high-quality, high-frequency sleep staging data. However, there are multiple formats of PSG with an array of drawbacks and benefits.

Full or diagnostic PSG provides the most comprehensive assessment of brain and behavior during sleep and typically requires, at minimum, 3-lead EEG (e.g., O1 and O2, C3 and/or C4), chin electromyography (EMG), electrooculography (EOG), electrocardiogram (EKG), nasal/oral airflow, thoracic effect, abdominal effort, oxygen saturation, and body position (Kushida et al., 2005). Full PSG is used primarily for assessment of intrinsic sleep disorders like sleep-related breathing disorders, restless legs syndrome or periodic limb movement disorder, REM behavior disorder, and narcolepsy (Kushida et al., 2005). Depending on the aim of the study, certain elements may or may not be included.

Although in-lab PSG has traditionally been considered the "gold standard" of objective sleep measurement, it has substantial limitations. A primary limitation of in-lab PSG is the financial burden—a full night of PSG costs, on average, $1,190 per participant (Chervin, Murman, Malow, & Totten, 1999). Further, an in-lab PSG requires 8-12 hours of labor from a trained technician, and a physical lab space equipped for one or more sleeping participants with accompanying monitoring equipment. The equipment is uncomfortable and cumbersome, often requiring one or more nights for the participant to adapt and sleep normally. Once data is collected, it must be scored by trained technologists, which typically costs an additional sum and takes about 2 hours per record (Malhotra et al., 2013). Although PSG data is high-quality objective data, it cannot provide information about the participant's subjective experience, and therefore is not necessarily the "gold standard" for every type of sleep assessment. For example, experts have recommended full PSG is not an essential component of clinical assessment of insomnia, and instead emphasize a focus on clinical interview and sleep diaries (Buysse, Ancoli-Israel, Edinger, Lichstein, & Morin, 2006). Finally, data achieved by in-lab PSG, although high-quality, reflects the sleep of an individual sleeping in a laboratory environment and not in their

typical environment. This affects the ecological validity of the method and may limit study conclusions.

Ambulatory PSG has been able to partially mitigate some of the challenges posed by in-lab PSG but still suffers similar limitations. Ambulatory PSG, or in-home sleep testing, typically involves a participant coming to a laboratory for a reduced PSG hookup (e.g., EEG, EOG, EMG) by a trained technician, and then going home to sleep for the night. Every night of testing, the participant must return to the lab to be hooked up to equipment again. Ambulatory PSG units are typically more lightweight and portable than traditional PSG systems, but the cost and discomfort of the equipment are similar. The burden of scoring the collected data remains. Although automatic scoring systems have been developed, these systems still require a trained technician to hook up the full array of PSG sensors (Kaplan, 2014). In short, although home sleep testing can reduce the amount of technician labor and improve ecological validity compared to in-lab PSG, many of the limitations remain.

In response to the limitations of in-lab and ambulatory PSG, single-channel EEG devices (i.e., 3 electrodes) with automatic data scoring capabilities have recently emerged. The Zmachine Insight+ (General Sleep Inc.) is a portable, single-channel, EEG recording device that costs approximately $600 per unit. In comparison to the 10+ electrodes applied by a trained technician required for full PSG, the Zmachine uses only three electrodes that can be easily applied by the participant in their own home. The placement of the electrodes for the Zmachine (behind each ear and on the back of the neck) allows for minimal participant discomfort during sleep. Because participants do not have to return to the lab each night of a study for equipment hookup, the Zmachine can be used for upwards of three weeks without researcher interference. The

Zmachine offers high quality sleep staging data, reduced participant burden, reduced cost, and greater ecological validity compared to in-lab and ambulatory PSG.

Despite the promising features of the Zmachine, published research validating the Zmachine and data scoring algorithm is extremely limited. The first published study of the Zmachine (R. F. Kaplan, Wang, Loparo, Kelly, & Bootzin, 2014) examined the validity of an algorithm processing single-channel EEG data from the differential-mastoid location to determine sleep and wake. Participants were 99 healthy men and women aged 18-60 years, with or without a sleep complaint. Compared to full PSG rated by 2-4 scorers, the Zmachine demonstrated 96% sensitivity and 93% specificity for sleep-wake detection. Agreement between PSG and Zmachine for sleep parameters was also high: total sleep time $r = 0.95$, sleep efficiency $r = 0.93$, sleep onset latency $r = 0.96$, and wake after sleep onset $r = 0.89$.

Actigraphy is an objective measure of total sleep time that uses a wrist-worn accelerometer to capture motion and sometimes light to determine sleep/wake. Computer software uses an algorithm to analyze activity and estimate sleep parameters such as total sleep time, sleep onset latency, number of awakenings, wake after sleep onset, and terminal wakefulness (Ancoli-Israel et al., 2003). Actigraphy has a moderate up-front cost (e.g., $1000 - $5000 per actigraph and scoring equipment) but no per-study associated costs. Actigraphy is an ambulatory measurement device that allows for measurement over days, weeks, or months with minimal user burden (Ancoli-Israel et al., 2003).

Actigraphy provides several benefits over other types of sleep measurement. Actigraphy is an unobtrusive measure that does not substantially interfere with an individual's routine which allows for increased ecological validity compared to full PSG (de Souza et al., 2003). This reduced burden also allows for long periods of data collection (e.g., several months), which can

provide more accurate information about habitual sleep habits compared to full PSG. Actigraphy offers increased accuracy over subjective measures and can provide data about participant nighttime movement/activity unlike single-channel EEG. Although costlier than sleep diaries and surveys, actigraphy is substantially less expensive than PSG. Finally, actigraphy data is less time-consuming to collect and score compared to PSG (R. F. Kaplan et al., 2014).

Limitations of actigraphy primarily stem from concerns regarding comprehensiveness and accuracy of collected sleep data. Unlike PSG, actigraphy does not capture sleep staging data. Actigraphy total sleep time estimates correlate well (*r*s .70 – .90) with polysomnography in a laboratory setting (Ancoli-Israel et al., 2003; Sadeh & Acebo, 2002), but actigraphy has much greater sensitivity (i.e., sleep detection) than specificity (i.e., wake detection; Ancoli-Israel et al., 2003). This renders actigraphy data less accurate and useful in sleep-disordered populations because these individuals are likely to have more periods of wakefulness during the night that actigraphy is unable to detect (Ancoli-Israel et al., 2003). Related to this, a further concern is that actigraphy typically overestimates total sleep time compared to PSG by an average of 0.8 hours (Lauderdale, Knutson, Yan, Liu, & Rathouz, 2008).

Accuracy of actigraphy can be affected by type of device, type of computer software and algorithm, and participant adherence. Several mainstream devices used to assess total sleep time and other sleep variables differ on factors such as type of accelerometer or software compatibility (Ancoli-Israel et al., 2003). Devices from different manufacturers estimate total sleep time differently (Cellini, Buman, McDevitt, Ricker, & Mednick, 2013), although agreement between devices is generally high (>90%; Ancoli-Israel et al., 2003). Algorithms for scoring actigraphy data have not been standardized, and few published studies have examined differences between different algorithms (Ancoli-Israel et al., 2003). Participant adherence (e.g.,

completing concurrent sleep diaries, pressing a button on the watch indicating intended sleep time and final wake time, wearing the device consistently), device functioning (e.g., battery charge) and environmental factors (e.g., bed partners, pets, bedroom light exposure) can also impact accuracy (Sadeh & Acebo, 2002). The accuracy of actigraphy may be improved when it is used in combination with a subjective measure (e.g., sleep diaries; Kushida et al., 2001). Limitations of actigraphy yield it less useful in research or clinical situations requiring high specificity or sleep staging data. See Table 1 for a summary of studies comparing actigraphy to PSG.

Sleep diaries are a subjective measure of total sleep time that uses structured self-report to capture estimated periods of sleep/wake. Sleep diary information can be collected via paper-and-pencil means, via voicemails left by participants, or through electronic data capture software. Sleep diaries are a low- or no-cost ambulatory measurement that allows for indefinite periods of measurement (given user participation).

Sleep diaries are considered the gold standard of subjective sleep measurement (Carney et al., 2012). Sleep diaries yield most of the commonly examined sleep parameters when comprehensive forms are used (e.g., Consensus Sleep Diary; Carney et al., 2012). Sleep diaries allow for high ecological validity and can allow for even longer periods of data collection compared to actigraphy. Sleep diaries provide valuable information about an individual's subjective experience of sleep, which is a key feature of several sleep disorders (e.g., insomnia), and may be differentially related to some health outcomes compared to objective measures (Kurina et al., 2013). Sleep diaries offer improved accuracy and more comprehensive data compared to surveys, making them a favorable alternative.

Table 1

Summary of Studies Comparing Total Sleep Time and Sleep Efficiency Measured by Polysomnography and Actigraphy

First Author	Year	N	Population	Actigraphy	PSG	# Nights	Setting	correlation for TST	Mean diff. ±standard error for TST	correlation for SE%	Mean diff. ±standard error for SE%
Beecroft	2008	12	mechanically ventilated patients	MiniMitter AW64 Actiwatch	Dataq WinDaq	1	ICU	*r* = .14	--	*r* = .18	--
Blackwell	2008	68	Older adult women	Ambulatory Monitoring Inc. Sleepwatch-O	Compumedics Siesta	1	Home	*ICC* = .76	44.2 ±29.1	*ICC* = .61	9.8 ±6.5
Jean-Louis (study 1)	2001	39	post-menopausal women	Actillume actiwatch	Oxford Medilog 9200	1	Home	*r* = .98	21.0 ±10.5	*r* = .91	4.0 ±2.0
Jean-Louis (study 2)	2001	11	healthy adults	Actillume actiwatch	VIASYS Somnostar	5	lab	*r* = .92	5.0 ±5.5	*r* = .69	1 ±0.9
Jean-Louis	2001	5	healthy adults	Actillume actiwatch; Mini Motionlogger	Somnostar	5	Lab	*r* = .85	10 ±6.5	*r* = .69	2.0 ±2.0
Kaplan	2012	54	adults with bipolar and controls	AW64	Siesta	2	Lab	*r* = .92	--	*r* = .49	--
Lichstein	2006	57	Insomnia	AW64	Respironics' Alice 3	1	Lab	*r* = .70	--	*r* = 0.43	--
Matthews	2018	223	Community middle-aged and older adults	Respironics Actiwatch 16	Siesta	2	Home	*r* = .66	-7.2 ±51.0	--	--
McCall	2012	54	adults with insomnia and depression	AW64	SomnoStar	1	Lab	*r* = .54	12.8 ±7.5	*r* = 0.48	0.1 ±6.0
Rupp	2011	29	sleep-deprived community adults	Basic Mini-Motionlogger; AW64	not specified	1	Lab	--	14.7 ±2.9; 39.9 ±2.6	--	2.23 ±0.44; 6.05 ±0.39
Sanchez-Ortuno	2010	31	adults with insomnia; community adults	Mini Mitter	not specified	6	Lab and home	*r* = .92; *r* = .93	--	*r* = 0.77; *r* = 0.81	--
Signal	2005	21	Flight crew members	Actiwatch	Medcare Embla	1	Hotel	*r* = .86	20.1 ±97.9	*r* = 0.54	7.0 ±43.0
Siversten	2004	34	adults with insomnia (pre; post-tx)	Actiwatch Plus	Embla A10	2; 1	Lab	--	--	*r* = 0.54; *r* = 0.66	--
Taibi	2013	16	community women with insomnia	AW64	Embla Somnologica	8	Lab		89.2		18.9

Note. TST = total sleep time; SE% = sleep efficiency percentage.

Table 2

Summary of Studies Comparing Total Sleep Time and Sleep Efficiency Measured by Polysomnography and Subjective Measures (i.e., Diary, Survey)

First Author	Year	N	Population	Diary/Survey	PSG	# Nights	Setting	correlation for TST	Mean diff. ±standard error for TST	correlation for SE%	Mean diff. ±standard error for SE%
					Diary						
Kaplan	2012	54	adults with bipolar and controls	"standard sleep diaries"	Compumedics Siesta	2	Lab	*r* = .71	--	*r* = .48	--
Lichstein	2006	57	Insomnia	adapted from Lichstein, Riedel & Means (1999)	Respironics' Alice 3	1	Lab	*r* = .59	--	*r* = .48	--
Matthews	2018	223	Community middle-aged and older adults	modified Pittsburgh sleep diary	Siesta	2	Home	*r* = .56	15.6 ±67.2	--	--
McCall	2012	54	adults with insomnia and depression	unspecified	VIASYS SomnoStar	1	Lab	*r* = .33	54.5 ±14.1	--	--
					Survey						
Matthews	2018	223	Community middle-aged and older adults	Pittsburgh Sleep Quality Index Question	Siesta	2	Home	*r* = .14	22.2 ±87.0	--	--
Signal	2005	21	Flight crew	"post-sleep subjective questionnaire"	Embla	1	Hotel	*r* = .84	12.9 ±111.6	*r* = .41	-24.2 ±47.0

Note. TST = total sleep time; SE% = sleep efficiency percentage.

Limitations of sleep diaries primarily stem from participant adherence/burden and accuracy. Sleep diaries have demonstrated acceptable agreement with PSG in a community sample ($\kappa = .87$; Rogers, Caruso, & Aldrich, 1993), but limited validation research exists. Sleep diaries tend to overestimate total sleep time overall, but sleep diaries perform differently in individuals with and without sleep disorders. Individuals with sleep disorders like insomnia tend to underestimate their total sleep time, whereas normal sleepers tend to overestimate their total sleep time.

Accuracy of sleep diaries can be affected by factors such as type of sleep diary and participant compliance (e.g., filling out sleep diaries every morning upon awakening). Although a standardized research sleep diary exists (Carney et al., 2012), this sleep diary is not used consistently across studies. Additionally, one study in a related field (i.e., chronic pain) demonstrated daily-collected diaries are associated with 93.6% participant compliance compared to 10.9% with weekly-collected diaries (Stone, Shiffman, Schwartz, Broderick, & Hufford, 2003). Low compliance may affect accurate reporting due to expectations about sleep, recall bias, or memory impairments (Martin & Hakim, 2011). These limitations indicate use of a concurrent objective measure, to assess subjective difficulties with sleep diaries, may be warranted. See Table 2 for a summary of studies comparing sleep diaries to PSG.

Surveys and sleep questionnaires (e.g., Pittsburgh Sleep Quality Index; Buysse et al., 1989) are subjective, retrospective measures of total sleep time that can use 1-20 questions (e.g., “On average, how many hours do you sleep per night?”) to assess estimated periods of sleep/wake. Surveys are a no-cost measurement that can capture long periods of retrospective sleep data at a single time point.

The primary benefits of surveys are the convenience, low cost, and simplicity of their use. Typically, survey questions of sleep focus on total sleep time and use three or fewer questions, making them easy to insert into larger single time-point studies with relatively low participant burden. Similarly, answers to survey questions require minimal scoring burden. Further, survey measures can be used to capture long periods of time in a very brief assessment period. Surveys are commonly used in large epidemiological studies that assess many variables (Schwarz, 2007), and surveys are the most frequently used measure of total sleep time in existing research (Cappuccio et al., 2010).

Despite their frequent use, surveys are a poor choice for accurately and comprehensively assessing sleep. Retrospective survey estimates of total sleep time are only moderately correlated (r = 0.45) with objective measures of sleep. One study (Lauderdale et al., 2008) found single-point retrospective estimates tend to be systematically biased towards over-reporting total sleep time compared to an objective measure (i.e., actigraphy). This effect was more pronounced for individuals reporting shorter total sleep time (e.g., 1.3 hours overestimated for 5 hours of sleep vs. 0.3 hours overestimated for 7 hours of sleep). Furthermore, surveys often give ordinal (e.g., "Less than 5 hours, 5-6 hours, etc.) response sets, which greatly limits response variability and therefore power. Surveys typically only ask questions about total sleep time and/or sleep disturbances and fail to capture other important aspects of sleep (e.g., timing, variability). See Table 2 for a summary of studies comparing survey measures to PSG.

Survey measures of sleep are not standardized across studies, and varying question formats can make comparisons difficult and impact data validity and reliability. First, using two questions to differentiate between weekend and weekday sleep, as opposed to a single question assessing a daily average, yields higher average total sleep time (Lauderdale, 2014). Second,

assessing total sleep time via an hour estimate (e.g., “How many hours do you sleep per night?”) versus bedtime and risetime estimates (e.g., “What time do you typically go to bed/wake up?”) can yield consistently different results. A recent meta-analysis examining total sleep time and mortality found studies using an hour estimate demonstrated a U-shaped relationship between total sleep time and overall mortality, whereas studies that used bedtime and risetime estimates to calculate average total sleep time failed to find the U-shaped relationship (i.e., found a linear relationship or no relationship). This indicates there may be a systematic bias across studies determined by the way total sleep time questions are asked (Kurina et al., 2013). Taken together, these findings suggest the ease of using survey measures of sleep may come at the price of accuracy, and a more thorough method of assessment is indicated.

Impact of Insomnia on Measurement

Previous research has demonstrated subjective estimates of sleep variables are impacted by individual characteristics, most notably insomnia. Manconi et al. (2010) compared restrospective sleep estimates of total sleep time (“How long, on average, do you think you slept this night?”) to PSG over one night in adults with (n =159) and without (n = 288) insomnia. Individuals with insomnia underestimated their total sleep time by about 157 minutes on average (95% limits of agreement = 240 minutes) whereas individuals without insomnia overestimated their total sleep time by about 20 minutes on average (95% limits of agreement = 530 minutes). Kay, Buysse, Germain, Hall, and Monk (2015) compared daily sleep diaries to actigraphy over seven nights in older adults with (n = 63) and without (n = 51) insomnia. They found individuals with insomnia overestimated their nighttime wakefulness (SOL + WASO) by about 36 minutes on average and those without insomnia underestimated nighttime wakefulness by about 19

minutes on average. Although actigraphy-derived total sleep time differed minimally between individuals with and without insomnia, sleep diary-derived total sleep time was substantially lower for individuals with insomnia. Taken together, these studies suggest individuals with insomnia underestimate total sleep time on subjective measures compared to objective measures whereas individuals without insomnia are fairly accurate or may slightly overestimate total sleep time.

There are few studies comparing accuracy of actigraphy-derived total sleep time estimates with PSG in individuals with and without insomnia. Sánchez-Ortuño, Edinger, Means, and Almirall (2010) compared total sleep time assessed by actigraphy and PSG across 3 nights in the lab in a sample of adults with (n=31) and without (n=31) insomnia. On average, actigraphy estimates were higher than PSG estimates of total sleep time for both individuals with insomnia (8.3 minutes) and without insomnia (4.9 minutes). McCall and McCall (2012) compared total sleep time assessed by actigraphy and PSG in a single-night in-lab study with N=54 middle-aged men and women with depression and insomnia. They found actigraphy overestimated sleep by, on average, 12.8 minutes (SD = 7.5) compared to PSG. Although information on which to base comparisons is limited, these studies study suggests insomnia status does not result in substantial differences between PSG- and actigraphy-derived estimates of total sleep time.

Gaps in the Literature

Validation of global sleep/wake data for actigraphy, sleep diaries, and surveys compared to PSG are limited in scope (see Table 1). Previous studies of actigraphy validation have primarily assessed total sleep time over 1-7 days in a laboratory environment, or over 1-3 nights in the home environment. These studies are important, but they cannot capture the night-to-night

variability of sleep/wake in an ecologically valid setting. Previous research suggests a minimum of 7 days of assessment of sleep is necessary to obtain an accurate estimate of habitual total sleep time (Van Someren, 2007; Wohlgemuth et al., 1999). There are few previous studies of sleep diary validation, and the existing studies have not used PSG as the gold standard, or have focused on specialized populations like individuals with bipolar disorder. Finally, there are very few previous studies of retrospective survey or sleep questionnaire validation and existing studies have not used PSG as a gold standard. To our knowledge, no studies have examined measurements of sleep timing, duration, and efficiency using actigraphy, sleep diary, or retrospective sleep questionnaires against PSG for the recommended 7-night duration in an ecologically valid setting (i.e., the individual's typical sleep environment).

The Current Study

The current study examined sleep duration (total sleep time), sleep efficiency (percentage of intended sleep time that is actually filled with sleep), and sleep timing (circadian midpoint) across seven nights using four measurement methods (i.e., Zmachine, actigraphy, sleep diary, retrospective survey) in a community sample. The primary hypothesis was that actigraphy and sleep diaries would demonstrate acceptable agreement with Zmachine for circadian midpoint and sleep efficiency, but not for total sleep time, and a sleep survey would fail to demonstrate acceptable agreement with Zmachine for sleep timing, duration, or efficiency. The secondary hypothesis was that insomnia severity would be a predictor of total sleep time differences between a) actigraphy and Zmachine, with greater insomnia severity predicting higher actigraphy-assessed total sleep time compared to Zmachine, b) sleep diary and Zmachine, with greater insomnia severity predicting lower sleep diary-assessed total sleep time compared to

Zmachine, and c) survey and Zmachine, with greater insomnia severity predicting lower survey-assessed total sleep time compared to Zmachine.

CHAPTER 2

METHOD

Participants

Participants for the proposed study were recruited broadly in the Denton County area. Recruitment efforts aimed to increase generalizability by including a wide age range and a diverse racial/ethnic breakdown similar to the community. Recruitment materials directed interested individuals to informed consent and a brief online screening survey that collected contact information and assessed the following inclusion criteria: a) willingness to participate for at least 7 days, b) ability to travel to the Sleep and Health Research lab, c) English language fluency, d) over the age of 18, e) had a phone number at which they could be regularly reached, and f) had regular (daily) internet and personal email access.

Data were collected from February 2017 through August 2017. Initially, 120 people expressed interest in the study, 101 completed the screening questionnaire, and 87 completed the baseline questionnaire. A total of 81 participants attended the first lab appointment and completed some measures and a final $N = 80$ were included in any analyses. Participant characteristics are presented in Table 3. The majority of participants were female, non-Hispanic White, married or in a relationship, well-educated, and employed.

Procedures

All procedures were approved by the University of North Texas Institutional Review Board prior to the start of data collection. After completing the brief screening measure, eligible participants were contacted and given the opportunity to complete the one hour of baseline

measures online at home via a secure online data collection tool (REDCap). Participants were then scheduled for their first in-person appointment in the Sleep and Health research laboratory.

Table 3

Sample Participant Characteristics

		Mean (*SD*)	*n* (%)
Age (y; range 19-69)		32.7 (10.1)	
Sex	Male		30 (37.5)
	Female		50 (62.5)
Race/ethnicity	NH Black		3 (3.8)
	NH White		68 (85.0)
	Asian		4 (5.0)
	Biracial/Other		5 (6.2)
Married/Committed relationship	Yes		58 (72.5)
	No		22 (27.5)
Educational attainment	High school or less		2 (2.5)
	≤4 years post-high school education		36 (45.0)
	>4 years post-high school education		42 (52.5)
Employment status	Full time		49 (61.3)
	Part Time		19 (23.8)
	Retired/Unemployed		12 (15.1)
STAI trait anxiety		39.8 (10.2)	
PANAS-X	General Positive Affect	30.5 (7.8)	
	General Negative Affect	18.9 (7.1)	
PCL-5 posttraumatic stress symptoms*		13.8 (15.3)	
	>33		10 (13.3)
	≤33		65 (86.7)
PROMIS Sleep-Related Impairment (SF)		20.6 (5.9)	
STOP sleep apnea screening score	>1		20 (25)
	≤1		60 (75)
ISI insomnia severity		8.9 (5.7)	
Insomnia diagnosis by clinical interview	Yes		24 (30)
	No		56 (70)

Note: NH = non-Hispanic; *for PCL-5,* n = *75 due to missing responses. For all other descriptives,* n = *80.*

During this appointment, participants were asked to provide a copy of their drivers' license and contact information for two individuals who could be contacted if the participant could not be reached and did not return equipment. Participants were trained in completion of the Zmachine via videos provided by the equipment manufacturer and hands-on demonstration. Participants were trained in use of actigraphy via verbal instruction from the research assistants and hands-on demonstration. Participants were trained in use of daily sleep diaries via a sample survey sent to their internet-enabled device and hands-on demonstration. Participants were then given a Zmachine, actigraph, paper sleep diaries (in case of website malfunction) and written instructions for all items (including online sleep diary instructions). Finally, participants' understanding of study procedures was verified via a brief written quiz and their second in-lab appointment was scheduled.

Participants used the Zmachine, actigraph, and sleep diary in their typical sleep environment for 7 days. After the first study night, research assistants messaged participants to ensure that equipment worked properly and check if they had any questions. At the first lab visit, participants and research assistants mutually chose a time to receive the first sleep diary reminder. Participants received up to two additional reminders, at three-hour intervals, if they did not complete the sleep diary. Additionally, if they had not completed the diary by noon, research assistants messaged the participants to remind them to complete. After 7 days of data collection, participants returned to the lab to return the equipment and complete final measures including retrospective sleep surveys. The compensation offered for participation in the study was a) $20, b) a comprehensive report of the participant's sleep over the study duration and sleep disorders resources, and c) a decorative refrigerator magnet.

Measures

Retrospective Survey

The Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) a 19-item self-rated questionnaire designed to measure seven domains of sleep. Domain scores range from 0 (no difficulty) to 3 (severe difficulty), and combine to produce a global score of sleep quality ranging from 0-21. The original validation study established a cutoff score of >5 indicating a significant sleep disturbance (Buysse et al., 1989). In the current study, only questions that queried average total sleep time, sleep efficiency and variables used in the assessment of circadian midpoint over the past week were used for analyses.

Sleep Diaries

The sleep diary is a subjective measure that asks participants to give an estimate of their sleep on the previous night (e.g., bed time, sleep onset latency, wake time). These variables allow for the calculation of additional sleep metrics (i.e., sleep efficiency, total sleep time, sleep onset latency, wake after sleep onset). In addition, the sleep diary included additional questions that queried nightmare frequency and severity. Sleep diaries have been shown to correlate moderately well with both polysomnography and actigraphy (Carney et al., 2012). Sleep diaries were collected using electronic data capture software (REDCap; Harris et al., 2009).

Actigraphy

Actigraphs are wrist-worn, wristwatch-like devices that use an accelerometer to capture motion as a proxy for activity. Computer software uses an algorithm to analyze activity and estimate sleep parameters such as total sleep time, sleep onset latency, number of awakenings,

wake after sleep onset, and terminal wakefulness (Ancoli-Israel et al., 2003). In the current study, the actigraphs used were Philips Respironics Actiwatch Spectrum devices, and data was analyzed with Respironics Actiware v6.0. Data was be scored by two trained scorers using an internally-developed scoring hierarchy, and discrepancies were resolved by a third scorer. Settings used for data export in Actiware were the following: low threshold (activity count: 10), 20 epochs inactivity for sleep onset/offset.

Zmachine

The Zmachine is an ambulatory device that processes a single channel of EEG data using information from two mastoid-placed electrodes and one neck-placed ground electrode. The Zmachine is capable of differentiating between wake, light sleep (stages N1 and N2), deep sleep (stage N3), and rapid eye movement sleep (R. F. Kaplan et al., 2014). The Zmachine electrodes are single-use and were self-applied by the participant 30 or more minutes prior to bedtime.

Insomnia Status and Severity

Insomnia status was determined via the Structured Clinical Interview for Sleep Disorders (SCISD; Taylor et al., 2018). Interviewers were undergraduate-level students trained to fidelity by the first author (JRD). In regression analyses, the Insomnia Severity Index (ISI) was used as a continuous measure. The ISI is a 7-item self-report measure that assesses perceived severity of insomnia (Bastien, Vallieres, & Morin, 2001). Each item uses a Likert scale ranging from 0 to 4 with higher scores indicating greater severity of insomnia symptoms. The items are summed to produce a total score (range 0–28). The ISI has good internal consistency in the current study (coefficient $\alpha = 0.88$) with item-total correlations ranging from $r = 0.39$-$.83$ (all ps < 0.001).

Data Analysis

The following sleep parameters were computed across all four measurement methods, when possible: circadian midpoint (bedtime – waketime/2), total sleep time (time in bed – [sleep onset latency + wake after sleep onset + terminal wakefulness]), and sleep efficiency (total sleep time/time in bed). For comparison to survey methods, data was averaged across the week for Zmachine, actigraphy, and diary. Averages were only calculated for a given measure if ≥5 days of data existed for that measure.

Hypothesis 1

In order to test whether actigraphy, sleep diary, and survey methods demonstrate acceptable agreement with Zmachine for sleep timing, duration, and efficiency, parameters were compared across measurement method using the Bland and Altman technique to examine systematic bias and agreement (Bland & Altman, 2010). One deficit in the sleep measure validation literature is the erroneous use of product-moment correlation coefficients (*r*) and other global indices to demonstrate agreement between two measures. Correlation coefficients are inadequate for assessing agreement for four primary reasons: a) high correlation between two methods indicates the *strength* of a relationship instead of the agreement between the methods; b) a difference in scale of measurement affects agreement but not correlation (e.g., a parameter measured consistently lower by one method [e.g., 400 min vs 500 min] could be perfectly correlated with a gold standard even though the two measures never agree), and therefore data that has low agreement can still demonstrate high correlations; c) correlation fluctuates depending on the range in the sample such that a wide-ranging parameter will inherently have a greater correlation than a narrow-ranging parameter; and d) significance testing is irrelevant in

testing the strength of agreement, as it is expected that two measures would have a certain degree of relationship (Bland & Altman, 2010). Bland and Altman (1999) state, although frequently done, it is inappropriate to assess agreement between measures using correlation, regression, comparison of means, structural equations, or intra-class correlation methods. Instead, plots of two methods against each other (prediction plots), plots of means against mean differences (Bland-Altman plots), and estimates of where 95% of differences between measures are expected to fall (limits of agreement) are suggested, as these items give information as to potential systematic bias and variability of estimates in addition to mean differences.

Analyses for Hypothesis 1 were conducted in R software version 3.1.3 (R Core Team, 2015) using the using the `MethComp` package version 1.22.2 (Carstensen, Gurrin, Claus Ekstrom, & Figurski, 2016). Separate analyses were used for each sleep parameter and each measure comparison. Data was first examined via two plots: a) prediction plots: gold standard on the X-axis against comparison method on the Y-axis with a line of equality (i.e., perfect agreement between the methods); and b) Bland-Altman plots, the mean of both methods ([method 1 + method 2]/2) on the X-axis against difference between methods (method 1 – method 2) on the Y-axis. These plots allow for visual examination of agreement between methods and detection of systematic or unsystematic bias. For comparison between Zmachine, actigraphy, and diary (repeated measures) a mixed model was used to estimate the 95% limits of agreement while controlling for nesting of repetitions within participants (data was considered "linked" or paired replicates; Carstensen, Simpson, & Gurrin, 2008).This method includes measure (i.e., Zmachine, actigraphy, diary) and participant as fixed effects, and the measure × participant interaction as a random effect.

For comparison between survey and the other measures' averages, traditional calculations

for Bland-Altman plots were created and visually examined (Bland & Altman, 1999). Given non-constant bias across levels of the construct (i.e., total sleep time, sleep efficiency, and circadian midpoint), recommendations by Carstensen (2010) were used to calculate coefficients that can be used to convert one method to another and prediction intervals rather than traditional limits of agreement.

Hypothesis 2

Analyses for Hypothesis 2 and descriptives were conducted in SPSS software version 21. In order to test whether insomnia severity was a predictor of differences between average actigraphy and Zmachine-assessed total sleep time, sleep diary and Zmachine, and survey and Zmachine, three hierarchical linear regressions were used. Demographics (gender, marital status, age, education level) were entered into the first block, psychosocial covariates (negative affect, anxiety) were entered into the second block, and ISI score was entered into the third block as predictors for all models. In the first model, the dependent variable was mean difference between Zmachine and actigraphy total sleep time, with positive numbers representing underestimation by actigraphy, and negative numbers representing overestimation by actigraphy compared to Zmachine. In the second model, the dependent variable was mean difference between Zmachine and sleep diary total sleep time. In the third model, the dependent variable was mean difference between Zmachine and survey total sleep time.

Power Analysis

There are few recommendations for power analysis in method comparison studies using Bland Altman analyses. Therefore, power analyses were calculated for repeated measures

ANOVA, within factors effects with a small effect size of $d = 0.15$, power of 0.80, alpha of .01 (to control for multiple comparisons), correlation among repeated measures of 0.43 (Lichstein et al., 2006) and 7 nights of repeated measures. This analysis indicated a total sample size of 70 participants was needed, but we included 80 to allow for 10-15% percent attrition and data loss. Carstensen (2010) argues power analysis calculations for measure comparison studies are irrelevant and instead recommends a sample of at least 50, with at least 3 days of measurement per person. Data collected far exceeds that recommendation, even given missing data.

CHAPTER 3

RESULTS

Data was cleaned by examining and applying necessary corrections for outliers, variable normality, and missing data in accordance with recommendations from Tabachnick and Fidell (2013). Participants were considered "completers" and therefore included for analysis if they had ≥5 days of usable data on at least two measures. One participant was excluded for <5 days of data on all measures. For Zmachine, 65 days were excluded for bad data and 19 days were missing, for a total of 84 excluded days for Zmachine (due to the nature of the data errors, circadian midpoint was retained for 19 days with bad data). A total of 11 participants had data removed for Zmachine due to <5 days of usable data (n = 69 with complete data). For actigraphy, 21 days were excluded for bad data and 15 days were missing, for a total of 36 days missing (circadian midpoint retained for 15 days). A total of 5 participants had data removed for actigraphy due to <5 days of usable data (n = 75 with complete data). For diary, 6 days were missing but no participants were removed (n = 80). For survey, 1 participant's data was completely missing, and two additional participants' data was removed for sleep efficiency due to impossible values given (n = 77 for all survey variables).

Correlations between Variables of Interest

Means for sleep variables of interest, divided by insomnia diagnosis and combined, are presented in Table 4. Preliminary correlations (Pearson r) were conducted to examine relationships between each variable of interest (total sleep time, sleep efficiency, circadian midpoint) as measured by Zmachine, actigraphy, diary, and survey (Table 5). Sole interpretation of these values as evidence for concordance or lack thereof would be inappropriate for reasons

described above. However, these values are presented to ease comparisons between the current results and prior research, which commonly report correlations between measures.

Table 4

Means (SD) for Individuals With and Without Insomnia and Whole Sample for Zmachine, Actigraphy, Diary, and Survey Measures of Total Sleep Time, Sleep Efficiency, and Circadian Midpoint

		Insomnia		No Insomnia		Combined	
		M (SD)	*N*	*M (SD)*	*N*	*M (SD)*	*N*
Total Sleep Time (h)	Zmachine	6.10 (0.81)	19	6.16 (0.74)	50	6.14 (0.75)	69
	Actigraphy	6.38 (1.03)	22	6.33 (0.76)	53	6.35 (0.84)	75
	Diary	7.08 (0.85)	24	7.08 (0.88)	56	7.08 (0.86)	80
	Survey	6.73 (1.14)	24	6.96 (1.10)	55	6.89 (1.20)	79
Sleep Efficiency (%)	Zmachine	80.57 (6.09)	19	81.47 (6.81)	50	81.22 (6.59)	69
	Actigraphy	80.17 (6.50)	22	82.18 (5.30)	53	81.59 (5.71)	75
	Diary	87.65 (7.49)	24	91.54 (4.10)	56	90.37 (5.59)	80
	Survey	79.76 (13.02)	22	86.66 (10.52)	55	84.69 (11.63)	77
Circadian Midpoint (t)	Zmachine	3.70 (1.26)	22	3.58 (1.26)	50	3.62 (1.25)	72
	Actigraphy	3.80 (1.26)	23	3.69 (1.17)	55	3.72 (1.19)	78
	Diary	3.79 (1.22)	24	3.76 (1.24)	56	3.77 (1.22)	80
	Survey	3.23 (1.43)	24	3.17 (1.36)	55	3.19 (1.37)	79

Generally, Zmachine demonstrated the highest correlations with actigraphy, followed by diary, and then survey across all variables. Correlations between Zmachine and actigraphy, diary, and survey were highest for circadian midpoint, followed by total sleep time, then sleep efficiency.

Hypothesis Testing

Hypothesis 1

Plots of Zmachine vs. actigraphy and diary for total sleep time, sleep efficiency, and circadian midpoint are presented in Figure 1. Bland Altman plots of Zmachine vs. actigraphy and diary for total sleep time, sleep efficiency, and circadian midpoint are presented in Figure 2.

Table 5

Correlations among Survey, Diary, Actigraphy and Single-Channel EEG Measures of Total Sleep Time, Sleep Efficiency, and Circadian Midpoint across 7 Nights at Home

		Actigraphy	Diary	Survey
Total Sleep Time (h)	Zmachine	.75***	.71***	.57***
	Actigraphy	--	.75***	.68***
	Diary		--	.63***
	Survey			--
Sleep Efficiency (%)	Zmachine	.25*	.18	.15
	Actigraphy	--	.18	.38**
	Diary		--	.42***
	Survey			--
Circadian Midpoint (t)	Zmachine	.99***	.98***	.94***
	Actigraphy	--	.99***	.93***
	Diary		--	.92***
	Survey			--

* $p < .05$. ** $p < .01$. *** $p < .001$.

These plots were examined and found to demonstrate constant bias (i.e., homoscedasticity of differences) at all levels so slope was fixed to 1 and bias and limits of agreement were calculated with mixed model adjustments for repeated measures (see Table 6; Carstensen et al., 2008). The 95% limits of agreement are displayed as thin blue lines on these plots, signifying that 95% of differences between measures are expected to fall between these lines. Mean difference, or bias, is represented as a thicker blue line on these plots, with positive values indicating underestimation of the second measure (e.g., actigraphy, diary) compared to the

first (Zmachine) and negative values indicating overestimation. For example, sleep efficiency was overestimated by sleep diary compared to Zmachine by 7.1%, on average (center image of Figure 2).

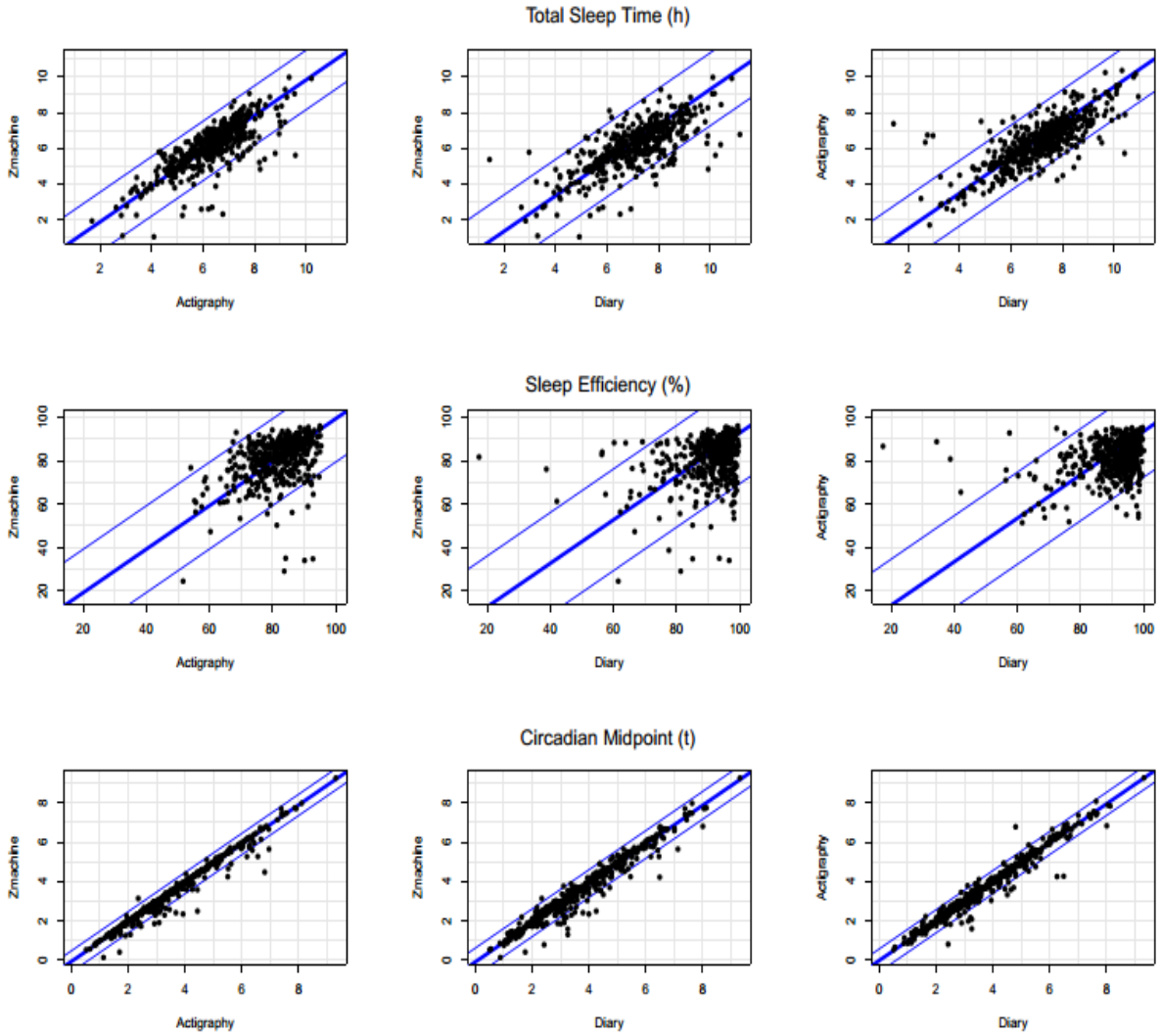

Figure 1. Prediction plots for repeated measures (7 days) of Zmachine- (EEG), actigraphy-, and sleep diary-assessed total sleep time, sleep efficiency, and circadian midpoint.

For total sleep time, the comparison between Zmachine and actigraphy demonstrated a slight tendency toward overestimation by actigraphy (8 minutes) with wide 95% limits of agreement (>3 hours). A similar pattern was seen for the comparison between Zmachine and

diary, with diary overestimating total sleep time by, on average, 40 minutes and wide limits of agreement (>4 hours).

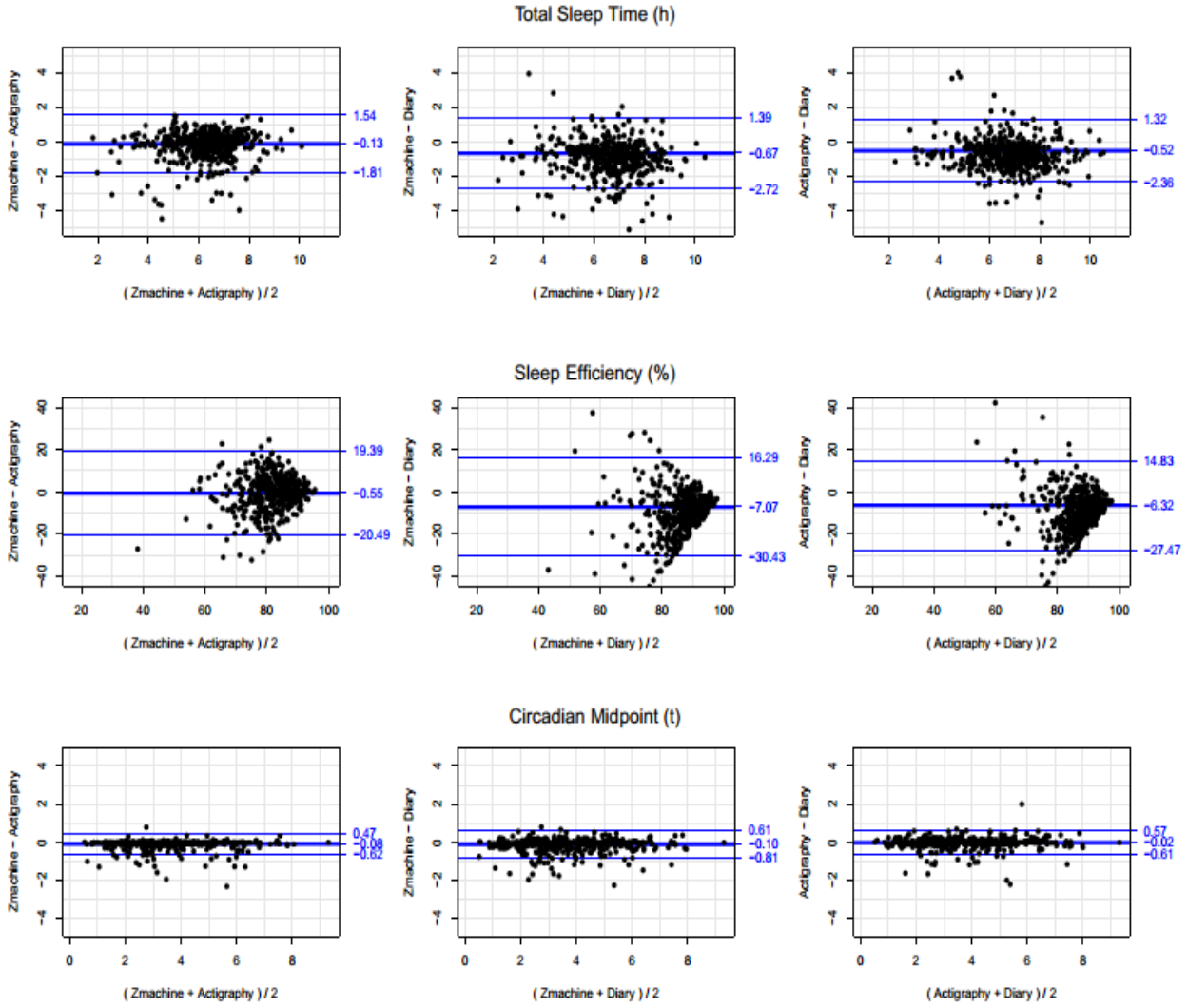

Figure 2. Bland-Altman plots adjusted for repeated measures (7 days) of Zmachine- (EEG), actigraphy-, and sleep diary-assessed total sleep time, sleep efficiency, and circadian midpoint with blue lines indicating mean bias and 95% limits of agreement.

For sleep efficiency, the comparison between Zmachine and actigraphy demonstrated almost no bias (<1% overestimation by actigraphy) and wide limits of agreement (>35%). The comparison between Zmachine and diary revealed an average overestimation of diary by approximately 7%, with wide limits of agreement (>45%). Notably, all comparisons for sleep

efficiency demonstrate substantial heteroscedasticity, with greater variability in differences at lower average values of sleep efficiency. In other words, days with greater sleep efficiency had more precise estimates compared to Zmachine.

For circadian midpoint, the comparison between Zmachine and actigraphy demonstrated almost no bias (<5 minutes overestimation by actigraphy) and rather narrow limits of agreement (<1.5 hours). The comparison between Zmachine and diary revealed a similar pattern with little bias (6 minutes) and rather narrow limits of agreement (<1.5 hours).

Bias, precision and 95% limits of agreement for survey vs. Zmachine are displayed in Table 6. Plots of Zmachine, actigraphy and diary vs. survey for averaged-across-week total sleep time, sleep efficiency, and circadian midpoint are presented in Figure 3. Bland Altman plots of averaged-across-week values for Zmachine, actigraphy, and diary versus survey for total sleep time, sleep efficiency, and circadian midpoint were examined and found to demonstrate non-constant bias (i.e., significant slope; all *p*s <.05). Therefore, plots were recomputed allowing differences to depend on averages in a linear rather than constant fashion. These plots are presented in Figure 4. Given non-constant bias, traditional limits of agreement could not be calculated (Carstensen, 2010). Instead coefficients for the regression line allowing linear conversion between each measure vs. survey were calculated using the `DA.reg` function of the `MethComp` R package (Carstensen et al., 2016). These coefficients are provided along with standard deviation of the prediction (precision) and 95% limits of prediction in Table 7. Conversion from survey to the other method can be achieved using the following formula from Carstensen (2010):

$$y_{2|1} = \alpha_{2|1} + \beta_{2|1} y_1 \pm 2 \times \text{s.d.}(y_{2|1}).$$

For total sleep time, the prediction interval for the comparison between Zmachine and survey was >2.5 hours. Additionally, the slope indicated that differences were not the same for individuals with high and low total sleep time. For individuals with lower total sleep time, survey was predicted to underestimate total sleep time compared to Zmachine. For individuals with higher total sleep time, survey was predicted to overestimate total sleep time compared to Zmachine. Practically, this means for an individual reporting 4 hours total sleep time on survey the predicted Zmachine value would be 4.4 hours and for an individual reporting 8 hours total sleep time on survey the predicted Zmachine value would be 6.8 hours.

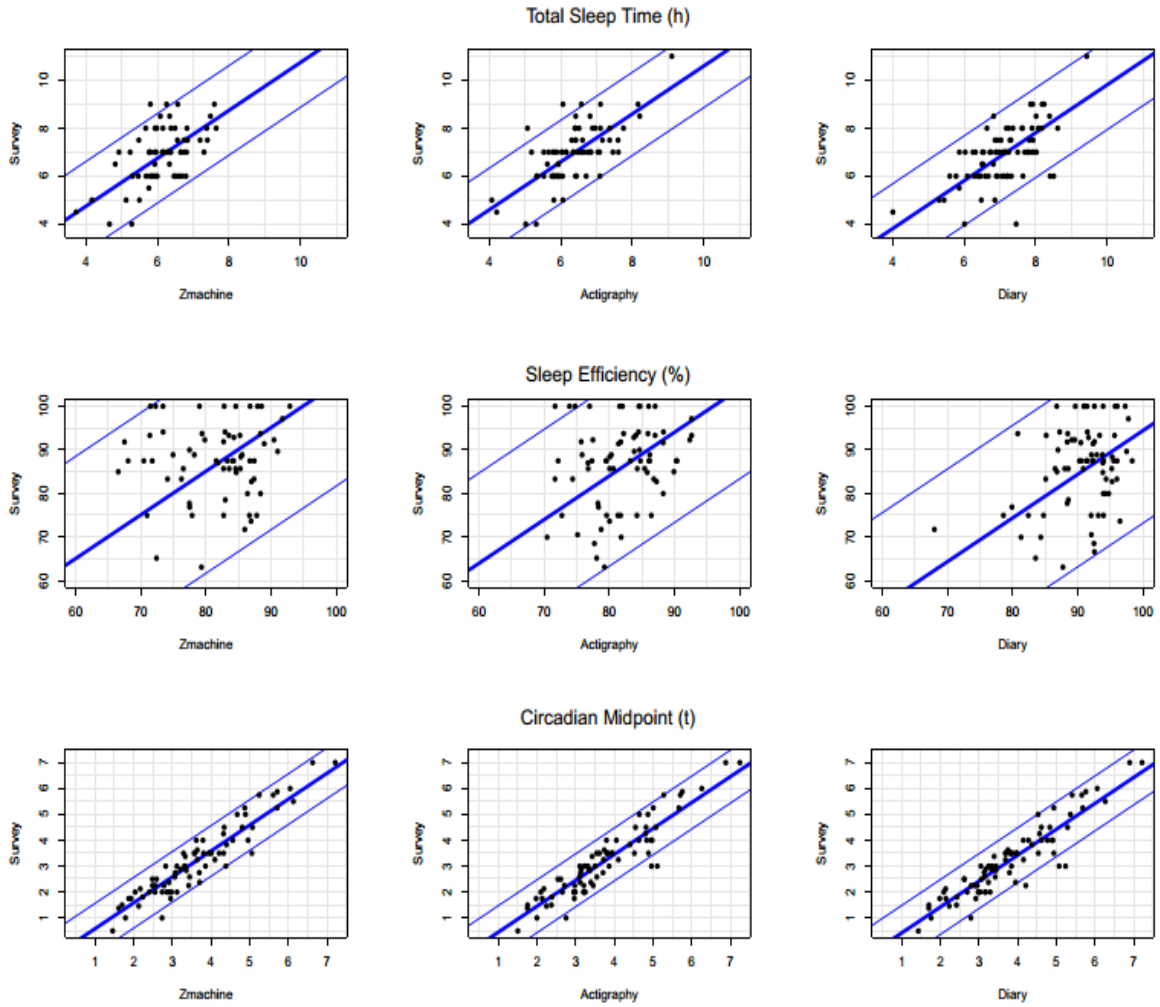

Figure 3. Prediction plots for averaged-across-week Zmachine- (EEG), actigraphy-, sleep diary- and survey-assessed total sleep time, sleep efficiency, and circadian midpoint.

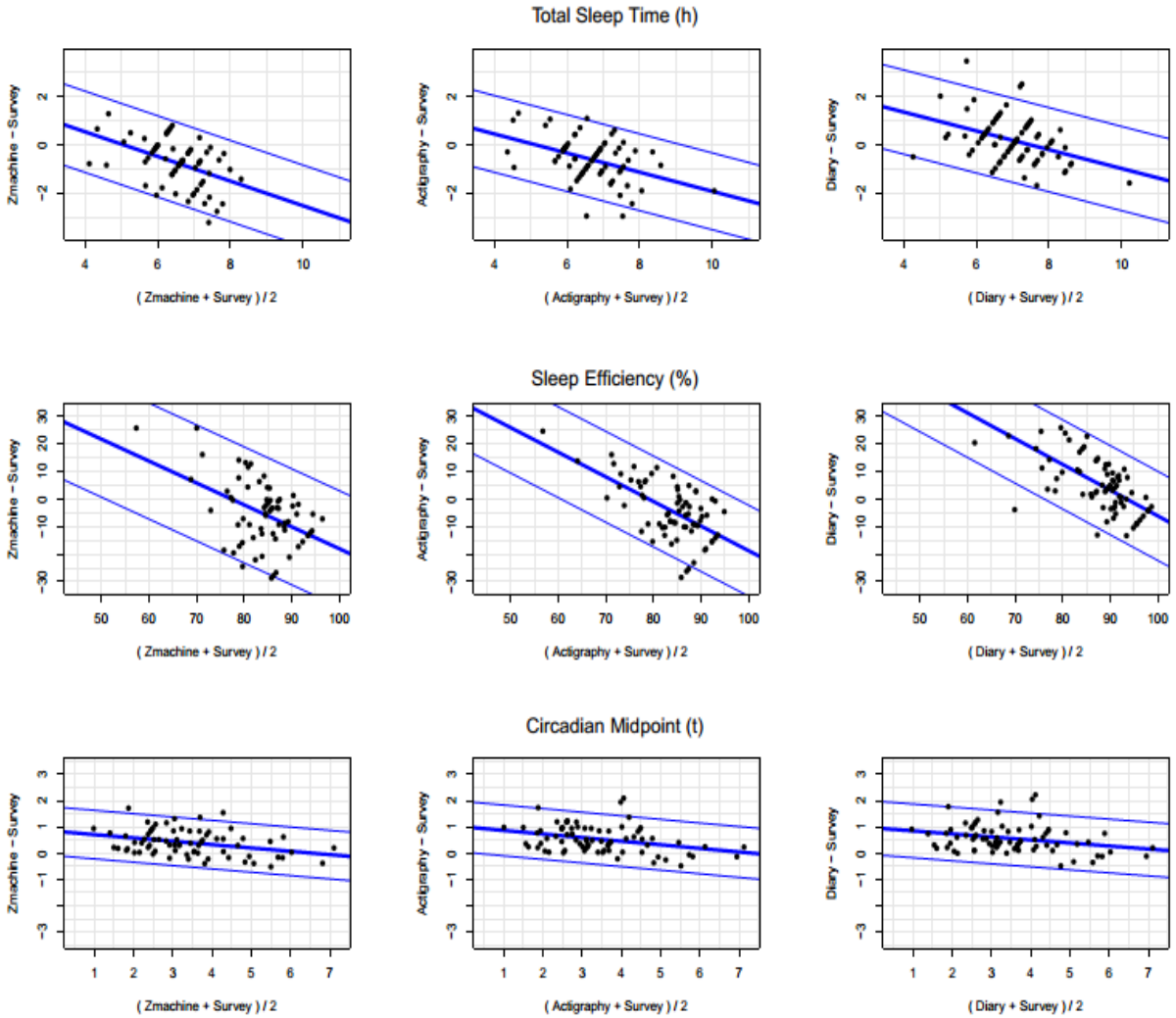

Figure 4. Bland-Altman plots for averaged-across-week Zmachine- (EEG), actigraphy-, sleep diary- and survey-assessed total sleep time, sleep efficiency, and circadian midpoint with slopes allowed to vary linearly and blue lines indicating mean bias and 95% prediction intervals.

For sleep efficiency, the prediction interval for the comparison between Zmachine and survey was >25%. Sleep efficiency demonstrated a similar pattern to total sleep time, such that for individuals with lower sleep efficiency, survey was predicted to underestimate their sleep efficiency compared to Zmachine. For individuals with higher sleep efficiency, survey was predicted to overestimate total sleep time compared to Zmachine. For an individual reporting 70% sleep efficiency on survey the predicted Zmachine value would be 74% and for an

individual reporting 95% sleep efficiency on survey the predicted Zmachine value would be 85%. Notably, most individuals in this study had sleep efficiency >75%, so information about lower sleep efficiencies is somewhat limited.

For circadian midpoint, the prediction interval for the comparison between Zmachine and survey was >1 hour. Circadian midpoint displayed a slight slope such that individuals sleeping earlier during the nighttime period were predicted to overestimate sleep timing on survey, and individuals sleeping later during the nighttime were predicted to slightly overestimate sleep timing on survey. For example, for an individual reporting a 2:00am circadian midpoint on survey the predicted Zmachine value would be 2:33am and for an individual reporting a 6:00am circadian midpoint on survey the predicted Zmachine value would be 6:04am.

Table 6

Bias, Precision, and 95% Limits of Agreement for 7 Days of Total Sleep Time, Sleep Efficiency, and Circadian Midpoint as Compared between Zmachine, Actigraphy, and Diary Compared to Survey

		α	σ	LoA lower	LoA upper
Total Sleep Time (h)	Zmachine v. Actigraphy	-0.13	0.84	-1.81	1.54
	Zmachine v. Diary	-0.67	1.03	-2.72	1.39
	Actigraphy v. Diary	-0.52	0.92	-2.36	1.32
Sleep Efficiency (%)	Zmachine v. Actigraphy	-0.55	9.97	-20.49	19.39
	Zmachine v. Diary	-7.07	11.68	-30.43	16.29
	Actigraphy v. Diary	-6.32	10.58	-27.47	14.83
Circadian Midpoint (t)	Zmachine v. Actigraphy	-0.08	0.27	-0.62	0.47
	Zmachine v. Diary	-0.10	0.35	-0.81	0.61
	Actigraphy v. Diary	-0.02	0.30	-0.61	0.57

Note. LoA = limit of agreement.

Hypothesis 2

Final models of linear regression results for hypothesis 2 are presented in Table 8. As hypothesized, insomnia severity was not a significant predictor of differences between Zmachine- and actigraphy-assessed total sleep time. The only significant predictors were gender and relationship status. Women had higher total sleep time on actigraphy than Zmachine (M_{diff} = 0.31) and men had lower total sleep time on actigraphy than Zmachine (M_{diff} = 0.07). Participants in a relationship had higher total sleep time on actigraphy than Zmachine (M_{diff} = -0.23) and single participants had lower total sleep time on actigraphy than Zmachine (M_{diff} = 0.05). Contrary to hypothesis, insomnia severity nor any other tested variables were significant predictors of differences between Zmachine- and diary-assessed total sleep time.

Table 7

Slopes, Intercepts, Standard Deviation of Prediction, and 95% Prediction Intervals for Averaged-Over-Week Total Sleep Time, Sleep Efficiency, and Circadian Midpoint as Measured by Zmachine, Actigraphy, and Diary Compared to Survey

		α	β	σ	95% p.i.
Total Sleep Time (h)	Zmachine v. Survey	2.04	0.60	0.67	±1.31
	Actigraphy v. Survey	1.68	0.67	0.66	±1.30
	Diary v. Survey	2.42	0.68	0.73	±1.42
Sleep Efficiency (%)	Zmachine v. Survey	43.92	0.43	7.52	±14.74
	Actigraphy v. Survey	48.69	0.38	5.69	±11.15
	Diary v. Survey	59.41	0.36	5.50	±10.79
Circadian Midpoint (t)	Zmachine v. Survey	0.79	0.88	0.44	±0.85
	Actigraphy v. Survey	0.95	0.87	0.45	±0.89
	Diary v. Survey	0.92	0.89	0.49	±0.95

Note. p.i. = prediction interval.

As hypothesized, insomnia severity significantly predicted differences between Zmachine- and survey-assessed total sleep time. Individuals with greater insomnia severity

demonstrated greater positive differences between survey- and Zmachine-assessed total sleep time, indicating lower total sleep time on survey than on Zmachine. Relationship status was also a significant predictor of differences between survey and Zmachine. Both participants in a relationship and single participants had higher total sleep time on actigraphy than Zmachine but the difference was greater for those in a relationship ($M_{diff} =$ 0.85) versus those single ($M_{diff} =$ 0.38). When the same analysis was conducted using absolute value of the differences between Zmachine and survey as the outcome variable, insomnia severity was again a significant predictor of differences suggesting greater insomnia severity is related to greater bias between survey and Zmachine estimations overall.

Table 8

Regression Results for Insomnia Severity, Demographics, and Psychosocial Variables Predicting Differences between Predictions by Zmachine Compared to Actigraphy, Diary and Survey for Total Sleep Time

Variable	Zmachine v. Actigraphy (n = 67)				Zmachine v. Diary (n = 69)				Zmachine v. Survey (n = 69)			
	B	*SE B*	β	r_s	*B*	*SE B*	β	r_s	*B*	*SE B*	β	r_s
ISI	0.01	0.01	.11	-.06	0.02	0.02	.20	.42	0.08	0.02	.47**	.63
PANAS NA	-0.02	0.02	-.26	-.17	-0.01	0.02	-.14	-.31	-0.02	0.03	-.13	.02
STAI	0.00	0.01	-.06	-.12	-0.01	0.01	-.11	-.27	-0.03	0.02	-.28	-.09
age	-0.01	0.01	-.20	-.43	0.00	0.03	.05	.34	0.01	0.01	.07	.24
gender	-0.28	0.13	-.33**	-.65	-0.15	0.15	-.12	-.39	-0.33	0.21	-.17	-.20
relationship status	0.45	0.18	.34*	.44	0.20	0.21	.14	.33	0.73	0.29	.33*	.49
education level	0.02	0.03	.08	.16	0.00	0.03	.00	.02	0.03	0.05	.07	.12
R^2	.25				.06				.27			
F (df)	2.83*				0.59				3.27**			

Note. r_s = structure coefficients (r/R); in a relationship = 0; male = 0. * $p < .05$; ** $p < .0$.

CHAPTER 4

DISCUSSION

The primary objective of the current study was to examine the accuracy of total sleep time, sleep efficiency, and circadian midpoint as assessed by actigraphy, sleep diary, and survey compared to a gold standard objective measure (Zmachine) in a community sample (N = 80) across 7 nights in the naturalistic sleep environment. Actigraphy generally demonstrated the best agreement with Zmachine across sleep variables, followed by sleep diary and then survey. Circadian midpoint was the most accurately assessed sleep parameter across measures, followed by total sleep time and then sleep efficiency.

Total Sleep Time

Actigraphy-assessed total sleep time did not differ substantially, on average, from Zmachine (+8 minutes). Previous findings are mixed, but the current results fall within the range of similar research. Current results are most similar to findings by McCall and McCall (2012) who compared PSG and actigraphy across one night in-lab in 54 patients with insomnia and depression and found actigraphy-assessed total sleep time was, on average, 13 minutes higher than PSG. Jean-Louis, Kripke, Mason, Elliott, and Youngstedt (2001) compared different actigraphy scoring modalities across 5 nights in lab in a small sample (N = 5) of healthy young adults and found for four out of five scoring modalities, actigraphy-assessed total sleep time was higher than PSG by 1-12 minutes. Similarly, Blackwell et al. (2008) compared 3 actigraphy scoring modalities across one night in lab in N = 68 older women and found for two of three modalities, actigraphy-assessed total sleep time was higher than PSG by 17-33 minutes and limits of agreement were similar in magnitude to the current study. Conversely, Matthews et al. (2018) found actigraphy-assessed total sleep time was lower than PSG by 7 minutes in a PSG-

actigraphy comparison of total sleep time across two nights of in-home PSG in a community sample (N = 215). Similarly, Signal, Gale, and Gander (2005) found actigraphy values were lower than PSG by 6 minutes, on average, when using a low wake threshold setting for an in-flight night, and 36 minutes lower, on average, during a layover night in a study comparing PSG and actigraphy in N = 21 flight crew members. In the current study, limits of agreement were broad due to poor precision (i.e., a large standard deviation of differences around the mean), with 95% of differences between the two measures falling within a range of >3 hours. McCall and McCall (2012) reported similarly broach 95% limits of agreement, with 95% of differences expected to fall between -1.5 to +2 hours, a total of a 3.5 hour span. Taken together, these results suggest actigraphy concordance with PSG/EEG appears to be dependent on numerous factors, likely including actigraphy scoring algorithm and settings, concurrent use of diary to set actigraphy scoring intervals, and setting of the study (e.g., in-lab vs. at home).

Given the lack of consistency in reporting and high degree of variability in actigraphy algorithm scoring settings, direct comparison is difficult. Regardless, in the current study actigraphy was, on average, an accurate measure of total sleep time compared to Zmachine. However, broad limits of agreement suggest a high degree of variability in the differences between these measures in individual cases and thus for a given individual the accuracy of actigraphy may be highly variable (i.e., over- or under-estimated by up to 2 hours). The average accuracy of actigraphy is likely due, in part, to a balance between over- and under-estimation which cancel out across an average. Although it is outside the scope of this study to suggest why the variability in accuracy was so high, one mechanism may be the variability in sleep complaints across the sample. For example, individuals with insomnia may be more likely to lie still while awake in bed, which would be incorrectly designated as sleep by actigraphy and correctly designated as wake by Zmachine (Lichstein et al., 2006). Conversely, individuals with

other sleep disorders (e.g., upper airway resistance syndrome and periodic limb movement disorder) may demonstrate the opposite effect (Kushida et al., 2001). It is likely that "good sleepers" and good nights of sleep produce greater accuracy whereas individuals with sleep or medical disorders may be more prone to discrepancies between actigraphy and Zmachine estimates of total sleep time.

In the current study, sleep diary-assessed total sleep time was 40 minutes higher than Zmachine on average. This was similar to findings by K. A. Kaplan, Talbot, Gruber, & Harvey (2012), who found sleep diary overestimated total sleep time compared to PSG by 50 minutes (night 1) and 28 minutes (night 2) in a group of N =27 healthy middle-aged adults. Matthews et al. (2018) found diary overestimated total sleep time compared to PSG by only 15 minutes on average across 2 nights in N = 223 middle-aged and older adults. In contrast, McCall and McCall (2012) found sleep diary estimates were 55 minutes lower, on average, than PSG in a sample of N = 54 young, middle-aged, and older adults. This extreme difference from the current findings is potentially attributable to a difference in participant characteristics, as McCall and McCall (2012) had a sample of people with depression and insomnia which is a group known to underestimate sleep on subjective measures. In the current study, limits of agreement were broad (i.e., 95% of differences fall within a range of >4 hours). McCall and McCall (2012) found even broader 95% limits of agreement (6.8 hours). Given the current and previous studies demonstrating sleep diary substantially overestimates total sleep time compared to PSG and Zmachine, and produced broad limits of agreement, it is reasonable to suggest that sleep diary is not an interchangeable method for measuring total sleep time. Instead, subjective total sleep time may reflect an overlapping yet distinct construct than objectively-assessed total sleep time and should be treated as such. Emerging research suggests subjective aspects of sleep like "depth" or "restfulness" are not closely related to objective sleep assessments (K. A. Kaplan et al., 2017). In

other words, self-assessment of sleep may not reflect its underlying biological functions (K. A. Kaplan et al., 2017). However, both objective and subjective assessments of sleep are important variables to consider depending on the domain of interest.

Survey estimates of total sleep time demonstrated non-constant bias such that lower values of total sleep time had positive bias (survey estimates were lower than Zmachine), and the reverse was true for higher values of total sleep time (survey estimates were higher than Zmachine). Additionally, precision was poor for Zmachine versus survey comparisons of total sleep time (i.e., 95% of predictions fall within a range of >2.5 hours). To our knowledge, no studies have examined non-constant bias between survey- and PSG-assessed sleep parameters. However, previous studies demonstrated a slight trend toward overestimation by survey compared to PSG, which was consistent with the current results. The current findings shed light on the potential for bias in survey estimates depending on level of the estimated parameter. Given this bias, daily sleep diary estimates are the preferred subjective measure given that they do not result in non-constant bias, which is difficult to adjust for either clinically or statistically.

Sleep Efficiency

Actigraphy-assessed sleep efficiency did not differ substantially (<1%), on average, from Zmachine. These findings are similar to Jean-Louis et al. (2001), McCall and McCall (2012) and Lichstein et al. (2006) who found minimal differences between measures in estimating sleep efficiency. Previous studies have also shown actigraphy to overestimate (Blackwell et al., 2008; 4-7%) or underestimate (Blackwell et al., 2008; 6%; Rupp & Balkin, 2011; 4-6%; Signal et al., 2005; 3-9%), depending on actigraphy modality, device, and sleep situation. In the current study, precision was poor, resulting in broad 95% limits of agreement (almost 40%). This suggests although actigraphy estimates of sleep efficiency are similar to Zmachine on average across a

sample, for an individual the accuracy of actigraphy is highly variable. Further, this variability was greater for lower estimates of sleep efficiency. Therefore, to better approximate Zmachine estimates of sleep efficiency, it is recommended to use actigraphy in large samples across several nights. Additionally, at the individual level, actigraphy estimates of sleep efficiency are likely to be more accurate for individuals or nights with greater sleep efficiency, and less accurate for lower sleep efficiency (e.g., people with insomnia).

Diary overestimated sleep efficiency compared to Zmachine by 7% on average. Two existing studies demonstrated minimal differences between sleep diary and PSG estimates of sleep efficiency (Lichstein et al., 2006; Kaplan et al., 2012). These differences may relate to variations in the form of sleep diary used across studies or the format of administration, as previous studies used paper-based sleep diaries. Similar to actigraphy, variability was greater for lower values of sleep efficiency. Additionally, precision was poor resulting in broad 95% limits of agreement (>45%). Diary performed only slightly worse than actigraphy for both mean and precision, underlining that as a subjective measure

Survey estimates of sleep efficiency demonstrated non-constant bias such that lower values of sleep efficiency had positive bias (survey estimates were lower than Zmachine), and the reverse was true for higher values of sleep efficiency (survey estimates were higher than Zmachine). Additionally, precision was poor for Zmachine versus survey comparisons of sleep efficiency (i.e., 95% of predictions fall within a range of almost 30%). To our knowledge, this is one of only two studies to compare sleep efficiency estimates between survey and PSG. Signal et al. (2005) found survey underestimated sleep efficiency by 22% on average compared to PSG, although their questionnaire substantially differed from the one used in the current study. On average sleep efficiency estimates were quite similar between survey and Zmachine, however the

non-constant bias and broad prediction interval suggest caution should be used in survey estimates of sleep efficiency.

Circadian Midpoint

Both actigraphy- and diary-assessed circadian midpoint was quite similar, on average, to Zmachine. Survey estimates of circadian midpoint demonstrated slight non-constant bias such that earlier circadian midpoints had positive bias (survey estimates were earlier than Zmachine) and the reverse was true for later circadian midpoints (survey estimates were later than Zmachine). Precision was acceptable for all methods. To our knowledge, this is the first study to compare accuracy across measures of circadian midpoint. High accuracy for actigraphy is likely due to the anchoring events for both devices (i.e., pressing the event marker and putting on/removing the Zmachine electrodes). The high accuracy of diary for circadian midpoint is likely attributable, in part, to the attention that individuals pay to bedtime and wake time and the proximity of these events to periods of sustained wakefulness. Because sleep is an inherently amnesiac state, memory for nighttime periods of wakefulness (e.g., SOL, WASO) are likely to be recalled less accurately. Further, setting alarms for morning wake-up provides a timestamp for bedtime and an anchor point for wake time, which can improve recall ability. The stability of circadian midpoint for both actigraphy and diary suggests that they are excellent proxies for Zmachine-assessed sleep timing, whereas survey estimates are slightly biased and should be used with caution.

Prediction of Measure Discrepancies

The second objective of the current study was to assess whether insomnia severity and other participant characteristics predicted differences between measures of total sleep time.

Insomnia severity was only a significant predictor of differences between Zmachine- and survey-assessed total sleep time but did not predict differences between Zmachine and other measures. Greater insomnia severity was related to greater underestimation of total sleep time on survey compared to Zmachine, suggesting people with insomnia subjectively underestimate their total sleep time in a week-long retrospective. These findings extend the work of Manconi et al. (2010), who found individuals with insomnia substantially underestimated total sleep time on retrospective survey compared to PSG across one night in lab. In contrast, insomnia severity was not a significant predictor of differences between actigraphy- or diary-assessed total sleep time compared to Zmachine. This suggests that when measured with another objective measure, or a daily-observed subjective measure, insomnia severity does not significantly impact total sleep time estimates. It is only on the week-long retrospective that a bias for underestimation on survey is observed. This suggests when measuring individuals with insomnia using subjective measures, daily diary may reduce underestimation bias and therefore be the preferred method over retrospective survey.

Strengths and Limitations

To our knowledge, this study was the first in the field of sleep measurement to adjust for repeated measures using a mixed model approach. Few studies examined Bland-Altman plots or limits of agreement to assess accuracy of measures, and among those that did, none adjusted for repeated measures. Adjusting for repeated measures improves confidence in the accuracy of the limits of agreement calculated (without adjustments for repeated measures, limits of agreement in the current study were calculated to be much wider). Further, this study extends the scarce literature comparing diary and survey measures to PSG. This is a crucial area of study, as many studies present subjective measures of sleep parameters as substitutes for objective measures

when in reality they appear to reflect different constructs. Finally, this was the first study to examine accuracy of sleep measures for circadian midpoint. The encouraging findings suggested all examined measures accurately reflected circadian midpoint and can be used to estimate sleep timing.

The current study was not without limitations. First, although this sample of participants was drawn from the community, the sample demographics do not reflect the larger community and substantially limit generalizability of results. In particular, this was a highly educated and largely non-Hispanic White convenience sample, so results may not apply to individuals with lower education or with different racial/ethnic identities. Second, use of a single-channel EEG device as a gold standard sacrifices some accuracy in assessing the true value of the measured construct in trade for greater ecological validity. The exact impact of this trade-off cannot be assessed within the current study but use of this device may limit comparisons with full PSG. Obviously, the results here cannot be generalized to studies conducted in a laboratory environment. Finally, retrospective surveys were collected after a week of careful attention to sleep via diary which likely increased accuracy of survey measures compared to typical use.

Implications and Future Directions

This was one of few studies to examine the validity of actigraphy, sleep diary, and Zmachine measures across a week in an ecologically valid setting, and the results highlight the importance of paying careful attention to measurement method when assessing total sleep time, sleep efficiency, and circadian midpoint because accuracy varies across method and outcome. These results also demonstrate the importance of using statistical techniques that describe accuracy (e.g., Bland Altman plots) in addition to relationship (e.g., correlation). Unsurprisingly, the current study demonstrated actigraphy is the best Zmachine proxy for all outcomes examined

and appears valid when viewed in aggregate across participants. However, poor precision for total sleep time and efficiency suggests individual days or participants may not demonstrate acceptable accuracy. Sleep diary was a worse Zmachine proxy for all outcomes and substantially overestimated total sleep time and efficiency. This is unsurprising given that sleep diary is a subjective measure and cannot be expected to agree perfectly with objective measures. Survey, the most commonly used sleep assessment in health research, performed poorly and demonstrated problematic bias in differences across levels of total sleep time, sleep efficiency, and circadian midpoint. Results suggest insomnia may be one characteristic that explains the poor performance of survey compared to Zmachine—insomnia severity impacted recall across the past week. Optimally, use of an objective measure in conjunction with a subjective measure (preferably sleep diary) will produce the greatest confidence in results and capture a fuller picture of sleep.

Future studies should continue to examine potential predictors of discrepancies between sleep measures, particularly in the comparison of subjective and objective measures. For example, with more information steps can be taken to adjust for known causes of discrepancies (e.g., for a person with insomnia, an upward adjustment for subjective total sleep time). Future studies should also continue to examine diary and survey methods in comparison to other measures, as relatively little is known about the accuracy of these tools. In particular, the current work could be extended by collecting objective sleep data in conjunction with a retrospective survey without a sleep diary in order to more closely mimic typical survey use.

Future studies should continue to explore measurement validity in the naturalistic sleep environment, as most current validation research focuses on a laboratory environment, whereas a large amount of sleep research occurs in the home environment. In particular, individuals from low-socioeconomic status backgrounds may be more likely to have discrepancies between their

typical sleep environment and the laboratory environment. Thus, it is important to continue this work in individuals with diverse characteristics, particularly on known sources of sleep disparities such as socioeconomic status, race/ethnicity, and sex/gender.

REFERENCES

Alvarez, G. G., & Ayas, N. T. (2004). The impact of daily sleep duration on health: A review of the literature. *Progress in Cardiovascular Nursing, 19*(2), 56-59.

Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., & Pollak, C. (2003). The role of actigraphy in the study of sleep and circadian rhythms. *Sleep, 26*(3), 342-392.

Bastien, C. H., Vallieres, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2*(4), 297-307.

Bei, B., Wiley, J. F., Trinder, J., & Manber, R. (2016). Beyond the mean: A systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. *Sleep Medicine Reviews, 28*, 104-120.

Blackwell, T., Redline, S., Ancoli-Israel, S., Schneider, J. L., Surovec, S., Johnson, N. L., . . . Study of Osteoporotic Fractures Research Group. (2008). Comparison of sleep parameters from actigraphy and polysomnography in older women: The SOF Study. *Sleep, 31*(2), 283-291. doi: 10.1093/sleep/31.2.283

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research, 8*(2), 135-160.

Bland, J. M., & Altman, D. G. (2010). Statistical methods for assessing agreement between two methods of clinical measurement. *International Journal of Nursing Studies, 47*(8), 931-936.

Buysse, D. J. (2014). Sleep health: Can we define it? Does it matter? *Sleep, 37*(1), 9-17.

Buysse, D. J., Ancoli-lsrael, S., Edinger, J. D., Lichstein, K. L., & Morin, C. M. (2006). Recommendations for a standard research assessment of insomnia. *Sleep, 29*(9), 1155-73.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Research, 28*(2), 193-213.

Cappuccio, F. P., D'Elia, L., Strazzullo, P., & Miller, M. A. (2010). Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep, 33*(5), 585.

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., & Morin, C. M. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep, 35*(2), 287-302. doi: 10.5665/sleep.1642

Carskadon, M. A., & Dement, W. C. (2011). Normal human sleep: An overview. In M. H. Kryger, T. Roth & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (5 ed., pp. 13-23). St. Louis, MO: Elsevier.

Carstensen, B. (2010). Comparing methods of measurement: Extending the LoA by regression. *Statistics in Medicine, 29*(3), 401-410. doi: 10.1002/sim.3769

Carstensen, B., Gurrin, L., Claus Ekstrom, C., & Figurski, M. (2016). MethComp: functions for analysis of agreement in method comparison studies. (Version 1.22. 2). Retrieved from http://CRAN.R-project.org/package=MethComp

Carstensen, B., Simpson, J., & Gurrin, L. (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *International Journal of Biostatistics, 4*(1). doi: 10.2202/1557-4679.1107

Cellini, N., Buman, M. P., McDevitt, E. A., Ricker, A. A., & Mednick, S. C. (2013). Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiology International, 30*(5), 691-698. doi: 10.3109/07420528.2013.782312

Chervin, R. D., Murman, D. L., Malow, B. A., & Totten, V. (1999). Cost-utility of three approaches to the diagnosis of sleep apnea: polysomnography, home testing, and empirical therapy. *Annals of Internal Medicine, 130*(6), 496-505.

Costa, G. (1996). The impact of shift and night work on health. *Applied Ergonomics, 27*(1), 9-16.

Czeisler, C. A. (2015). Duration, timing and quality of sleep are each vital for health, performance and safety. *Sleep Health, 1*(1), 5-8.

de Souza, L., Benedito-Silva, A. A., Pires, M. N., Poyares, D., Tufik, S., & Calil, H. M. (2003). Further validation of actigraphy for sleep studies. *Sleep, 26*(1), 81-85.

Grandner, M. A., Hale, L., Moore, M., & Patel, N. P. (2010). Mortality associated with short sleep duration: the evidence, the possible mechanisms, and the future. *Sleep Medicine Reviews, 14*(3), 191-203. doi: 10.1016/j.smrv.2009.07.006

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377-381. doi: 10.1016/j.jbi.2008.08.010

Hublin, C., Kaprio, J., Partinen, M., & Koskenvuo, M. (2001). Insufficient sleep--a population-based study in adults. *Sleep, 24*(4), 392-400.

Jean-Louis, G., Kripke, D. F., Mason, W. J., Elliott, J. A., & Youngstedt, S. D. (2001). Sleep estimation from wrist movement quantified by different actigraphic modalities. *Journal of Neuroscience Methods, 105*(2), 185-191. doi: 10.1016/S0165-0270(00)00364-2

Kaplan, K. A., Hirshman, J., Hernandez, B., Stefanick, M. L., Hoffman, A. R., Redline, S., . . . Zeitzer, J. M. (2017). When a gold standard isn't so golden: Lack of prediction of subjective sleep quality from sleep polysomnography. *Biological Psychology, 123*, 37-46.

Kaplan, K. A., Talbot, L. S., Gruber, J., & Harvey, A. G. (2012). Evaluating sleep in bipolar disorder: comparison between actigraphy, polysomnography, and sleep diary. *Bipolar Disorders, 14*(8), 870-879. doi: 10.1111/bdi.12021

Kaplan, R. F., Wang, Y., Loparo, K. A., Kelly, M. R., & Bootzin, R. R. (2014). Performance evaluation of an automated single-channel sleep-wake detection algorithm. *Nature and Science of Sleep, 6*, 113-122.

Kay, D. B., Buysse, D. J., Germain, A., Hall, M., & Monk, T. H. (2015). Subjective–objective sleep discrepancy among older adults: associations with insomnia diagnosis and insomnia treatment. *Journal of Sleep Research, 24*(1), 32-39. doi: 10.1111/jsr.12220

Kimoff, R. J. (1996). Sleep fragmentation in obstructive sleep apnea. *Sleep, 19*(9 Suppl), S61-66.

Kurina, L. M., McClintock, M. K., Chen, J.-H., Waite, L. J., Thisted, R. A., & Lauderdale, D. S. (2013). Sleep duration and all-cause mortality: a critical review of measurement and associations. *Annals of Epidemiology, 23*(6), 361-370.

Kushida, C. A., Chang, A., Gadkary, C., Guilleminault, C., Carrillo, O., & Dement, W. C. (2001). Comparison of actigraphic, polysomnographic, and subjective assessment of sleep parameters in sleep-disordered patients. *Sleep Medicine, 2*(5), 389-396.

Kushida, C. A., Littner, M. R., Morgenthaler, T., Alessi, C. A., Bailey, D., Coleman Jr, J., . . . Kramer, M. (2005). Practice parameters for the indications for polysomnography and related procedures: an update for 2005. *Sleep, 28*(4), 499-521.

Lauderdale, D. S. (2014). Survey questions about sleep duration: does asking separately about weekdays and weekends matter? *Behavioral Sleep Medicine, 12*(2), 158-168. doi: 10.1080/15402002.2013.778201

Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K., & Rathouz, P. J. (2008). Self-reported and measured sleep duration: how similar are they? *Epidemiology, 19*(6), 838-845.

Lewy, A. J., & Sack, R. L. (1989). The Dim Light Melatonin Onset as a Marker for Orcadian Phase Position. *Chronobiology International, 6*(1), 93-102.

Lichstein, K. L., Stone, K. C., Donaldson, J., Nau, S. D., Soeffing, J. P., Murray, D., . . . Aguillard, R. N. (2006). Actigraphy validation with insomnia. *Sleep, 29*(2), 232.

Malhotra, A., Younes, M., Kuna, S. T., Benca, R., Kushida, C. A., Walsh, J., . . . Pien, G. W. (2013). Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. *Sleep, 36*(4), 573-582.

Manconi, M., Ferri, R., Sagrada, C., Punjabi, N. M., Tettamanzi, E., Zucconi, M., . . . Ferni-Strambi, L. (2010). Measuring the error in sleep estimation in normal subjects and in patients with insomnia. *Journal of Sleep Research, 19*(3), 478-486. doi: 10.1111/j.1365-2869.2009.00801.x

Martin, J. L., & Hakim, A. D. (2011). Wrist actigraphy. *CHEST Journal, 139*(6), 1514-1527.

Matthews, K. A., Patel, S. R., Pantesco, E. J., Buysse, D. J., Kamarck, T. W., Lee, L., & Hall, M. H. (2018). Similarities and differences in estimates of sleep duration by polysomnography, actigraphy, diary, and self-reported habitual sleep in a community sample. *Sleep Health, 4*(1), 96-103. doi: 10.1016/j.sleh.2017.10.011

McCall, C., & McCall, W. V. (2012). Comparison of actigraphy with polysomnography and sleep logs in depressed insomniacs. *Journal of Sleep Research, 21*(1), 122-127. doi: 10.1111/j.1365-2869.2011.00917.x

Patel, S., Malhotra, A., & Gottlieb, D. (2006). Correlates of long sleep duration. *Sleep, 29*(7), 881-889.

Reid, K. J., & Zee, P. C. (2009). *Circadian Rhythm Disorders.* Paper presented at the Seminars in neurology.

Rogers, A. E., Caruso, C. C., & Aldrich, M. S. (1993). Reliability of sleep diaries for assessment of sleep/wake patterns. *Nursing Research, 42*(6), 368-371.

Rupp, T. L., & Balkin, T. J. (2011). Comparison of Motionlogger Watch and Actiwatch actigraphs to polysomnography for sleep/wake estimation in healthy young adults. *Behavior Research Methods, 43*(4), 1152-1160. doi: 10.3758/s13428-011-0098-4

Sadeh, A., & Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews, 6*(2), 113-124.

Sánchez-Ortuño, M. M., Edinger, J. D., Means, M. K., & Almirall, D. (2010). Home is where sleep is: an ecological approach to test the validity of actigraphy for the assessment of insomnia. *Journal of Clinical Sleep Medicine, 6*(1), 11-22.

Schwartz, J. R., & Roth, T. (2006). Shift work sleep disorder. *Drugs, 66*(18), 2357-2370.

Schwarz, N. (2007). Retrospective and concurrent self-reports: The rationale for real-time data capture. *The science of real-time data capture: Self-reports in health research*, 11-26.

Signal, T. L., Gale, J., & Gander, P. H. (2005). Sleep measurement in flight crew: comparing actigraphic and subjective estimates to polysomnography. *Aviation, Space, and Environmental Medicine, 76*(11), 1058-1063.

Smith, C. S., Reilly, C., & Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *Journal of Applied Psychology, 74*(5), 728.

Spielman, A. J., Saskin, P., & Thorpy, M. J. (1987). Treatment of chronic insomnia by restriction of time in bed. *Sleep, 10*(1), 45-56.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials, 24*(2), 182-199.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Boston, MA: Pearson.

Taylor, D. J., Wilkerson, A. K., Pruiksma, K. E., Williams, J. M., Ruggero, C. J., Hale, W., . . . Litz, B. T. (2018). Reliability of the Structured Clinical Interview for DSM-5 Sleep Disorders Module. *Journal of Clinical Sleep Medicine, 14*(03), 459-464. doi: 10.5664/jcsm.7000

Vallieres, A., Ivers, H., Bastien, C. H., Beaulieu-Bonneau, S., & Morin, C. M. (2005). Variability and predictability in sleep patterns of chronic insomniacs. *Journal of Sleep Research, 14*(4), 447-453.

Van Dongen, H., Vitellaro, K. M., & Dinges, D. F. (2005). Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. *Sleep, 28*(4), 479-496.

Van Someren, E. J. (2007). Improving actigraphic sleep estimates in insomnia and dementia: how many nights? *Journal of Sleep Research, 16*(3), 269-275.

Vgontzas, A. N., Fernandez-Mendoza, J., Liao, D., & Bixler, E. O. (2013). Insomnia with objective short sleep duration: the most biologically severe phenotype of the disorder. *Sleep Medicine Reviews, 17*(4), 241-254.

Wetter, T. C., & Pollmächer, T. (1997). Restless legs and periodic leg movements in sleep syndromes. *Journal of Neurology, 244*(1), S37-S45.

Wohlgemuth, W. K., Edinger, J. D., Fins, A. I., & Sullivan, R. J. (1999). How many nights are enough? The short-term stability of sleep parameters in elderly insomniacs and normal sleepers. *Psychophysiology, 36*(2), 233-244.