TITLE: Clinical Validation of New Commercial Sleep Monitoring Devices

IRB Protocol #: 1509016542 **ClinicalTrials.gov ID: NCT02779543**

Version Date: Dec 14, 2015

Principal Investigator:

Name: Ana Krieger
Address: WCMC Center for Sleep Medicine, 425 East 61st St 5th Floor, New York, NY 10065
Telephone: 646-962-7378
Fax: 646-962-0455
Email: ack2003@med.cornell.edu

Administrative Contact / Research Assistant:

Name: Erik Zavrel
Address: WCMC Center for Sleep Medicine, 425 East 61st St 5th Floor, New York, NY 10065
Telephone: 646-749-5237
Email: erz2004@med.cornell.edu

SUMMARY

This study seeks to evaluate three new commercial sleep monitoring devices: Fitbit, Jawbone UP, and Microsoft Band by comparison against concurrent polysomnography, considered the gold standard assessment for sleep physiology.

TABLE OF CONTENTS

1. OBJECTIVES

1.1. Primary Objectives

This study seeks to evaluate three new commercial sleep monitoring devices: Fitbit, Jawbone UP, and Microsoft Band by comparison against concurrent polysomnography (PSG) in a clinical population. The primary objective is to compare total sleep time (TST) between each of the sleep monitoring devices and PSG.

1.2. Secondary Objectives

a. To assess agreement between sleep monitoring devices and PSG by characterizing epoch-by-epoch data into pooled contingency tables for kappa analysis.

b. Evaluate agreement between sleep monitoring devices and PSG for other sleep parameters: Sleep Onset Latency (SOL), Latency to Persistent Sleep (LPS), Wake After Sleep Onset (WASO), Sleep Efficiency (SE), and time spent in particular sleep stages (Light, Deep, REM).

c. We will assess inter-device agreement in a sub-population of participants who are willing to simultaneously wear two different sleep monitoring devices, e.g., a Fitbit and a Jawbone UP or a Fitbit and a Microsoft Band.

2. BACKGROUND AND RATIONALE

2.1 Disease

Participants will be recruited from those patients attending a regularly scheduled overnight PSG sleep study at the WCMC Center for Sleep Medicine. Patients are referred for an overnight study when a sleep disorder is suspected. Sleep disorders encompass numerous problems including sleep apnea, insomnia, narcolepsy, periodic limb movement, restless leg syndrome, circadian rhythm issues, among others. PSG enables confirmation and severity assessment, necessary for determination of proper treatment.

Obstructive sleep apnea (OSA) is a condition characterized by temporary diminutions or cessations of breathing caused by repetitive collapse of the upper airway (UA) during sleep (Eastwood 2006). OSA is a common disorder associated with abnormalities in pharyngeal anatomy and physiology in which the muscles of the airway, which normally relax during sleep, fail to provide sufficient dilatory force to balance the contractive force from inspiratory activity (Fogel 2002). This force imbalance serves to either partially or completely collapse the UA, thereby preventing sufficient air from reaching the lungs. These pauses in breathing lead to blood oxygen desaturation and induce neurological arousal resulting in sleep disruption and fragmentation. The cycle of airway collapse and arousal can repeat hundreds of times per night (Eastwood 2011). According to the National Sleep Foundation, OSA affects 18-22 million Americans, 80% of whom are undiagnosed. OSA is more prevalent among overweight and older individuals and those with reduced muscle tone, skeletal anomalies such as micrognathia or retrognathia, and airways crowded by redundant or enlarged soft tissue structures.

OSA is associated with significant physiological and psychological problems. OSA results in excessive daytime sleepiness, fatigue, memory impairment, and reduced reaction time, increasing the

risk for motor vehicle (Peppard 2000) and workplace (Yaggi 2005) accidents. In addition, OSA sufferers face increased cardiovascular risk including hypertension, heart disease, and stroke (Eikermann 2007). OSA has even recently been linked to increased cancer incidence (Campos-Rodriguez 2013) and mortality (Martínez-García 2014), presumably through hypoxia-induced angiogenesis.

The Harvard Medical School released a report in 2010 entitled, “The Price of Fatigue: The Surprising Economic Costs of Unmanaged Sleep Apnea,” in which it estimated the annual economic cost of moderate to severe OSA in the United States to be $65-165B (compared to $60B for drunk driving and $150B for not wearing seatbelts), including $10-40B in OSA-related traffic accidents and $5-20B in OSA-related workplace accidents (Harvard Medical School, 2010).

Existing treatments for OSA include lifestyle modifications (Busetto 2005) such as weight loss, position restriction, and avoidance of muscle relaxants such as alcohol and benzodiazepine drugs. Oral appliances including mandibular advancement devices and tongue retaining devices have been increasingly employed. OSA is also treated through surgery, including tonsillectomy and uvulopalatopharyngeoplasty (UPPP) to reduce tissue crowding of the UA lumen, genioglossal advancement, and maxillomandibular advancement. The most widespread and generally effective treatment for OSA, however, remains the use of various devices for maintaining positive airway pressure (PAP) such as Continuous Positive Airway Pressure (CPAP), BiLevel Positive Airway Pressure (BiPAP), and Autotitrating (AutoPAP) devices.

Insomnia is among the most common health complaints. The defining characteristics of insomnia are widely recognized as difficulty falling asleep, difficulty staying asleep, and sleep that is non-restorative (DSM-IV 2000). Estimates of the prevalence of insomnia vary widely depending on the criteria used and the population assessed but the majority of evidence suggests that about 10% of the adult population complains of a chronic insomnia problem (sleep difficulty at least 3 nights per week lasting longer than 1 month) (Ohayon 2002). With aging, increasingly disturbed sleep and less satisfaction with sleep quality are reported. More than 50% of elderly people have difficulty initiating and maintaining sleep (Kamel 2006).

This common problem has wide ranging cognitive and behavioral consequences including excessive daytime sleepiness, chronic fatigue, difficulty concentrating, memory impairment, diminished motivation, compromised social interaction, increased absenteeism, diminished job performance, and higher healthcare utilization. It has also been well established that insomnia is highly co-morbid with psychiatric disorders and is a risk factor for anxiety, depression, and suicide (Jansson-Froejmark 2008). Insomnia is also increasingly associated with broader health problems. In one large study, insomnia was associated with a high risk of hypertension (Vgontzas 2009). Another study has shown that insomnia is predictive of increased mortality (Capuccio 2010). In elderly with dementia, sleep disruption can lead to nursing home placement (Neubauer 1999).

2.2 Investigational Agent or Device

Fitbit
Jawbone UP
Microsoft Band

These are accelerometer-based wrist-worn actigraphy devices, which measure limb movement of the wearer. These devices are not FDA approved as they are not intended to diagnose or treat any disease or medical condition; however, two of them (Fitbit and Jawbone UP) have been used in previous validation studies.

2.3 Rationale and Background

The ability to investigate sleep is of scientific and clinical interest. Sleep patterns are studied with PSG, subjective self-reported diaries and questionnaires, and actigraphy. Each method of measuring sleep has advantages and limitations.

Historically, the gold standard of sleep measurement has been laboratory PSG, which utilizes a combination of electroencephalography (EEG), electrooculography, and electromyography (EMG) to determine sleep stages and sleep-related phenomena such as arousals. Laboratory PSG recordings require a physical space to conduct the sleep assessment and an on-site overnight staff to both apply and remove the physiological sensors and to ensure the integrity of acquired data. Data are often scored visually, in 30-second epochs, by registered PSG technicians. PSG boasts the advantage of excellence in terms of individualized sleep staging accuracy; however, the financial costs and time associated with conducting the data acquisition and subsequent scoring of the sleep records, as well as the burden to participants or patients, can outweigh this benefit. In those research studies and clinical screening applications in which in-home sleep monitoring over many days or weeks is required, the use of portable multichannel PSG is often financially and logistically impractical.

Subjective sleep assessments are a common method for obtaining information about the sleep of participants in research and clinical populations. Although sleep diaries and questionnaires offer a way to monitor sleep and habits easily and affordably over long periods of time, such tools suffer severe limitations. The very act of sleeping precludes conscious knowledge of time and events, and reliance on subjective self-reports of sleep, without objective data, can lead to an incomplete and potentially misleading picture. The data collected by such tools may not always be accurate accounts of sleep in either healthy or sleep-disordered populations. For instance, it is well known that insomniacs dramatically underestimate the amount of time they spend asleep throughout the night and overestimate sleep onset latency (SOL, the amount of time spent awake before falling asleep). In addition, sleep diaries are limited to sleep and wakefulness with no ability to discriminate between sleep stages. Subjective measures of sleep also require participants to manually enter data, making compliance challenging. The tools used for subjective sleep assessments also vary widely (Epworth Sleepiness Scale / ESS, Fatigue Severity Scale / FSS, Insomnia Severity Index / ISI, Functional Outcomes of Sleep Questionnaire / FOSQ, bed partner snoring logs), making comparisons between reports challenging.

The cost and inconvenience of PSG and the subjectivity and limitations of sleep diaries and questionnaires have motivated the development of devices capable of evaluating sleep in the home. Actigraphy systems are accelerometer-based devices that infer sleep and wake from the presence or absence of movement. Actigraphy is less expensive and less cumbersome than PSG, is relatively easy to use in ambulatory settings, and utilizes automated scoring algorithms that reduce the need for

manual interpretation of the gathered data. These advantages enable its use for multiple nights in longitudinal studies. Because actigraphy uses limb movement as a proxy for wakefulness, it traditionally has been limited to the detection of sleep and wakefulness (unable to perform sleep staging); however, some new consumer sleep monitoring devices purport to differentiate among sleep stages.

The development of low-cost, portable, easy-to-use sleep recording devices, with automated algorithms to distinguish between sleep stages, has important implications for sleep medicine and research. Clinical validation using a large diverse population of participants will provide substantial insight into the accuracy of these new commercial sleep monitoring devices. This study follows a validation study (Protocol Title: Validation of a New Sleep Measurement Device / IRB Protocol Number 1210013184) of another consumer sleep monitoring device we conducted in 2014. This study provided a wealth of data and we are currently preparing to publish our results.

A recent validation study (Montgomery-Downs 2012) found Fitbit to have high sensitivity but poor specificity (tending to misidentify wake as sleep and thus overestimate total sleep time). The participants in this study (N=24) were healthy (non-disordered) young adults. The study concluded with a recommendation of additional validation using a clinical population of varying ages. A separate validation study (Zambotti 2015) found Jawbone UP likewise to have high sensitivity but low specificity in a population (N=28) of middle-aged females. Another validation study (Zambotti 2015) found Jawbone UP to have good agreement levels with PSG for total sleep time, sleep efficiency, and wake after sleep onset in a population of healthy adolescents (N=65). The study concluded with a recommendation that additional validation is needed in other age groups and clinical populations before advocating use of these inexpensive and easy-to-use devices in clinical sleep medicine and research. Our study will involve a diverse clinically relevant (referred for suspected sleep disorder, e.g., sleep apnea, hypersomnia, insomnia, restless leg syndrome) population of a wide age range (18-80). Our study will also involve a substantially larger sample size (N=100 per device). Lastly, we will use the latest available version of each device as design and detection / discrimination algorithms are continually evolved and refined.

3. PATIENT SELECTION

3.1 Inclusion Criteria

Attending a regularly scheduled overnight PSG sleep study at the WCMC Center for Sleep Medicine

Age: 18-80 years old

3.2 Exclusion Criteria

Participant-reported pregnancy

Inability to provide informed consent

Any medical or psychiatric condition that, in the opinion of the site investigator, would interfere with study participation

4. STATISTICAL CONSIDERATIONS

4.1 Study Design

Participants (male or female, 18-80 years old) will be recruited from patients attending a regularly scheduled overnight PSG sleep study at the WCMC Center for Sleep Medicine. They will be approached at the time they present to the sleep clinic.

A researcher will meet with the prospective participant and verbally explain the purpose of the study and what participation involves. If a willingness to participate is expressed, the researcher will provide the prospective participant with a hard copy of the informed consent form to read through and answer any questions.

The investigator will review the participant's record for additional data including the reason for the PSG study. As the devices being evaluated rely on limb movement for wake / sleep determination, it may be of interest to examine the results of participants with certain sleep disorders, e.g., periodic limb movement, separately (in previous validation studies, only healthy participants were studied).

After providing informed consent, participants will be randomized to one of three devices and a secondary device if consenting to wear two devices. Allocation to the primary device will be 1:1:1. A randomization list will be generated by an outside investigator and maintained at the sleep center. The investigators will not be aware of the next assignment in advance of randomization.

The participant will be fitted with the assigned commercial sleep monitoring devices on the wrist of their non-dominant hand when prepared for the sleep study by a technician. Participants willing to wear more than one may be fitted with a second sleep monitoring device, e.g., a Fitbit and a Jawbone UP or a Fitbit and a Microsoft Band.

The commercial sleep monitoring devices have a manual mode that allows the user to start and stop tracking sleep. When the sleep study is to begin and the participant's room lights are turned off, the researcher will activate sleep tracking mode on the device(s). At the conclusion of the sleep study when the participant's room lights are turned on, the researcher will deactivate sleep tracking mode on the device(s).

Sleep monitoring device(s) will be removed from the participant's wrist in the morning at the conclusion of the sleep study.

Data will be transferred from the sleep monitoring device(s) for later statistical analysis and comparison with scored polysomnogram.

Sleep monitoring device(s) will be cleaned (using antiseptic wipes), reset, recharged, and otherwise prepared for reuse.

4.2 Sample Size / Accrual Rate

A total of 300 participants will be recruited (100 per device).

This number is based on several considerations. First, a similar sample size was used in previous validation studies of similar sleep monitoring devices: N=150 in our 2014 validation of Zeo Personal Sleep Coach (Protocol Title: Validation of a New Sleep Measurement Device / IRB Protocol Number: 1210013184), N=99 in Kaplan's 2014 validation of Zmachine, and N=65 in Zambotti's 2015 validation of Jawbone UP. In addition, this number is guided by logistical considerations of average laboratory throughput. Based on our experience, we realistically expect to recruit an average

of 5 participants per night, 30 participants per week, or 120 participants per month.

This sample size will allow us to describe the difference in total sleep time with a 95% confidence interval of +/-2.5 minutes, assuming a SD of 12.4 as was observed in a prior study of Fitbit.

The total number of complete data sets collected may be less than 300 as a participant may be unable to fall asleep while in the lab or may elect to end the study prematurely. However, the number of such eventualities should be low, e.g., < 5%. Participants with incomplete data will not be replaced.

4.3 Analysis

Overnight PSG will be performed at the WCMC Center for Sleep Medicine. Multichannel data acquisition equipment (Grass Technologies TWin PSG Clinical Software) will be used to record four channels of electroencephalography (O1/O2, C3/C4), chin electromyography, bilateral electrooculography, electrocardiogram, and anterior tibialis electromyography. Audio and video will be simultaneously digitally recorded. All PSG records will be manually scored and double-checked by analysts blind to the results of the devices being evaluated, using American Academy of Sleep Medicine criteria. Principal PSG outcome measures are sleep efficiency and total sleep time.

Analysis will be conducted by comparing data gathered by the consumer sleep monitoring device(s) against manually scored polysomnogram. Epoch-by-epoch data will be gathered into pooled contingency tables for analysis. Data will be analyzed according to several degrees of rigor:

Stage-by-stage (with N1 and N2 combined into Light)
Wake / REM / NREM (with N1, N2, N3 combined into NREM)
Wake / Sleep (with N1, N2, N3, and REM combined into Sleep)

Percent agreement, sensitivity, and specificity will be calculated for pooled epoch data (overall and for each stage: Wake, Light, Deep, and REM).

Cohen's kappa will be calculated. What kappa value is considered indicative of weak or strong correlation is subjective and different classifications have been proposed:

A kappa value of:

0 – 0.2 is considered essentially slight/no agreement
0.2 – 0.4 is considered fair/low agreement
0.4 – 0.6 is considered moderate agreement
0.6 – 0.8 is considered substantial/high agreement
0.8 – 1.0 is considered nearly perfect agreement

Agreement levels of several sleep metrics will be assessed using Intraclass Correlation Coefficients: Total Sleep Time (TST), Sleep Onset Latency (SOL), Latency to Persistent Sleep (LPS), Latency to REM (REML), Wake After Sleep Onset (WASO), Sleep Efficiency (SE), and time spent in particular sleep stages (Light, Deep, REM).

Positive Predictive Value (PPV) will be calculated for each stage type: Wake, Light, Deep, REM.

Data will be analyzed using SPSS software.

4.4 Study Monitoring

The investigators will follow weekly accrual statistics including number of participants approached for participants, proportion giving consent, and reasons for refusal. We expect 20-30 participants to enroll per week. We expect to complete accrual in mid-2016.

We are not planning formal interim analyses. We will perform interim evaluation of study data after 10 participants have worn each device to assure that there are no issues with data transfer and to develop the detailed analysis plan. We do not expect any adverse events related to study participation. The investigators will review the occurrence of any device related adverse events. An outside review may be convened if the investigators are concerned about the occurrence of adverse events.

REFERENCES

Eastwood, P., et al. (2006). Relationship Between Upper Airway Inspiratory Pump Muscle Force in Obstructive Sleep Apnea, *Journal of Sleep Medicine*, 130(6), 1757-64.

Fogel, R., et al. (2002). The Male Predisposition to Pharyngeal Collapse, *American Journal of Respiratory and Critical Care Medicine*, 166, 1388-95.

Eastwood, P., et al. (2011). Treating Obstructive Sleep Apnea with Hypoglossal Nerve Stimulation, *Sleep*, 34(11), 1479-86.

Peppard, P., Young, T., Palta, M., & Skatrud, J. (2000). Prospective study of the association between sleep-disordered breathing and hypertension, *New England Journal of Medicine*, 342, 1378-84.

Yaggi, H., Concato, J., Kernan, W., Lichtman, J., Brass, L., & Mohsenin, V. (2005). Obstructive sleep apnea as a risk factor for stroke and death, *New England Journal of Medicine*, 353, 2034-41.

Eikermann, E., Jordan, A., Chamberlin, N., Gautam, S., Wellman, A., Lo, Y., White D., & Malhotra, A. (2007). The Influence of Aging on Pharyngeal Collapsibility During Sleep, *Chest*, 131, 1702-9.

Campos-Rodriguez, F., Martinez-Garcia, M., Martinez, M, Duran-Cantolla, J., Peña, L., Masdeu, M., Gonzalez, M., Campo, F., Gallego, I., Marin, J., Barbe, F., Montserrat, J., & Farre, R. (2013).

Association between obstructive sleep apnea and cancer incidence in a large multicenter Spanish cohort, *American Journal of Respiratory Critical Care Medicine*, 187(1), 99-105.

Martinez-Garcia, M., Campos-Rodriguez, F., Martinez, M, Duran-Cantolla, J., Peña, L., Masdeu, M., Gonzalez, M., Campo, F., Gallego, I., Marin, J., Barbe, F., Montserrat, J., & Farre, R. (2014).

Obstructive sleep apnea is associated with cancer mortality in younger patients, *Journal of Sleep Medicine*, 15(7), 742-8.

Harvard Medical School. *The Price of Fatigue: The Surprising Economic Costs of Unmanaged Sleep Apnea*, McKinsey & Company, December 2010.

Busetto, L., Enzi, G., Inelmen, E., Costa, G., Negrin, V., Sergi, G., & Vianello, A. (2005).

Obstructive sleep apnea syndrome in morbid obesity: effects of intragastric balloon, *Chest*, 128(2), 618-23.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Arlington, VA: American Psychiatric Publishing.

Ohayon, M. (2002). Epidemiology of insomnia: What we know and what we still need to learn, *Sleep Medical Reviews*, 6(2), 97-111.

Kamel, N., & Gammack, J. (2006). Insomnia in the Elderly: Cause, Approach, and Treatment, *The American Journal of Medicine*, 119, 463-9.

Jansson-Froejmark, M., & Lindblom, K. (2008). A bi-directional relationship between anxiety and depression and insomnia? A prospective study in the general population, *Journal of Psychosomatic Research*, 64(4), 443-9.

Vgontzas, A., et al. (2009). Insomnia with objective short sleep duration is associated with a high risk for hypertension, *Sleep*, 32(4), 491-7.

Capuccio, F., D'Elia, F., Strazzullo, P., & Miller, M. (2010). Sleep Duration and all-cause mortality: a systematic review and meta-analysis of prospective studies, *Sleep*, 35(5), 585-92.

Neubauer, D. (1999). Sleep problems in the elderly, *American Family Physician*, 59(9), 2551-60.

Montgomery-Downs, H., et al. (2012). Movement toward a novel activity monitoring device, *Sleep Breath*, 16(3), 913-7.

Zambotti, M., et al. (2015). Evaluation of a consumer fitness-tracking device to assess sleep in adults, *Chronobiol Int*., 32(7), 1024-8.

Zambotti, M., et al. (2015). Validation of Sleep-Tracking Technology Compared with Polysomnography in Adolescents, *Sleep*, 38(9), 1461-8.

Kaplan, R., e al. (2014). Performance evaluation of an automated single-channel sleep–wake detection algorithm, *Nature and Science of Sleep*, 6, 113-22.