The Development and Accuracy of the THIM Device for Estimating Sleep and Wakefulness

by

Hannah Scott

Thesis
Submitted to Flinders University
for the degree of

Doctor of Philosophy

College of Education, Psychology and Social Work

July 2020

Table of Contents

Dissertation Summary

Chronic insomnia is a pervasive and burdensome sleep disorder that is not being adequately treated at present in Australia. There is a need for practical yet effective treatments for insomnia to address this serious personal and societal burden. Intensive Sleep Retraining (ISR) is a brief but effective laboratory-based behavioural treatment for sleep onset insomnia that involves a series of rapid sleep onsets facilitated by near-total sleep deprivation. In conjunction with Re-Time Pty. Ltd., we have developed a wearable device called THIM which promises to administer ISR in the home environment. To successfully administer ISR, THIM must be able to accurately estimate sleep onset and wake the patient at the appropriate time to achieve rapid sleep onsets. Additionally, THIM can passively monitor sleep and wakefulness during the sleep period. If THIM accurately monitors sleep, this data could be incorporated into insomnia treatment. This dissertation discussed the development and accuracy of THIM for estimating sleep onset and for monitoring nocturnal sleep and wakefulness compared to the gold standard of objective sleep measurement, polysomnography (PSG).

Chapter 2 was the first systematic review to examine the accuracy of wearable devices for the estimation of sleep onset latency (SOL) compared to PSG. The review concluded that devices measuring behavioural sleep onset were most suitable for the administration of ISR because they consistently overestimated PSG-determined SOL, but with little variability between individuals compared to other wearable devices. This finding justified the method that THIM relies upon to estimate sleep onset for the purposes of ISR: the measurement of behavioural responsiveness to minimal intensity vibratory stimuli.

Chapter 3 described the development and accuracy of the THIM device for

estimating SOL in comparison to PSG. From the findings of Study 1, the THIM algorithm was refined and its accuracy confirmed in Study 2. THIM showed much closer agreement to PSG than other behavioural devices that use auditory stimuli and larger hand/wrist movements as behavioural responses. The final version of the algorithm had a discrepancy with PSG-SOL of less than one minute on average, which was consistent across two nights of testing. THIM appears to be accurate enough to administer ISR, but its accuracy for individuals with insomnia needs to be investigated in future research.

Chapter 4 presented a quantitative electroencephalography analysis of the data from Chapter 3 to characterise sleep microstructure through a more fine-grained lens than traditional 30-second epoch sleep staging. This study was the first to examine the correspondence between sleep microstructure and responses to minimal intensity vibratory stimuli during the process of falling asleep. The findings indicated that participants had increases in higher frequency brainwaves (alpha, sigma and beta) when they responded to the vibratory stimulus and increases in delta activity when they did not respond to the stimulus across all sleep stages. This suggests that a shift to wakefulness or an arousal occurred prior to, or coincident with, the onset of the vibratory stimulus, which may explain why participants responded to the stimulus. Thus, THIM was able to detect brief arousals during sleep stages that traditional sleep scoring criteria would overlook, which were a common occurrence during N1 sleep. THIM is accurate for detecting brief periods of wakefulness. These findings further supported the conceptualisation of N1 sleep as a transitional, fluctuating state between wake and sleep.

Chapter 5 described the refinement of the THIM sleep tracking function. It was the first study to test the accuracy of THIM for estimating sleep and wakefulness

over a nocturnal sleep period compared to PSG. THIM was similar in accuracy for estimating sleep and wake compared to wrist-based actigraphy devices. However, THIM showed a greater tendency to underestimate total sleep and overestimate total wake compared to other actigraphy devices. There was high variability in the accuracy of THIM between individuals, yet this was not explained by whether participants were good or poor sleepers.

Chapter 6 extended the findings of Chapter 5 by examining the consistency in the accuracy of THIM over three nights compared to PSG. THIM showed consistently high sensitivity, specificity and accuracy compared to PSG across all nights. However, THIM produced consistently and significantly lower estimations of sleep efficiency due to higher estimations of wake after sleep onset. The improvement of the accuracy of THIM for estimating wake is required to render the device useful for objective sleep monitoring.

Together, the findings of this dissertation indicate that THIM may be able to successfully administer ISR. The findings also suggest that improvements to the THIM sleep tracking algorithm are required for the device to provide accurate enough sleep tracking data to support the treatment of insomnia. Future research is required to investigate the efficacy of THIM-administered ISR and the accuracy of THIM sleep tracking in the home environment for individuals with insomnia. The long-term goal of this research program is to create an effective yet practical device to support the treatment of insomnia. This dissertation is the proof-of-concept step in the development of THIM.

Declaration

I certify that this thesis:

1. does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; and

2. to the best of my knowledge and belief, does not contain any material previously published or written by another person except where due reference is made in the text.

Hannah Scott, B. Psych. (Hons.) 20th March 2020

Acknowledgements

This dissertation would not have been possible without the people that I am very fortunate enough to be surrounded by. They are too numerous to mention, but I will try.

To Leon Lack – you may recall the first time that we met in your office in 2015. I was a third-year placement student. You talked extensively about the project that I would be working on (CRC), and embarked on many tangents that I soon came to realise was typical of your conversations and crucial for generating new ideas. I have enjoyed and benefitted immensely from every conversation we have had since then. Your constant enthusiasm is admirable and aspirational. Thank you for your guidance and support through all of the (many) hurdles that we faced together through the past few years. I am honoured to have completed this PhD under your supervision.

To Nicole Lovato – thank you for your guidance throughout this challenging degree. You went above and beyond the call of duty. Your steadiness and humour got me over the line in the last couple of months, thank you.

To my sleep psychology family:

Kelsey– I was a small third-year student when we first met. Thank you for helping me become a small PhD student. More than that, you always found a way to be supportive and my buddy through this 'journey', even through the hard times. Thank you for your friendship.

Gorica, Claire, Tess, Alex, and Jenny – thank you for your knowledge, your helping hands, and your good humour. I have really enjoyed working with you all.

Ashwin, Megan, and Alex C – this research would not have been possible without your hard work. I admire your dedication to research and to your studies. I am very fortunate to have worked with you all and am looking forward to seeing what you accomplish next.

Placement students and volunteers – thank you for your enthusiasm to learn, your valuable assistance through data collection, and your willingness to help me stay awake during overnights in the lab. Much appreciated.

A special thank you to Michaela for scoring the EEG data, and Bastien and Claire for their help with qEEG analyses.

To the Adelaide Institute for Sleep Health (AISH) team - I consider myself incredibly

lucky to have been welcomed into the AISH family. You are all outstanding researchers and human beings. The culture at AISH is truly special – thank you for letting me be part of it.

To Re-Time Pty. Ltd. (Ken, Louise, Ben and Vera) – thank you for the opportunity to collaborate on THIM. To see our conversations at the beginning of 2017 come to fruition was rewarding beyond belief. I cannot thank you all enough for your expertise and hard work that made THIM become a reality.

To my wider Flinders University family: thank you for the guidance, expertise and resources to conduct this research. I will endeavour to thank you all individually soon. A special thank you to Peter Catcheside and Sarah Cohen-Woods for reviewing and contributing to my PhD research proposal. Your feedback was incredibly valuable and shaped this research.

Thank you to the Australian Government for the scholarship that made this candidature possible.

To my non-university family:

Mia – Thank you for putting up with me and listening (or pretending to listen) to my incessant ramblings. Your constant support and what appears to be an infinite amount of strength got us through this degree. I can't wait for the next years of our lives together.

Dad – Thank you for supporting me through my education. You gave me everything that I needed to grasp opportunities that were not available to you. What a gift.

Mum, Gemma, and friends – thank you for nodding along and pretending to understand my research. I appreciate the support immensely.

List of Tables

List of Figures

Glossary of Abbreviations

AASM	American Academy of Sleep Medicine
ANOVA	Analysis of Variance
BMI	Body mass index
CBT-I	cognitive behavioural therapy for insomnia
CST	Consumer sleep technology
ECG	electrocardiography
EEG	electroencephalography
EMG	electromyography
EOG	electrooculography
ISI	Insomnia Severity Index
ISR	Intensive Sleep Retraining
LMM	Linear Mixed Modelling
M	Mean
MeSH	Medical Subject Heading
MSLT	Multiple sleep latency test
MWT	Maintenance of wakefulness test
N	Sample size
N1	Non-rapid eye movement Stage 1
N2	Non-rapid eye movement Stage 2
OSA	Obstructive sleep apnea
PAT	peripheral arterial tone
PLMS	Periodic limb movements of sleep
PPG	photoplethysmography
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PSG	polysomnography
qEEG	quantitative electroencephalography
SCT	stimulus control therapy
SD	Standard deviation
SE	Standard error

SOL	sleep onset latency
SRT	sleep restriction therapy
TST	total sleep time
WASO	wake after sleep onset

List of Manuscripts and Publications from this Dissertation

Scott, H., Lack, L., & Lovato, N. (2020). A systematic review of the accuracy of wearable devices for estimating sleep onset. *Sleep Medicine Reviews, 49*(101227), 1-12. doi: 10.1016/j.smrv.2019.101227.

Scott, H., Whitelaw, A., Canty, A., Lovato, N., & Lack, L. (2020). The accuracy of the THIM wearable device for estimating sleep onset latency. *under review*.

Scott, H., Lovato, N., & Lack, L. (2020). The development and preliminary validation of the THIM wearable device for estimating sleep and wakefulness. *under review*.

List of Published Abstracts and Referred Conference Proceedings from this Dissertation

Scott, H., Whitelaw, A., Canty, A., Lovato, N., & Lack, L. The Accuracy of a Novel Sleep Ring Device for Estimating Sleep Onset with Good and Poor Sleepers. *SLEEP 2020*, 13-17th June 2020, Philadelphia: US.

Scott, H., Whitelaw, A., Canty, A., Lovato, N., & Lack, L. The Accuracy of a New Sleep Ring Device for Tracking Sleep and Wakefulness Overnight Using Actigraphy. *SLEEP 2020, 13-17th June 2020, Philadelphia: US.*

Scott, H., Whitelaw, A., Canty, A., Lovato, N., & Lack, L. The accuracy of the THIM device for estimating sleep onset with good and poor sleepers. *Australasian Sleep Association Sleep Down Under*, 16-20th October 2019, Sydney: Australia.

Scott, H., Lovato, N., & Lack, L. The accuracy of THIM for passively estimating sleep and wakefulness overnight with good and poor sleepers. *Australasian Sleep Association Sleep Down Under*, 16-20th October 2019, Sydney: Australia.

Scott, H., Lack, L., & Lovato, N. The development of the THIM wearable device and smartphone application for the treatment of chronic insomnia. *Global Telehealth 2019*, 5th July 2019, Adelaide: Australia.

Scott, H., Lack, L., Lovato, N., & Whitelaw, A. The validity of a novel wearable device for estimating sleep onset. *Australasian Sleep Association Sleep Down Under*, 19th October 2018, Brisbane: Australia.

Scott, H., Lack, L., Lovato, N., & Whitelaw, A. The validity of the THIM wearable device for estimating sleep onset. *Adelaide Sleep Retreat*, 8th November 2018, Brisbane: Australia.

Scott, H. The development of a behavioural device for measuring sleep onset in the home environment. *Adelaide Sleep Retreat*, 15th November 2017, Adelaide: Australia.

Chapter 1: Introduction and Dissertation Aims

Chronic Insomnia

Chronic insomnia is the persistent inability to sleep despite adequate sleep opportunity, leading to negative daytime consequences. According to the International Classification of Sleep Disorders (ICSD-3), a diagnosis of Chronic Insomnia Disorder requires that the individual experiences insomnia symptoms and associated daytime impairments, such as fatigue and mood disturbances, at least three times per week, persisting for at least three months (American Academy of Sleep Medicine, 2014; Roth, 2007). Difficulties in initiating sleep within 30 minutes after attempting sleep is referred to as sleep onset insomnia (Freedman & Sattler, 1982). Additional presentations of insomnia include difficulties maintaining sleep (sleep maintenance insomnia, characterised by >30 minutes of wakefulness after sleep onset) and difficulties with waking too early from sleep (early morning awakening insomnia). Some patients may experience one or a combination of these sleep difficulties and the ICSD-3 does not differentiate between these presentations in the diagnosis of Chronic Insomnia Disorder (American Academy of Sleep Medicine, 2014). The diagnostic criteria also do not differentiate between primary and secondary insomnia, recognising that insomnia, regardless of the cause, is a serious burden that warrants treatment in its own right.

Chronic insomnia is the most common sleep disorder (referred herein as insomnia) and is associated with serious adverse health consequences (Deloitte Access Economics, 2017). The prevalence of insomnia is difficult to determine because epidemiological studies have produced a range of estimations, with this variability potentially explained by the use of various criteria to define insomnia. When using the ICSD-3 criteria, an estimated 11.3% of the Australian population

experience insomnia (Deloitte Access Economics, 2017). This compares to an estimate of 20% for insomnia prevalence in Australia (Adams et al., 2017). The prevalence of insomnia in Australia is comparable to similar countries, including New Zealand (Arroll et al., 2012), Spain (Ohayon & Sagales, 2010), and the UK (Morphy, Dunn, Lewis, Boardman, & Croft, 2007). People with insomnia typically experience dysphoria, fatigue and reduced quality of life (Benca, 2005; Kyle, Morgan, & Espie, 2010). There is also increasing awareness of medical comorbidities associated with insomnia, including a 2.1 times greater risk of depression (Baglioni et al., 2011), and more frequent health problems and hospitalisations (Wade, 2010). Insomnia is also associated with reduced work productivity (Daley et al., 2009b). Adams et al. (2017) observed that 20% of those with insomnia reported missing 1-2 workdays in the past month. Similarly, Reynolds et al. (2017) reported from the same Australian survey that 29% of adults reported making errors at work due to sleepiness or sleep problems. It is clear that insomnia has a substantial impact on health and work productivity.

Insomnia is also a serious economic burden. In 2016-2017, the estimated cost of inadequate sleep in Australia was $66.3 billion ($8,968 per person with inadequate sleep) comprised of $26.2 billion in direct costs (Deloitte Access Economics, 2017). This is higher than previous estimates of approximately $36 billion in Australia in 2013, with a third of that attributable to insomnia, or approximately $8,000/year per person with insomnia (Hillman & Lack, 2013). These are comparable to global estimates, such as approximately $5,010/year per person with insomnia in North America (Daley, Morin, LeBlanc, Grégoire, & Savard, 2009a). Given the associated personal and societal costs, insomnia is a substantial burden that warrants the development of effective and accessible treatments to adequately

address the disorder. This project aims to reduce this personal and societal burden by delivering a device to effectively treat insomnia in the home environment.

Current Treatments for Insomnia

The most common medical treatment for insomnia is pharmacotherapy. In a recent survey of Australian family physicians, 90% of patients presenting with insomnia were prescribed medication, predominantly benzodiazepines (Miller et al., 2017). Despite their frequent use, hypnotic medication is not best practice for insomnia (Riemann & Perlis, 2009). There are problems of tolerance, dependence, and side-effects such as rebound insomnia, deleterious effects on respiratory, hepatic, renal, and cardiac disorders, and daytime effects of dysphoric mood and sleepiness (Benca, 2005). Impaired motor and intellectual functioning are also experienced by individuals with insomnia, which is particularly problematic for older adults because this places them at a greater risk of falls (Stone, Ensrud, & Ancoli-Israel, 2008). Epidemiological evidence suggests that chronic use of benzodiazepines provides little long-term benefit to sleep (Ohayon & Caulet, 1995), and is associated with increased morbidity and mortality (Kripke, 2016; Lovato & Lack, 2018). The prevalence of prescriptions for pharmacological treatments has remained constant since 2000, potentially due to their ease of use, perceived rapid treatment response, and dependency associated with long-term use (Miller et al., 2017). These findings emphasise the need for practical but effective treatments for insomnia.

The recommended first-line treatment for insomnia is Cognitive Behavioural Therapy for Insomnia ([CBT-I], Edinger & Carney, 2014; Morin & Espie, 2007; Perlis, Jungquist, Smith, & Posner, 2005). This treatment incorporates a range of cognitive and behavioural techniques, and Stimulus Control Therapy (SCT) is the most widely

studied and endorsed behavioural component (Bootzin & Epstein, 2011; Bootzin, 1972). The aim of SCT is to decrease the duration of time spent awake in bed. Instructions include asking patients to only go to bed when they feel sleepy, to get out of bed if not asleep within 15 minutes, to go back to bed only when sleepy again, and to maintain the same wake-up time regardless of sleep duration. Complying with these instructions decreases total sleep time (TST) over the first few weeks of therapy (3-4 weeks), thereby increasing homeostatic sleep drive until eventually, sleep becomes inevitable. Following the SCT instructions is thought to extinguish the conditioned insomnia response that has become a habit learned through repeated failures to initiate sleep in the past (Lack, Scott, Micic, & Lovato, 2017; Perlis, Shaw, Cano, & Espie, 2011). Consequently, over time, the bedroom environment becomes conducive for sleep rather than wakefulness.

The effectiveness of SCT, and indeed CBT-I, has been shown repeatedly in clinical trials (Espie, Lindsay, Brooks, Hood, & Turvey, 1989; Riedel et al., 1998). A recent meta-analysis including 87 randomised controlled trials reported that CBT-I is effective for treating insomnia, when utilising either the full CBT-I treatment package or the main behavioural components (van Straten et al., 2018). CBT-I results in moderate-large improvements in sleep onset latency (SOL), and small-moderate effects in TST and wake after sleep onset ([WASO], Morin et al., 1999; Smith et al., 2002). Compared to pharmacological treatments, CBT-I produces comparable short-term improvements in sleep, but a considerable advantage of CBT-I is that it results in greater long-term benefits beyond treatment (Morin, LeBlanc, Daley, Gregoire, & Merette, 2006). CBT-I is also effective for treating insomnia in the presence of other health conditions, including obstructive sleep apnea [OSA], Sweetman, Lack, Lambert, Gradisar, & Harris, 2017), many psychiatric conditions (Taylor & Pruiksma,

2014), and medical conditions in older adults (Lovato, Lack, Wright, & Kennaway, 2014; McCurry, Logsdon, Teri, & Vitiello, 2007; Rybarczyk et al., 2005).

Shifting insomnia healthcare practice towards CBT-I is a priority for sleep medicine. However, the main barrier to the achievement of this goal is the lack of resources available to treat patients using CBT-I. To undergo face-to-face CBT-I in Australia at present, patients experiencing insomnia symptoms typically visit their general practitioner in the first instance. After screening for potential physical health conditions or other sleep disorders, the physician may refer the patient to a sleep specialist. Waitlists to see sleep specialists can be lengthy and access to specialised sleep disorder specialists is limited in remote areas of Australia. Furthermore, few physicians are qualified to effectively deliver CBT-I. Of those sleep specialists who claim to treat insomnia, it is unknown whether they treat insomnia effectively, particularly whether they use CBT-I in an efficacious manner. Therefore, despite its high efficacy, CBT-I is not readily available to treat many individuals with insomnia. As such, many individuals seek - and/or their general practitioner recommends - pharmacological treatments, complementary alternative medicines, or sleep hygiene instructions that produce minimal long-term therapeutic benefits for insomnia (Chung et al., 2018; Miller et al., 2017; Wilt et al., 2016). Even though CBT-I is the recommended first line treatment for insomnia, the current healthcare model and lack of resources prevents this recommendation becoming a reality for the overwhelming majority of individuals with insomnia.

Digital Treatments for Insomnia

With the shortfall of necessary resources to effectively treat sleep disorders and a boom in the use of healthcare technology, there has been a rise in the number of digital sleep products available to consumers. These products span smartphone

applications, web-based programs, wearable devices, and ‘nearable’ devices placed near the sleeping individual (Bianchi, 2017). Many of these products claim to improve sleep, but few have robust empirical evidence publicly available to support their efficacy. These consumer sleep technologies (CSTs) are often classed as lifestyle/entertainment devices, thereby avoiding the need for validation data required for the approval of medical devices under the United States Food and Drug Administration regulatory authority. As such, claims of efficacy are often unsupported by evidence.

Consumers may believe marketing claims and purchase products that do not effectively treat sleep and have the potential to cause harm. Gavriloff et al. (2018) gave individuals with insomnia sham feedback from a sleep tracking device following a night’s sleep. Those who received negative feedback (low sleep efficiency) had reduced daytime function (d = 0.79) and increased sleepiness and fatigue (d = 0.55) the following evening compared to those who received positive feedback (high sleep efficiency), indicating that inaccurate feedback provided by sleep trackers can impact daytime functioning. In a position statement regarding CST, the American Academy of Sleep Medicine (AASM) recommended that CSTs should be validated if this technology is to be used in the diagnosis or treatment of sleep disorders (Khosla et al., 2018). Similarly, the Sleep Research Society (SRS) have made recommendations on the conduct and reporting of studies attempting to validate CSTs in the hope of generating further research in this area (Depner et al., 2019). This dissertation is the first step in the validation of the THIM wearable device for treating insomnia.

Some online CBT-I programs have strong empirical evidence available to support their efficacy. One popular internet-based treatment developed by insomnia

researcher, Colin Espie, is known as Sleepio (Sleepio, 2019). This treatment program delivers six weeks of CBT-I via web- and mobile-based platforms. Incorporating participants' sleep diaries, the program tailors the treatment to the individuals' current sleep and treatment goals. Participants are encouraged to contribute to the Sleepio community by seeking and providing feedback on chat forums with other Sleepio participants. In a large randomised controlled trial (N = 164), Espie et al. (2012) reported greater improvements in sleep diary outcomes and daytime functioning for the Sleepio treatment group compared to a placebo imagery relief therapy group and a waitlist control group. Effects sizes for change from baseline to two-month follow-up for sleep efficiency (*d* = 1.37), total wake time (*d* = 1.21) and SOL (*d* = 0.80) were similar to those found with face-to-face CBT-I (Espie et al., 2012). In relation to clinical significance, approximately 75% of those in the Sleepio treatment group had a sleep efficiency >80% post-treatment compared to approximately 30% and 20% in the imagery relief and waitlist control groups, respectively. In this study, 82% of participants completed the entire Sleepio program and 75% completed the two-month follow up assessment. Such high treatment adherence is considerably better than other studies investigating online CBT-I programs.

While online CBT-I programs can produce substantial benefits in sleep, treatment adherence in other studies investigating these treatments are often unsatisfactory. Espie et al. (2019) observed small improvements in psychological wellbeing and quality of life measures for individuals using Sleepio compared to a sleep hygiene control group in a community sample of 1,000 adults. Of those randomly assigned to the Sleepio group, 80.8% logged on for at least one treatment session, but only 48.4% completed all 6 sessions. Freeman et al. (2017) found that

69% of participants in the Sleepio group completed at least one session and only 18% completing all six sessions. Cheng et al. (2019) similarly observed that 38% of participants completed the Sleepio intervention.

This trend for low treatment adherence with Sleepio is also found with other online treatment programs. SHUT-I (Sleep Healthy Using the Internet) is a similar six-week non-tailored CBT-I program (BeHealth Solutions, 2018; Thorndike et al., 2008). In a randomised controlled trial of 303 adults with insomnia, SHUT-I produced substantial improvements in sleep outcomes that were maintained at 12-month follow up assessment, including in SOL, d = 1.41, and Insomnia Severity Index (ISI) questionnaire scores, d = 2.32 (Ritterband et al., 2017). At 12-month follow up, an intention to treat analysis revealed that 69.7% of the SHUTi group were classified as treatment responders (a reduction of >7 points on the ISI) compared to 43.0% of the patient education control group. However, treatment adherence was low, with only 60.3% of participants assigned to SHUT-I completing the program.

Considering that these programs are low cost and use less resources than face-to-face CBT-I, online CBT-I programs produce acceptable improvements in sleep. Coupled with practical administration, these programs may be useful to address the large-scale problem of providing effective insomnia treatment for little cost. In fact, Sleepio is now provided to England residents free-of-charge under the publicly-funded National Health Service (NHS North West London, 2018). However, low treatment adherence is an issue that needs to be resolved to improve treatment outcomes (Matthews, Arnedt, McCarthy, Cuddihy, & Aloia, 2013).

One potential reason for low adherence is the inability of individuals to endure behavioural components of the therapy, such as SCT or sleep restriction therapy (SRT). Both of these behavioural treatments build homeostatic sleep drive over the

first 3-4 weeks of treatment to the point where sleeping in the bedroom environment becomes inevitable. This necessary lag in treatment response is associated with early treatment sleepiness, discomfort from sustained sleep loss and great difficulty in changing lifestyle habits (Kyle, Morgan, Spiegelhalder, & Espie, 2011). High levels of motivation are necessary to overcome these typical difficulties experienced in the first few weeks of treatment, but this high motivation is difficult to achieve and maintain throughout treatment. In clinical practice with the patient unsupervised at-home, SCT often does not lead to the promised therapeutic benefits found in the closely supervised clinical trials (Matthews et al., 2013). Whilst clinicians may be able to provide the necessary support for many individuals to overcome these challenges in clinical practice, this degree of support is lacking in online treatment programs and may contribute to the low treatment adherence.

Developing treatments that *avoid* these challenges by producing more rapid improvements in sleep would presumably be more acceptable to individuals, likely lead to greater treatment adherence, and thereby produce greater therapeutic benefits. Such rapid treatments would also require less support from clinicians, reducing the burden on public healthcare resources. Relatedly, the allure of pharmacological treatments for insomnia may partly be due to the rapid therapeutic response (Miller et al., 2017). The development of an effective yet rapid behavioural treatment would be a useful and welcome technique for clinicians, researchers and consumers wanting a practical insomnia treatment.

Intensive Sleep Retraining

Intensive Sleep Retraining (ISR) is a promising alternative to other behavioural treatments for insomnia. ISR is a brief but rigorous behavioural treatment involving near-total sleep deprivation over 24 hours to facilitate a series of

rapid sleep onsets (Harris, Lack, Kemp, Wright, & Bootzin, 2012; Harris, Lack, Wright, Gradisar, & Brooks, 2007). The patient is required to lie in bed and attempt to fall asleep. After a brief period of light sleep according to polysomnography ([PSG], in the order of 2-3 minutes of sleep), the patient is awoken and given feedback about how long it took them to fall asleep. This process is known as a sleep onset trial. The patient is subsequently instructed to remain awake until the next half-hour timepoint, before attempting to fall asleep again on the next trial. Since brief episodes of light sleep (< 3 minutes) do not reduce homeostatic sleep drive (Tietzel & Lack, 2002), sleep deprivation is effectively maintained over the whole retraining session. It is the deprivation of recuperative sleep combined with a high circadian drive for sleepiness during the early hours of the morning that cause patients to fall asleep more rapidly with each subsequent sleep onset trial. Consequently, patients who report average pre-treatment SOLs > 60 minutes are able to fall asleep in < 5 minutes on dozens of attempts during the retraining session. This is thought to retrain patients to fall asleep more quickly by extinguishing the conditioned cortical arousal response hypothesised to interfere with the attempt to initiate sleep in the home environment (Lack, Scott, & Lovato, 2019).

Early pilot studies indicated that ISR produced immediate, significant and sustained improvements in sleep (Harris et al., 2007; Lack & Baraniec, 2002). For 17 participants with sleep onset insomnia, ISR significantly reduced SOL by 30 minutes and increased TST by 65 minutes (Harris et al., 2007). These improvements were maintained at two-month follow up. Although the ISR procedure may appear vexing, and participants sometimes had trepidations about whether they could manage the sleep loss, what they mainly experienced was increased sleepiness. This was not an aversive experience for insomnia patients as they rarely felt the sleepiness they

desperately sought, mitigating the common fear that their sleep mechanism is irreparably broken. Many participants were pleased with the experience of falling asleep quickly and no participants in any subsequent laboratory studies withdrew during the treatment procedure (Harris et al., 2012; Harris et al., 2007; Lack & Baraniec, 2002).

A randomised controlled trial compared the efficacy of ISR and SCT to a sleep hygiene control group (Harris et al., 2012). Seventy-nine individuals with chronic sleep onset insomnia were randomly allocated to four groups: ISR treatment group, SCT treatment group, a combined ISR followed by SCT treatment group, and a waitlist control group. Not only did ISR produce improvements that were as effective as what SCT achieved in four weeks, but it did so in as little as 24 hours. Treatment outcomes for the ISR group included a reduction in SOL (d = 0.61) and daytime impairment (d = 0.57), and an increase in TST (d = 0.53). Such a rapid treatment response is comparable to that promised by pharmacological treatments, yet ISR results in sustained improvements in sleep akin to CBT-I. At two-month follow up, 46.7% of participants in the ISR treatment group were classified as treatment responders, compared to 37.5% in the SCT group. The combination of ISR and SCT was particularly effective, with 61.1% of participants in this group identified as treatment responders at two-month follow up.

In practice, ISR may be more achievable and motivating for those who are unable to comply with SCT instructions over the more prolonged treatment period or other behavioural treatments, potentially leading to greater treatment adherence and subsequently, therapeutic benefit. With its strong efficacy and rapid therapeutic effect, ISR could become the treatment of choice for clinicians and consumers, if it can be effectively translated to the home environment.

To date, ISR has been administered as a laboratory-based treatment that requires the patient to remain in a sleep laboratory for 24 hours whilst undergoing PSG recording. A trained sleep technician is required to monitor PSG in real time to detect sleep onset during each trial so that the patient can be woken at the appropriate time. The prompt detection of sleep onset and subsequent waking of the patient is necessary to maintain a high homeostatic sleep drive to ensure the continuation of the rapid sleep onset experiences thought to be essential to the efficacy of the treatment (Lack et al., 2019; Lack et al., 2017). Consequently, the laboratory-based ISR procedure is costly, impractical for widespread implementation and not readily available to most insomnia patients. An alternative administration method is needed to translate this rapid and effective behavioural treatment to the home environment to make it viable for addressing the insomnia burden on a large scale.

Translation of Intensive Sleep Retraining to the Home Environment

Sleep On Cue

The Sleep On Cue smartphone application (app) was designed to translate the ISR protocol to the home environment (MicroSleep, 2015). To undergo ISR, the individual lies down in bed at their typical bedtime whilst holding their smartphone and wearing earphones, and attempts to fall asleep. The app emits a faint tone stimulus via the earphones approximately every 30 seconds and the individual is required to gently move the smartphone in response. When the individual fails to respond to the tone stimulus, the app assumes that they have fallen asleep and emits a strong vibration to wake the individual. The app and similar devices utilising the stimulus-response method of estimating sleep onset are known to be accurate compared to PSG. Behavioural responses to auditory stimuli tend to cease 2–3

minutes after the onset of electroencephalography (EEG) non-rapid eye movement Stage 1 (N1) sleep (Connelly, 2004; Lack & Mair, 1995; Scott, Lack, & Lovato, 2018). Therefore, the individual is likely to experience the desired 2-3 minutes of light sleep before they are woken up: aligning with the laboratory-based ISR administration protocol. After the strong vibration has woken the individual, they have a short break before initiating the next trial, and this cycle continues overnight until the following morning.

Because the administration of the ISR protocol is similar to the laboratory-based method, the Sleep On Cue app should be successful in administering the ISR procedure in the patient's home without the need for PSG or sleep technicians. The Sleep On Cue app developers have reported considerable anecdotal support from their customers, but experimental confirmation of treatment efficacy is still only preliminary. A pilot study compared sleep outcome measures at baseline to post-treatment and four-week follow-up in twelve sleep onset insomnia patients (Mair, Scott, & Lack, 2020). The study also tested a modified ISR protocol whereby sleep onset trials were only separated by a brief six-minute break. Using this modified protocol, participants completed an average of 36.33 (*SD* = 6.72) trials within a 12-hour nocturnal treatment session. Compared to baseline, strong improvements were demonstrated for SOL (*d* = 0.75), sleep efficiency (*d* = 0.77), and insomnia severity on the ISI (*d* = 1.82) at four weeks post-treatment (Mair et al., 2020). These effects are comparable to those of the earlier laboratory-based ISR studies. However, this study did not include a long-term follow up or a control group, so the cause of the observed improvements in sleep outcomes and whether these are maintained in the long-term are unclear. Nonetheless, it appears that Sleep On Cue-administered ISR can be used in the home environment with good adherence and without

experimenter supervision.

While Sleep On Cue can effectively administer the ISR protocol, there are issues pertaining to the user experience of the app. The use of tone stimuli is problematic because the stimulus should be maintained at the lowest perceptible intensity to minimise disruption, whilst remaining perceptible to the individual so they can initiate the required behavioural response. Maintaining this intensity using a tone stimulus emitted via earphones is difficult, as they may (and often did during testing) become dislodged overnight (Mair et al., 2020). Perhaps most problematic is the requirement that the individual holds their smartphone whilst undergoing the ISR procedure. The individual is required to move their phone with a back-and-forth motion of the wrist in response to the tone stimuli. The effort and motor activity required to produce this response may contribute to unnecessarily prolonged wakefulness during the ISR protocol, reducing the efficacy of the treatment due to less time for rapid sleep onset opportunities. Many individuals also do not like having their smartphone in the bedroom environment while they are attempting sleep. Overcoming these issues may create a more user-friendly experience, potentially leading to greater treatment adherence and satisfaction.

THIM

In close collaboration with Re-Time Pty. Ltd., we designed the THIM wearable device: a device that promises to administer ISR in the home environment. The individual connects THIM to the accompanying smartphone application via Bluetooth to instruct the device to administer ISR. THIM then administers the ISR protocol independently of the smartphone, meaning that individuals do not need to have their smartphone in the bedroom environment. The individual places THIM on the index finger of their dominant hand (see Figure 1-1) and attempts to fall asleep in bed

starting at their typical bedtime. The device emits a low intensity, short duration vibration approximately every 30 seconds to which the individual responds by tapping their finger. If THIM detects a response to the vibration, the device infers that the individual is awake and when they fail to respond to two consecutive stimuli, THIM assumes they have fallen asleep. Once THIM detects sleep onset, the device emits a high intensity vibration to wake the individual and signal the end of the sleep onset trial. After a five-minute break, THIM emits another high intensity vibration to signal that the individual should return to bed and attempt another trial. The retraining session continues for a duration of time specified by the individual during the configuration. Once this specified duration elapses, THIM begins the final sleep onset trial but once it determines sleep onset, it will not emit the high intensity alarm vibration. Instead, the individual sleeps uninterrupted until the morning. In many ways, the THIM-administered ISR protocol is designed to align with the laboratory-based and the Sleep On Cue-administered ISR protocols, with the individual falling asleep and waking up shortly thereafter on multiple occasions during the one overnight retraining session.

Figure 1-1. THIM placed on a sleeping individual, printed with permission.

For THIM to effectively administer ISR, it is crucial that the patient is woken at

the appropriate time relative to the onset of N1 sleep. Disturbing the patient too early may cause them to not experience sleep onset. Waking the patient too late may result in the individual experiencing enough sleep to significantly reduce sleep pressure, decreasing the likelihood of falling asleep quickly on subsequent trials. This is supported by studies investigating the consequences of brief naps, which have consistently found that 5-10 minutes of sleep can significantly reduce objective and subjective sleepiness post-nap (Brooks & Lack, 2006; Hilditch, Centofanti, Dorrian, & Banks, 2016; Tietzel & Lack, 2002). Because the rapid nature of the sleep onsets is thought to be the most therapeutic element of ISR, longer SOLs would presumably reduce the efficacy of the treatment (Lack et al., 2019). To wake the individual at the appropriate time, it is therefore essential that THIM can accurately detect sleep onset. Like Sleep On Cue, THIM utilises the stimulus-response method of estimating sleep onset. The accuracy of this method and its practicality for administering ISR is desirable compared to other objective methods used by other wearable devices, as will be discussed in Chapter 2.

While similar devices are accurate for measuring sleep onset using the stimulus-response method, THIM differs from previously tested devices in potentially important ways. THIM uses vibratory stimuli while other devices use auditory stimuli (Mair, 1994; Ogilvie, Wilkinson, & Allison, 1989; Scott et al., 2018), thereby relying on different information processing pathways for perception with reductions in perceptibility potentially occurring at differing time points during the onset of sleep. THIM also requires an easy-to-exert finger twitch behavioural response while other devices require more onerous hand/arm movements (Kuderian, Ogilvie, McDonnell, & Simons, 1991; Mair, 1994; Scott et al., 2018), which may become less possible to exert sooner into the onset of sleep than finger twitches. Furthermore, algorithm

parameters such as the stimulus duration, stimulus intensity and the number of consecutively missed responses to stimuli to register sleep onset may all differ from previously tested devices. These aspects may mean that the accuracy of THIM for estimating sleep onset differs from similar devices despite utilising the same underlying rationale that the cessation of behavioural responsiveness to external stimuli indicates sleep onset. The first aim of this dissertation was to refine and test the accuracy of THIM's sleep onset detection algorithm by comparing its estimations of sleep onset to the gold standard of objective sleep measurement, PSG. This aim is achieved in Chapters 3 and 4.

Sleep Tracking

THIM is entering a marketplace flooded with consumer sleep products. Health conscious and technology-savvy individuals regularly quantify and monitor their health and wellbeing, including their sleep (Robbins, Krebs, Rapoport, Jean-Louis, & Duncan, 2018). Individuals collect sleep data such as TST and sleep efficiency, which they can monitor for long-term trends. The Quantified Self Movement is a term coined to describe the growing trend of individuals using technology to quantify and monitor their health and wellbeing (Swan, 2012). This movement is a major contributor to the rapid growth of the CST industry, with one popular CST brand earning $US248 million in revenue in the first quarter of 2018 (Sawh, 2018). However, sales of these devices have slowed in recent years and long-term use of these devices are dwindling (Statista, 2020). One potential reason for this is that individuals, while initially intrigued by their sleep data, do not see the purpose of tracking sleep in the long-term, as merely tracking sleep does not lead to improvements in sleep (Russo & Bianchi, 2017).

During the treatment of insomnia, it is advantageous to measure sleep to

monitor the patient's response to treatment. In clinical practice, patients may complete sleep diaries to subjectively monitor their sleep and inform treatment. Over the long-term, individuals may struggle to consistently maintain a sleep diary. The use of CSTs to track sleep in the long-term may have greater adherence than sleep diaries. Sleep tracker data could also be automatically integrated into online treatment programs, without the need for user input, leading to greater personalisation of the treatment program and potentially greater therapeutic outcomes. There are many advantages and potential uses for CSTs in the treatment and management of insomnia and many other sleep disorders (Watson, Lawlor, & Raymann, 2019), with the main limitation being the validation of the current technology (Khosla et al., 2019). Therefore, a worthwhile goal for sleep medicine is the validation and incorporation of objective sleep measurements into online insomnia treatment programs.

Some online insomnia treatment programs incorporate sleep data to tailor the treatment instructions, including Sleepio which utilises subjective sleep diary data and objective data collected from a popular sleep tracker. However, for sleep trackers to be useful for this purpose, the device must provide accurate sleep data. The accuracy of the vast majority of sleep trackers is unknown as many companies do not publish their validation research, if they have indeed tested the product. Nonetheless, recent findings suggest that actigraphy may be useful for monitoring sleep for individuals with insomnia (Hamill et al., 2020; Kahawage, Jumabhoy, Hamill, de Zambotti, & Drummond, 2019).

The THIM device has many functions, not just the administration of ISR. THIM has the capacity to passively track sleep using the well-established actigraphy method, facilitate the optimal power nap, and emit a smart alarm to wake individuals

during light stages of sleep in the morning. While all functions require refinement and validation to ensure that the device is executing them appropriately, it is beyond the scope of this dissertation to investigate the power napping and smart alarm functions of the THIM device. However, the sleep tracking function will be investigated in this dissertation. This was to further the long-term goal of incorporating the sleep tracking function into a comprehensive mobile-based treatment of insomnia that utilises THIM-administered ISR. Therefore, the second and final aim of this dissertation is to develop and assess the accuracy of the THIM sleep tracking algorithm compared to PSG. This aim is achieved in Chapters 5 and 6.

Dissertation Aims

The aims of this dissertation were to develop and test the accuracy of the THIM wearable device 1) for estimating sleep onset, and 2) for estimating sleep and wakefulness during the nocturnal sleep period. Both the ISR and sleep tracking functions of the THIM device are being developed and assessed in this dissertation because they are necessary components to achieving the long-term goal of the effective and practical online treatment of insomnia.

Chapter 2 summarised the ability of current wearable technology to measure sleep onset. The review systematically identified studies which examined the accuracy of practical sleep wearable devices for estimating sleep onset. The aim was to identify the most suitable objective sleep measurement method for administering ISR in the home environment.

Chapter 3 discussed the development and accuracy of the THIM stimulus-response algorithm for estimating SOL. The aim was to refine the algorithm to ensure that it can accurately estimate sleep onset compared to PSG.

The aim of Chapter 4 was to further investigate the correspondence between

PSG and THIM estimations of SOL through quantitative electroencephalography (qEEG) analysis.

Chapter 5 shifts focus by describing the development of the THIM sleep tracking function. The main aim was to investigate whether THIM can accurately track sleep using actigraphy compared to PSG.

Relatedly, Chapter 6 aimed to assess the consistency in the accuracy of THIM for monitoring sleep over multiple nights compared to PSG.

Chapter 7 discussed the findings of this dissertation, their implications for sleep research and healthcare practice, and directions for future research into ISR and the THIM device.

Methodology Justification

The methodology of this research is discussed throughout the following chapters. However, the reasoning behind three methodological decisions are discussed here as they 1) relate to all studies described in this dissertation, and 2) are particularly important for the interpretation of the aims and findings of this dissertation.

Terminology: Accuracy versus Validity

This dissertation discusses the accuracy of the THIM device. The theoretical concept of accuracy refers to the degree to which the measure, THIM, aligns with the gold standard measure, PSG (Streiner & Norman, 2006). This differs from validity, which refers to the extent to which the measure represents what it is supposed to measure for a specific purpose (Messick, 1995). The goal of the validation of THIM is to assess whether the device can successfully administer ISR. Assessing the accuracy of the device is the first step in reaching this goal by confirming that THIM can substitute for PSG to monitor nocturnal sleep and wakefulness. Further testing

will be required to confirm the validity of the device for administering ISR, as is discussed in Chapter 7. Such testing would need to determine whether patients can adhere and follow instructions for THIM-administered ISR in the uncontrolled home environment and whether this would lead to improvements in insomnia symptoms. This dissertation presents findings about the accuracy, not the validity, of the THIM device for estimating sleep onset and for monitoring sleep and wakefulness.

Polysomnography as the Gold Standard Measure of Sleep

To assess the accuracy of THIM, the device's estimations of sleep and wake were compared to those derived from standardised PSG scoring criteria (American Academy of Sleep Medicine, 2018), except for Chapter 4 which involved a qEEG analysis. Despite limitations of this approach (Tryon, 2004), PSG is the current gold-standard method of objectively measuring sleep and is the recommended comparison to assess the validity of wearable devices (Depner et al., 2019). An alternative gold-standard sleep measure is self-reported sleep (i.e. sleep diaries). Whilst understanding the correspondence between THIM and self-reported sleep is important for some applications, including for the treatment of insomnia, this was not our focus. The goal was to develop a device that could substitute PSG and assess objective sleep, not subjective sleep, for the specific purposes of ISR and monitoring of objective sleep in the home environment. For these purposes, PSG scoring criteria is the gold-standard measure to which THIM should be compared to ascertain whether the device can suitably perform these purposes.

Good and Poor Sleepers Sample

All studies described in this dissertation tested THIM with samples of good and poor sleepers (subthreshold clinical insomnia). The decision was made to recruit this sample rather than individuals with insomnia. Whilst the long-term goal is to use

THIM in clinical practice for the treatment of insomnia, THIM is currently available to the general population. Thus, the device's ISR and sleep tracking functions will likely be used largely by individuals without clinical insomnia, at least in the short-term before its potential implementation in healthcare practice. As such, we decided to assess the accuracy of THIM with a sample of individuals who have varied sleep quality that is more representative of the sleep quality found in the general population (Adams et al., 2017). This was to ensure that THIM could conduct these functions appropriately with the population that are likely to implement THIM first. To achieve the long-term goal of using THIM specifically for the treatment of insomnia, the device will need to be tested with an insomnia sample. Due to resource restrictions, this was not possible within the scope of this dissertation. Nonetheless, the relevance of the findings for individuals with insomnia is discussed throughout this dissertation where appropriate.

Chapter 2: A Systematic Review of the Accuracy of Sleep Wearable Devices for Estimating Sleep Onset

Hannah Scott[1], Leon Lack[1,2], Nicole Lovato[2]

[1] College of Education, Psychology and Social Work, Flinders University.

[2] College of Medicine and Public Health, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, Flinders University.

Author contributions: HS contributed to study design, screening and data extraction, analysis and interpretation of the data, and manuscript preparation. LL and NL contributed to conception and study design, interpretation of the data and drafting of the manuscript.

Conflict of interests: Leon Lack is a shareholder in Re-Time Pty Ltd. The other authors have no conflicts of interest to disclose.

Citation: Scott, H., Lack, L., & Lovato, N. (2020). A systematic review of the accuracy of wearable devices for estimating sleep onset. *Sleep Medicine Reviews, 49*(101227), 1-12. doi: 10.1016/j.smrv.2019.101227.

Abstract

The accurate estimation of sleep onset is required for many purposes, including the administration of a behavioural treatment for insomnia called Intensive Sleep Retraining, facilitating power naps, and conducting objective daytime sleepiness tests. Specialised equipment and trained individuals are presently required to administer these applications in the laboratory: a costly and impractical procedure which limits their utility in practice. A wearable device could be used to administer these applications outside the laboratory, increasing accessibility. This systematic review aimed to identify practical wearable devices that accurately estimate sleep onset. The search strategy identified seventy-one articles which compared estimations of sleep onset latency from wearable devices against polysomnography. Actigraphy devices produced average estimations of sleep onset latency that were often not significantly different from polysomnography, but there was large inter-individual variability depending on participant characteristics. As expected, electroencephalography-based devices produced more accurate and less variable estimates. Devices that measured behavioural aspects of sleep onset consistently overestimated polysomnography-determined sleep onset latency, but to a comparatively low degree. This sleep measurement method could be deployed in a simple wearable device to accurately estimate sleep onset and administer Intensive Sleep Retraining, power naps, and objective daytime sleepiness tests outside of the laboratory setting.

Keywords: sleep; sleep onset; insomnia; power naps; Intensive Sleep Retraining; wearable technology; consumer sleep technology; polysomnography; actigraphy; systematic review.

There are many reasons why researchers, clinicians and consumers want to accurately measure sleep in the home environment. In particular, some applications require the accurate estimation of sleep onset. One such application is to administer a novel behavioural treatment for insomnia called Intensive Sleep Retraining (ISR): a brief but effective treatment involving near-total sleep deprivation to retrain the patient to fall asleep more quickly (Harris et al., 2012; Harris et al., 2007). To undergo ISR, the patient is required to lie down in bed and attempt to fall asleep. After a brief period of light sleep in the order of 2-3 minutes according to polysomnography (PSG), the patient is awoken and given feedback about how long it took them to fall asleep. The patient is subsequently instructed to remain awake for a brief period, before attempting to fall asleep again on the next trial. Since brief episodes (< 3 minutes) of light sleep do not reduce homeostatic sleep drive (Tietzel & Lack, 2002), sleep deprivation is effectively maintained over the whole, traditionally 24-hour, retraining session. It is the deprivation of recuperative sleep combined with a high circadian drive for sleepiness during the early hours of the morning that cause patients to fall asleep more rapidly with each subsequent sleep onset trial. Consequently, patients who report average pre-treatment sleep onset latencies (SOLs) > 60 minutes are able to fall asleep in < 5 minutes on dozens of attempts during the retraining session. The series of rapid sleep onsets is thought to extinguish the conditioned cortical arousal response hypothesised to interfere with the attempt to initiate sleep in the home environment (Lack et al., 2017).

For ISR to be administered appropriately in the home environment, it is essential that sleep onset is measured accurately. As described, current ISR procedures requires the patient to undergo PSG in the sleep laboratory (Harris et al., 2012; Harris et al., 2007). The prompt waking of the patient shortly after sleep onset

is thought to be necessary for achieving the strong therapeutic effect (Lack et al., 2017). If the patient is disturbed before sleep is established, the patient will not have the experience of having attained sleep and that particular sleep 'onset' trial is unlikely to produce any therapeutic benefit. If the patient is woken too late and has experienced >5 minutes of sleep, then sleep pressure would be somewhat alleviated and they would be expected to take longer to fall asleep in subsequent trials (Brooks & Lack, 2006; Hayashi, Motoyoshi, & Hori, 2005). Furthermore, as this treatment is conducted over a finite period of time, delayed awakenings past the point of sleep onset would reduce the number of rapid sleep onset and retraining experiences during the treatment period, which could reduce treatment efficacy. Additionally, learning theory would suggest that longer SOLs may reinforce, or at least fail to diminish or reduce the intensity of, the conditioned insomnia response. Therefore, the successful administration of ISR in the home environment requires that the administration tool accurately estimates sleep onset to wake them at the appropriate time, ideally after 2-3 minutes of sleep, but after no more than five minutes of sleep.

Similarly, the accurate measurement of sleep onset is required to facilitate power naps. Brief naps have consistently been shown to reduce both subjective and objective sleepiness and improve daytime functioning, including fatigue, attention, and memory (Brooks & Lack, 2006; Hilditch, Dorrian, & Banks, 2017; Takahashi, Fukuda, & Arito, 1998). Very brief naps (less than five minutes sleep duration) do not lead to significant improvements in alertness, while longer naps (30 minutes or longer sleep duration) result in sleep inertia upon waking that reduces immediate daytime functioning delaying the beneficial effects of the nap(Lovato & Lack, 2010a). Naps of ~10 minutes sleep duration are optimal because they avoid the detrimental effects of sleep inertia but significantly improve daytime functioning immediately after

waking, which are sustained for up to three hours (Brooks & Lack, 2006; Hilditch et al., 2016). To nap for 10 minutes at present, people may set an alarm to wake them after a pre-determined duration of time. This requires the individual to estimate how long it will take them to fall asleep but if this estimate is not fulfilled, they will fail to obtain the optimal 10-minutes of sleep. A device that could accurately detect sleep onset and wake the individual after the appropriate duration of sleep could allow individuals to experience the strongest and most consistent improvements in daytime functioning from power naps in their home or work environment. This is an obvious benefit in today's culture with many experiencing daytime impairments from inadequate nocturnal sleep, and for the many shift workers wanting to counteract fatigue (Hilditch et al., 2017).

There are also potential diagnostic uses for a device that could accurately estimate sleep onset. The multiple sleep latency test (MSLT) and maintenance of wakefulness test (MWT) are objective daytime sleepiness tests which involve measuring SOL (Carskadon, 1986). These tests are used to objectively measure excessive daytime sleepiness used for the diagnosis of sleep disorders such as obstructive sleep apnea, idiopathic hypersomnolence, and narcolepsy. Currently, these tests are conducted in the sleep laboratory with specialised equipment and trained individuals required to administer the protocol. As a result, both tests are not widely used in healthcare practice to measure excessive daytime sleepiness, with clinicians opting in favour of subjective sleepiness measures (Johns, 1991) that have their limitations and correlate weakly with MSLTs (Benbadis et al., 1999; Sangal, Mitler, & Sangal, 1999). It is desirable to develop an alternative method of administering MSLTs and MWTs outside the laboratory setting that utilises fewer healthcare resources. For these tests to be administered appropriately, SOL needs

to be measured to a high degree of accuracy to make appropriate decisions for patient care. A device that could accurately estimate sleep onset would be useful for this purpose. Similarly, an accurate device could also be useful to reduce costs of laboratory-based research, such as for conducting constant routine protocols where SOL is used as a measure of sleep propensity (Gradisar & Lack, 2004).

These applications have been administered previously using PSG, which is considered to be the gold standard method of objective sleep measurement (Hirshkowitz, 2017). PSG measures the physiological changes that occur during the transition from wake to sleep, including brain waves (electroencephalography, EEG), eye movements (electrooculography, EOG) and muscle activity (electromyography, EMG). Since the early twentieth century, researchers have observed physiological activity around sleep onset using PSG to better understand sleep by distinguishing it from wakefulness (Kleitman, 1929; Loomis, Harvey, & Hobart, 1935). Sleep onset is currently understood to be a complex transitional process involving physiological, behavioural and psychological changes that begins during relaxed wakefulness and ends in undeniable and sustained sleep (Carskadon & Dement, 2017; Ogilvie, 2001). As the individual relaxes, alpha waves subside as low voltage, mixed frequency theta waves emerge sometimes accompanied by slow rolling eye movements and a decrease of EMG activity indicating the onset of N1 sleep. This is usually followed by signs of Non-Rapid Eye Movement Stage 2 sleep (N2): k-complexes and sleep spindles. These physiological changes during N1 and N2 sleep coincide with behavioural concomitants of sleep onset, including reduced responsiveness towards auditory and visual stimuli, characteristic visual imagery, and an increasing likelihood of the perception of having fallen asleep if awoken around N2-sleep onset (Guilleminault, Phillips, & Dement, 1975; Ogilvie et al., 1989; Yang, Han, Yang, Su,

& Lane, 2010). These changes occur gradually, and individuals may fluctuate between sleep and wakefulness before finally entering sustained sleep.

While sleep onset is a complex process, current standardised PSG scoring criteria defines sleep onset as a specific point along the continuum (American Academy of Sleep Medicine, 2018). PSG N1-sleep onset is scored when EEG alpha comprise less than 50% of the 30-second epoch, which is typically accompanied by reductions in muscle tone and slow eye movements (American Academy of Sleep Medicine, 2018; Carskadon & Dement, 2017). This point is difficult to score especially for people with little alpha waves, even for experienced sleep technicians (Rosenberg & Van Hout, 2013). Nonetheless, PSG is the most accepted method researchers and clinicians have at their disposal for measuring objective sleep.

There are many practical limitations to PSG which limit its usefulness, particularly for measuring sleep in the home environment. PSG requires specialised equipment and trained people to administer, which is time consuming and expensive. Even though ambulatory PSG devices can be used in the home environment, the equipment and consumables are expensive and not readily available for use in many situations, or for many individuals. Aside from cost, people having their sleep monitored via PSG are often inconvenienced by having to attend a sleep laboratory to be setup for ambulatory monitoring and experience discomfort whilst attempting to sleep, at least on the first night. Many of these limitations are exacerbated when attempting to monitor sleep over multiple nights. In response, alternative methods and devices have been developed to measure sleep, many of which can be used practically outside the laboratory.

Previous literature reviews have summarised the validity of devices for objectively measuring sleep (Evenson, Goto, & Furberg, 2015; Van den Water,

Holmes, & Hurley, 2011). Van den Water et al. (2011) systematically reviewed the validity of devices for measuring objective sleep, including research-grade and consumer wearable and non-wearable sleep devices. The authors concluded that actigraphy devices were the most appropriate device for measuring sleep outside of the laboratory setting because they are the most widely used and validated method of sleep measurement. However, Van den Water et al. (2011) noted that further validation could indicate that other devices are more accurate than actigraphy devices. Similarly, Evenson et al. (2015) systematically reviewed the validity and reliability of consumer wearable sleep devices, including the popular Fitbit and Jawbone devices. These devices tended to overestimate sleep and underestimate wakefulness across the sleep period: a common finding of actigraphy-based devices. Importantly, these reviews have not specifically focused on the measurement of sleep onset.

A review of the accuracy of wearable sleep devices for estimating sleep onset is warranted because the accurate measurement of this sleep parameter in particular is crucial, specifically for the purposes of administering ISR, power naps, and daytime sleepiness tests outside the laboratory setting. Furthermore, the consumer sleep technology (CST) space in particular is evolving rapidly, as is research into the accuracy of these devices (Bianchi, 2017). This warrants an updated review of the evidence regarding the accuracy of wearable devices, focusing specifically on sleep onset.

The present study has two aims:

1) Use a systematic approach to evaluate the evidence of the accuracy of wearable devices for measuring SOL in adults compared to PSG.

2) Identify existing wearable devices and objective sleep measurement

methods that could be used to administer ISR, power naps and daytime sleepiness diagnostic tests outside the laboratory setting.

Method

This review was conducted in accordance with the PRISMA (preferred reporting items for systematic reviews and meta-analyses) guidelines (Moher, Liberati, Tetzlaff, Altman, & The, 2009).

Search Strategies

The database search strategy incorporated relevant title/abstract key words and medical subject headings (MeSH) or equivalent subject headings under three broad categories: sleep onset ('sleep latency', 'sleep onset period', etc.), measurement ('validity', 'detect', etc.) and devices ('actigraphy', 'mobile app', 'tracker', etc.). See Appendix 1 for the specific keywords and subject headings used to search in each database. The initial search strategy was developed in PubMed and translated to the remaining databases. On 12th June 2018, searches were conducted in PubMed, Web of Science, SCOPUS, PsycINFO and CINAHL EBSCOhost databases. Search results were restricted to articles published in English.

Additional search strategies included screening the reference lists of included articles and consulting with collaborators to identify relevant articles not captured by the database search strategy. These additional strategies were completed by 23rd July 2018.

Study Selection

After de-duplication, articles were screened according to the selection criteria summarised in Table 2-1 by the primary author, HS. The selection criteria were designed to include articles that assessed objective sleep wearable devices and

compared them to the gold-standard method, PSG. Articles that assessed the accuracy of the device using the same sample used to develop the algorithms (i.e. not an independent validation sample) were excluded.

Table 2-1. *Study selection criteria.*

Article Criteria
Original, full text, peer-reviewed articles
Written in English
Sample Characteristics Criteria
Adult, human sample
Sample size ≥ 10
Tested Device Criteria
Tested a device that used an objective method purporting to measure sleep
Wearable device that is practical for use outside the laboratory
Does not require expert knowledge or an extensive setup procedure. For instance, placing EEG electrodes at measured points on the scalp.
Accuracy Criteria
Tested against PSG in an independent sample
Data is presented pertaining to sleep onset

EEG = electroencephalography, PSG = polysomnography

Data Extraction

Key data fields were identified for extraction by HS. These fields included relevant study information (year, setting, study location), sample characteristics (age, gender, sleep characteristics), device specifications (type of device tested, algorithm, PSG scoring criteria) and any data relating to SOL. The statistics reported in the articles varied widely and included correlation coefficients, intra-class correlations, mean discrepancies between PSG and the tested device, and Bland-Altman plot values. The relevant data was extracted and inserted into summary tables (Tables A2, A3 and A4 in the Appendix).

Results

The PRISMA flow diagram in Figure 2-1 outlines the selection of studies at each stage of screening. The database search strategy identified a total of 4,200 articles and an additional 10 articles were found through other strategies. After removing duplicates, 1,855 records remained. 1,706 records were excluded after preliminary screening of the titles and abstracts, and a further 78 records were excluded during the full text screening (see Figure 2-1 for reasons for exclusion). A total of 71 articles met the selection criteria for inclusion in this review.

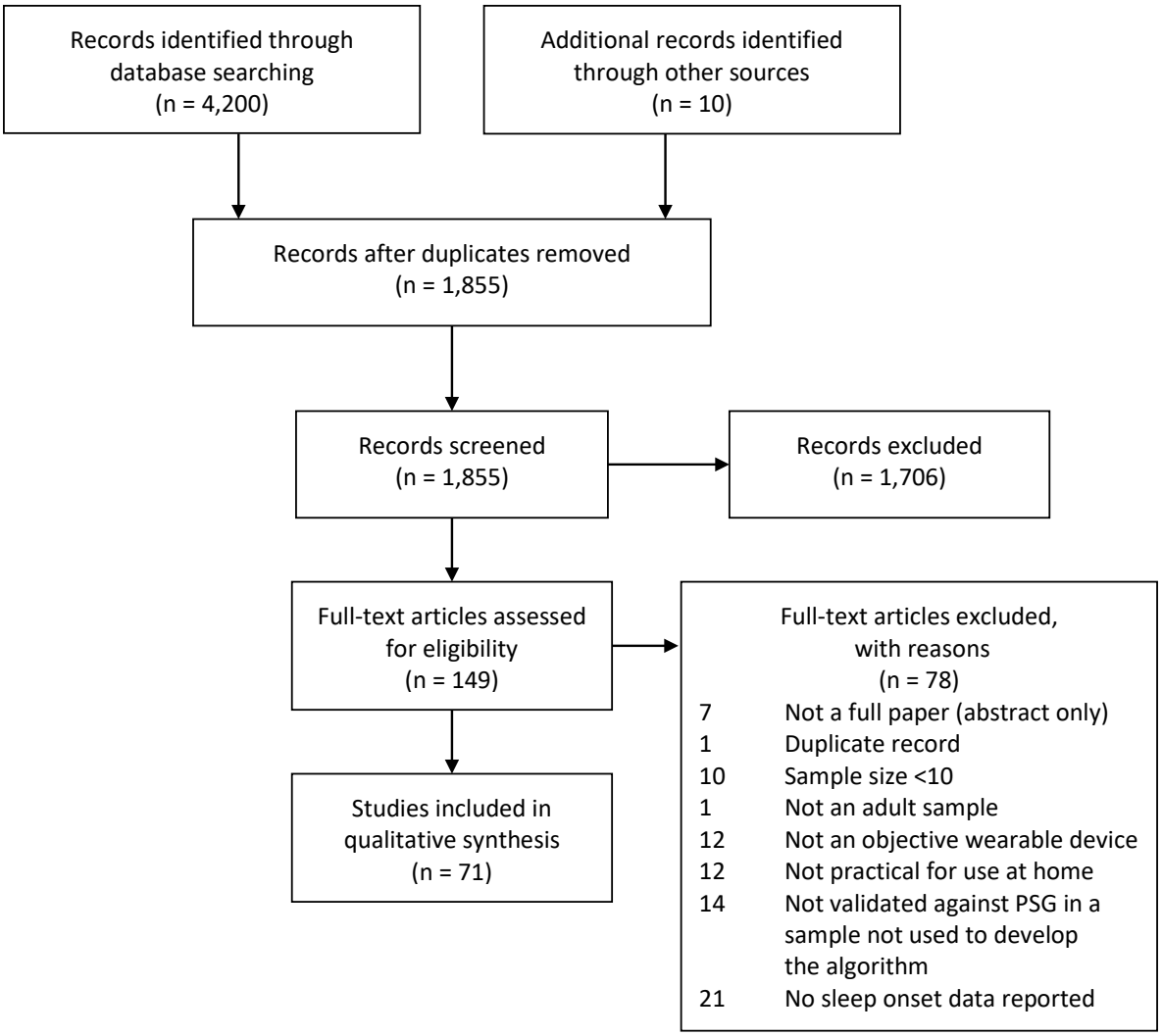


Figure 2-1. PRISMA flowchart detailing the study selection process.

Table A2 (see Appendix) presents detailed information relating to the sample, test setting and PSG specifications used in each study. Most studies were conducted

with young healthy individuals, those with sleep disorders or individuals with other significant physical or mental health conditions. Testing was typically conducted in the sleep laboratory with participants allowed their typical nocturnal sleep opportunity.

Methods used by the Wearable Devices

The wearable devices used a wide range of methods of estimating sleep onset, from relatively complicated physiology-based methods incorporating many signals to simple behaviour-based methods. The most common method of estimating sleep used by the tested devices was actigraphy, which comprised 75% of all devices tested in the included articles (Chae et al., 2009; Dunican et al., 2018; Rupp & Balkin, 2011). This method utilises in-built accelerometers to measure body movement to infer sleep and wakefulness (Sadeh & Acebo, 2002). Little/no movement is scored as sleep and a greater degree of movement is inferred as wakefulness. Devices also used physiology-based methods (15.9% of devices tested) including headbands that measured brain waves (Cellini, McDevitt, Ricker, Rowe, & Mednick, 2015; Kaplan, Wang, Loparo, Kelly, & Bootzin, 2014). Some devices incorporated both actigraphy and various physiological signals (6.8% of devices tested), such as EEG, heart-rate variability and eye movements (Edinger, Means, Stechuchak, & Olsen, 2004; Fonseca et al., 2017). Behavioural methods (2.3% of devices tested) included measuring behavioural responses to auditory stimuli and depressing a micro-switch (Hauri, 1999; Scott et al., 2018). Table A3 (see Appendix) provides information pertaining to the tested device specifications for each article in greater detail.

Accuracy of the Wearable Devices

Data pertaining to the accuracy of the tested devices is summarised in Table

A4 (see Appendix), grouped by the objective sleep measurement method used by the wearable device (actigraphy, actigraphy plus a physiological signal, physiology-based devices, and behaviour-based devices).

Actigraphy Devices

The accuracy of actigraphy devices is grouped by the characteristics of the sample, the device model and the algorithm used to derive sleep and wakefulness. Previous reviews have identified these factors as important influences on the accuracy of wearable devices for estimating sleep and wakefulness (Evenson et al., 2015; Van den Water et al., 2011). Figure 2-2 shows mean discrepancies between actigraphy device-derived SOL and PSG-determined SOL, as will be discussed in greater detail in the following sections.

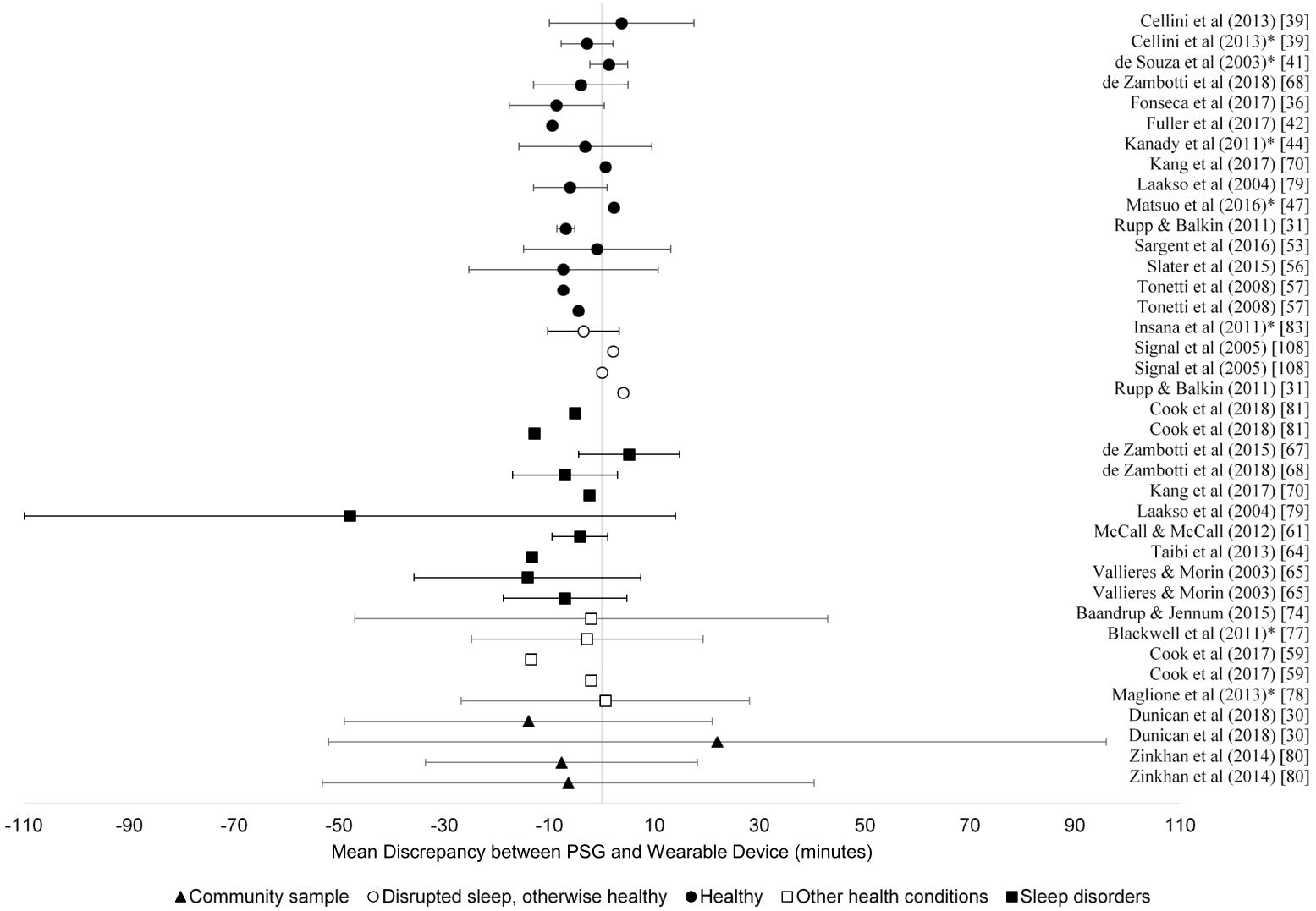


Figure 2-2. Graphical representation of the mean discrepancies between PSG-SOL and actigraphy-SOL.
The type of marker indicates sample characteristics. Negative values on the x axis indicate underestimations of actigraphy-derived SOL, positive values indicate overestimations of actigraphy-derived SOL compared to PSG. Error bars represent one standard deviation, where available. If a study tested the accuracy of multiple algorithms on the same dataset, only the most accurate algorithm was chosen for inclusion in this figure. * indicates studies that tested multiple algorithms.

Healthy participants

Studies conducted with healthy sleepers (Cellini, Buman, McDevitt, Ricker, & Mednick, 2013; Chakar et al., 2017; De Souza et al., 2003; Fonseca et al., 2017; Fuller, Juliff, Gore, Peiffer, & Halson, 2017; Gruwez, Libert, Ameye, & Bruyneel, 2017; Kanady, Drummond, & Mednick, 2011; Kosmadopoulos, Sargent, Darwent, Zhou, & Roach, 2014; Markwald, Bessman, Reini, & Drummond, 2016b; Matsuo et al., 2016; Nakazaki et al., 2014; O'Hare et al., 2015; Paquet, Kawinska, & Carrier, 2007; Pigeon et al., 2018; Reid & Dawson, 1999; Rupp & Balkin, 2011; Sargent, Lastella, Halson, & Roach, 2016; Scatena et al., 2012; Shambroom, Fabregas, & Johnstone, 2012; Slater et al., 2015; Tonetti, Pasquini, Fabbri, Belluzzi, & Natale, 2008) reported mean discrepancies between PSG and the actigraphy device-identified sleep onset ranging from an underestimation of 15.1 minutes to an overestimation of 23.2 minutes. Standard deviations ranged from 3.6 to 37.8 minutes suggesting that even amongst healthy populations, the variability in estimations from actigraphy devices can be substantial. Intra-class correlations ranged from -0.07 to 0.56, and Pearson correlations ranged from no correlation, r = .01, to strong positive correlations, r = 0.73.

Sleep disorders

Several studies examined the accuracy of actigraphy devices exclusively in samples with sleep disorders. Insomnia was the most common sleep disorder investigated in the included articles (Choi, Kang, Sung, & Joo, 2017; Cook, Prairie, & Plante, 2017; Lichstein et al., 2006; McCall & McCall, 2012; Mundt et al., 2016; Sivertsen et al., 2006; Taibi, Landis, & Vitiello, 2013; Vallieres & Morin, 2003). These studies generally showed that actigraphy devices underestimated SOL compared to PSG. Mean discrepancies ranged from -4.19 to -14.16 minutes (*SD* = 11.77-21.59), though few studies reported this statistic (Cook et al., 2017; McCall & McCall, 2012;

Taibi et al., 2013; Vallieres & Morin, 2003). One study reported a strong intra-class correlation of .70 (Choi et al., 2017), while Pearson correlations ranged from no correlation (r = .08) to a non-significant, weak correlation (r = .31). McCall and McCall (2012) tested the accuracy of an actigraphy device with individuals diagnosed with both Insomnia and Major Depressive Disorders. The authors observed a mean discrepancy between PSG and actigraphy of -4.19 minutes, though this was not a significant difference. There was a small, significant correlation between PSG and actigraphy-SOL, r = .31 (McCall & McCall, 2012).

Two studies were conducted with obstructive sleep apnea (OSA) patients (Chae et al., 2009; Dick et al., 2010b) and one study with those with sleep-disordered breathing more generally (Choi et al., 2017). Dick et al. (2010b) found that an actigraphy device significantly underestimated SOL compared to PSG, despite a strong positive correlation between the two measures, r = .89. For those with sleep-disordered breathing, Choi et al. (2017) found that actigraphy estimates of SOL did not significantly differ from PSG-SOL, but with no significant correlation between them.

Some studies investigated the accuracy of actigraphy devices with samples of both healthy participants and those with sleep disorders (de Zambotti, Claudatos, Inkelis, Colrain, & Baker, 2015a; de Zambotti, Goldstone, Claudatos, Colrain, & Baker, 2018; Hedner et al., 2004; Kang et al., 2017; Kuo et al., 2017; Sanchez-Ortuno, Edinger, Means, & Almirall, 2010; Wang et al., 2008). Neither Kang et al. (2017) nor Sanchez-Ortuno et al. (2010) found actigraphy devices to be more or less accurate for estimating SOL for the insomnia groups compared to the healthy groups. Similarly, Wang et al. (2008) found no significant differences between PSG-SOL and actigraphy-SOL for either the healthy or OSA groups. However, Hedner et

al. (2004) found that actigraphy significantly overestimated SOL compared to PSG in groups with mild and moderate OSA. Although, no differences were found in the healthy or severe OSA groups. Interestingly, de Zambotti et al. (2018) found that actigraphy significantly underestimated SOL compared to PSG in healthy participants, but no significant difference in those with Periodic limb movements of sleep (PLMS). It is important to note that only nine participants had PLMS in this study, so replication is required with a larger sample to make confident conclusions. Synthesising these findings, the available evidence suggests that the accuracy of actigraphy for estimating SOL does not differ between healthy participants and those with sleep disorders.

Other health conditions

Studies that included those with mental health conditions showed generally little difference between PSG and actigraphy-determined SOL (Baandrup & Jennum, 2015; Cole, Kripke, Gruen, Mullaney, & Gillin, 1992; Cook et al., 2017; Kaplan et al., 2014). Kaplan et al. (2014) and Baandrup and Jennum (2015) found no significant differences between PSG-SOL and actigraphy-SOL for people with bipolar disorder and bipolar/schizophrenia disorder, respectively. Cook et al. (2017) found a research-grade actigraphy device significantly underestimated SOL in those with Major Depressive Disorders, but a consumer actigraphy device did not differ from PSG when using the default scoring algorithm.

The accuracy of actigraphy devices for estimating SOL has also been tested in those with injuries, diseases or disabilities which impact motor movement (Alsaadi et al., 2014; Blackwell, Ancoli-Israel, Redline, & Stone, 2011; Laakso, Leinonen, Lindblom, Joutsiniemi, & Kaski, 2004; Maglione et al., 2013). When the most accurate algorithm scoring parameters were set, actigraphy-SOL did not significantly

differ from PSG-SOL for men with osteoporotic fractures (Blackwell et al., 2011) or individuals with Parkinson's disease (Maglione et al., 2013). Whereas, Laakso et al. (2004) found an actigraphy device to greatly underestimate SOL compared to PSG for individuals with sleep disorders comorbid with motor disabilities (M = -152, SD = 194): a much greater underestimation than found from those with sleep disorders (M = -48, SD = 62) or able-bodied participants (M = -6, SD = 7). Correlations between PSG and the actigraphy device-derived SOL were strong for the able-bodied, r = .82, and sleep-disordered groups, r = .73, but weak and not significant for the individuals with sleep disorders comorbid with motor disabilities, r = .17.

Community samples

Two studies have tested actigraphy with representative community samples. Dunican et al. (2018) included 50 middle-aged adults of varying health whom were participating in a larger longitudinal study, whereas Zinkhan et al. (2014) recruited a convenience sample of residents from four German cities aged 18-75 years old (N = 100). Both studies found reasonably small mean discrepancies between actigraphy devices and PSG (M *range* = 6.4 - 22). However, variability was considerably larger than found in other studies with actigraphy devices, with standard deviations ranging from 18.2 to 74. No potential explanations were found for this high degree of variability in either study.

In summary, the findings suggest that actigraphy devices underestimate SOL, typically within 10 minutes of PSG-SOL. However, variability in the discrepancy with PSG was high, as indicated by large standard deviations. The accuracy of actigraphy devices can differ depending on the characteristics of the sample. As shown in Figure 2-2, there is the greatest variability in studies with community samples, with the least variability typically observed with healthy, good sleepers.

Actigraphy algorithms

Whilst actigraphy devices have the same underlying rationale that motion equates to sleep/wake state, there may be differences in accuracy between actigraphy models. Thirteen studies used at least two different actigraphy devices simultaneously (Alsaadi et al., 2014; Cellini et al., 2013; Cook et al., 2017; Cook, Prairie, & Plante, 2018; Dunican et al., 2018; Gruwez et al., 2017; Kang et al., 2017; Kosmadopoulos et al., 2014; Matsuo et al., 2016; Pigeon et al., 2018; Rupp & Balkin, 2011; Tonetti et al., 2008; Zinkhan et al., 2014). Most of these studies have tested Philips Actiwatch devices (10/13 studies), with mean discrepancies ranging from a 2.3-minute overestimation to a 14-minute underestimation. In recent years, Fitbit devices using the optimal algorithm settings have shown promising validity for estimating SOL compared to the accuracy of Actiwatch devices (Cook et al., 2017; Kang et al., 2017). Not all actigraphy devices are comparable, despite similarities in hardware and using the same underlying rationale of measuring sleep.

The scoring parameters and algorithms used to determine sleep and wakefulness may produce the discrepancy in accuracy between different actigraphy models. De Souza et al. (2003) compared the accuracy of two commonly-used algorithms, the Cole-Kripke (Cole et al., 1992) and the Sadeh et al. (Sadeh, Sharkey, & Carskadon, 1994) algorithms. The authors observed that the algorithms were comparable in accuracy for estimating SOL with healthy individuals. Paquet et al. (2007) also tested four different actigraphy algorithms, finding that the accuracy of the algorithms did not vary considerably between the experimental conditions of nocturnal sleep, daytime recovery sleep, and daytime recovery sleep after consuming caffeine.

However, the accuracy of actigraphy devices is impacted by the scoring parameter settings. Blackwell et al. (2011) tested the accuracy of three different

modes for accumulating the accelerometer data: zero crossing mode (the number of instances that the accelerometer waveform crosses zero), time above threshold (duration of time that the waveform is above a certain threshold), and proportional integration mode (the area under the curve of the waveform). The zero crossing mode showed a higher mean discrepancy for estimating SOL compared to PSG (*M* = 17.56, *SD* = 51.37) than the proportional integration mode (*M* = -2.77, *SD* = 22.03) and time above threshold mode (*M* = -2.43, *SD* = 27.82).

Similarly, Matsuo et al. (2016) showed that the threshold setting for the amount of movement required to score an epoch as wakefulness impacted the accuracy of a device. SOL estimates ranged from 15.00 minutes (*SD* = 3.67) for a low threshold to 8.50 minutes (*SD* = 3.09) for a high threshold. Chae et al. (2009) tested the accuracy of a research-grade device in people either with OSA or a combination of OSA and PLMS. The authors varied the number of 'immobile minutes' required to estimate actigraphy-derived SOL: a common scoring parameter used in the algorithms of many research-grade actigraphy devices. This change in algorithm setting greatly impacted the accuracy of the actigraphy device with SOL estimates ranging from 3.59 (*SD* = 4.05) to 44.61 minutes (*SD* = 66.67). Five immobile minutes was optimal for estimating SOL, while Insana, Glowacki, and Montgomery-Downs (2011) and Maglione et al. (2013) found 10 immobile minutes to produce more accurate estimates of SOL for first-time parents experiencing disrupted sleep and for those with Parkinson's disease, respectively. This suggests that not only does the accuracy of actigraphy devices depend upon the scoring parameters, but also that the optimal setting depends on the characteristics of the individual.

Studies that have tested various algorithms and scoring parameters have

mostly utilised research-grade actigraphy devices. Few studies have conducted similar testing with consumer actigraphy devices as the algorithms are often proprietary and unavailable to researchers. Cook et al. (2017) found that Fitbit Flex devices using the 'normal' algorithm settings as set through the accompanying online software were more accurate for estimating SOL than the 'sensitive' algorithm setting for individuals with Major Depressive Disorder. Similarly, Kang et al. (2017) found higher intra-class correlations for the 'normal' settings compared to the 'sensitive' settings for both good sleepers and individuals with insomnia.

Actigraphy Devices: The Next Generation

The latest actigraphy models incorporate additional physiological signals into the sleep scoring algorithms to improve the accuracy of the devices. An increasingly common addition to new consumer actigraphy devices is photoplethysmography (PPG). This method involves emitting light onto the skin and measuring changes in light absorption to measure heart rate and heart rate variability. Fonseca et al. (2017) compared the accuracy of an actigraphy device with PPG and a device without PPG for estimating SOL in healthy individuals. The mean discrepancy and standard deviation between the device and PSG for the actigraphy plus PPG device (M = −7.48, SD = 6.64) was slightly lower than the actigraphy-only device (M = −8.59, SD = 9.05). This suggests that incorporating this physiological signal may improve the accuracy of actigraphy devices.

Hedner et al. (2011) investigated the accuracy of a similar wrist-worn device called the Watch-PAT, which is used to detect sleep-disordered breathing. The device measures various signals including peripheral arterial tone (PAT), actigraphy, pulse rate and oximetry. This device significantly underestimated SOL by approximately seven minutes compared to PSG, yet there was a reasonably high

degree of correspondence between the two measures (r = .57).

Other actigraphy devices have combined head movements with EOG to estimate sleep. Ajilore, Stickgold, Rittenhouse, and Hobson (1995) found one such device produced estimations of SOL that were not significantly different from PSG, whereas Edinger et al. (2004) found a similar device gave significantly higher estimations of SOL compared to PSG. The divergent results could be explained by differences between the wearable devices and their algorithms: Ajilore et al. (1995) conducted testing with the Nightcap device, while Edinger et al. (2004) used the REMview device.

Some devices have incorporated EEG signals with actigraphy (Fietze et al., 2015; Finan et al., 2016). Finan et al. (2016) tested a device that combined two EEG channels (AF7, AF8) and head movement. The authors observed that the automatic scoring algorithm and manual scoring performed by two experts on data from this device were similar in accuracy for estimating SOL. However, both the automatic and the manual scoring methods significantly underestimated SOL compared to PSG (automatic: M = 5.00, SD = 5.52; manual: M = 4.82, SD = 6.18). Fietze et al. (2015) observed that combining one EEG channel (F4-M1) with actigraphy resulted in a high degree of correspondence with PSG (r = .98) and a small mean bias (M = 3, SD = 6). The addition of chin EMG and EOG channels did not improve the accuracy of the device for estimating SOL. It is important to note that the wearable device data was manually scored by a qualified scorer using standardised PSG criteria (Rechtschaffen & Kales, 1968). Whether a similar degree of accuracy could be obtained when using an automatic scoring algorithm with this particular device is yet to be shown. Nonetheless, the standard deviation for the mean discrepancy in estimations was considerably smaller in Fietze et al. (2015) than found with

actigraphy devices, suggesting that incorporating EEG signals may reduce the variability in the accuracy of the estimations made by actigraphy-based devices.

Physiology-based Devices

Some wearable devices tested in the included studies have relied solely on physiological signals to estimate sleep. These devices typically measure EEG, such as the discontinued Zeo wireless system (Cellini et al., 2015; Griessenberger, Heib, Kunz, Hoedlmoser, & Schabus, 2013; Kosmadopoulos et al., 2014; Markwald et al., 2016b; Shambroom et al., 2012; Tonetti et al., 2013), the Zmachine (Kaplan et al., 2014), and the Sleep Profiler (Lucey et al., 2016). These devices are relatively accurate for estimating SOL. Studies have reported a high degree of correspondence between these devices and PSG, with intra-class correlations ranging from .42 to .67. While these devices have significantly underestimated SOL in some studies (Cellini et al., 2015; Kosmadopoulos et al., 2014; Markwald et al., 2016b), other studies have reported no significant difference between the two measures (Myllymaa et al., 2016; Shambroom et al., 2012; Tonetti et al., 2013).

More complex devices have incorporated multiple physiological signals. For instance, Senny et al. (2012) tested a device that measured mandible movements, oxygen saturation and nasal airflow. The device significantly overestimated SOL compared to PSG across all sleep disorder groups, with mean discrepancies ranging from 29.7 to 36.1 minutes. Standard deviations for each sleep disorder group were large, ranging from 59.5 to 88.7. This indicates large variation in the accuracy of this device across sleep pathologies. Contrastingly, White, Gibb, Wall, and Westbrook (1995) tested the device with the most complex setup included in this review: the NightWatch system. This device measures EOG, leg movement, oxygen saturation, nasal airflow, chest and abdominal wall motion, body position, movement and heart

rate. The incorporation of many signals resulted in similar estimations of SOL (*M* = 14, *SE* = 2) compared to PSG (*M* = 15, *SE* = 3), and a medium degree of correspondence, *r* = .54 [.22, .76]. Nonetheless, the incorporation of many physiological signals may not always result in greater accuracy than other devices which rely on one signal.

Behavioural Devices

Two studies tested devices which utilise behavioural methods of estimating SOL. Hauri (1999) tested a device similar to a stop-watch, whereby the participant is required to continuously depress a switch on the device to keep the counter running. When the participant relaxes their finger to the point where the switch is no longer depressed, the counter stops, displaying SOL. Hauri (1999) observed that despite a reasonable degree of correspondence between the device and N1-SOL (*r* = .60), this device significantly overestimated N1-SOL by 15.1 minutes. The cessation of muscle tension aligned closely with the onset of sustained PSG sleep, defined as 10 minutes of uninterrupted sleep (mean discrepancy of -1.8 minutes), *r* = .98.

We have previously tested the accuracy of a smartphone application for estimating SOL (Scott et al., 2018). This application measured behavioural responses to tone stimuli to estimate sleep onset, with the cessation of responses to the stimuli indicating that the participant had fallen asleep. Similar to Hauri (1999), the lowest mean discrepancy (*M* = 0.81, *SD* = 1.96) and highest degree of correspondence ($r_{(s)}$ = .92) between the smartphone application and PSG was observed for the onset of PSG-N2 sleep. The relatively small standard deviation for the mean discrepancies indicates that there was little variability in the accuracy of the smartphone application across the sample of healthy participants. However, the smartphone application was still accurate at estimating N1-SOL with a mean

discrepancy of 3.17 minutes (*SD* = 3.04) and high degree of correspondence between the two measures, $r_{(s)}$ = .79.

Comparison across Sleep Measurement Methods

Figure 2-3 is a chart of the reported mean discrepancies between the tested device and PSG across all types of sleep measurement devices reviewed in this article. As previously discussed, actigraphy device estimations of SOL are often not significantly different from PSG, but the standard deviations indicate large inter-individual variability in the discrepancies with PSG, the extent of which depends on the characteristics of the individual. Actigraphy devices that incorporated physiological signals may improve the variability of SOL estimates. EEG devices produced less variable estimates, particularly for those with sleep disorders. Unlike other devices included in this article, behavioural devices consistently overestimated N1-SOL, but to a comparatively small degree and with low variability in the discrepancy of their estimates compared to PSG.

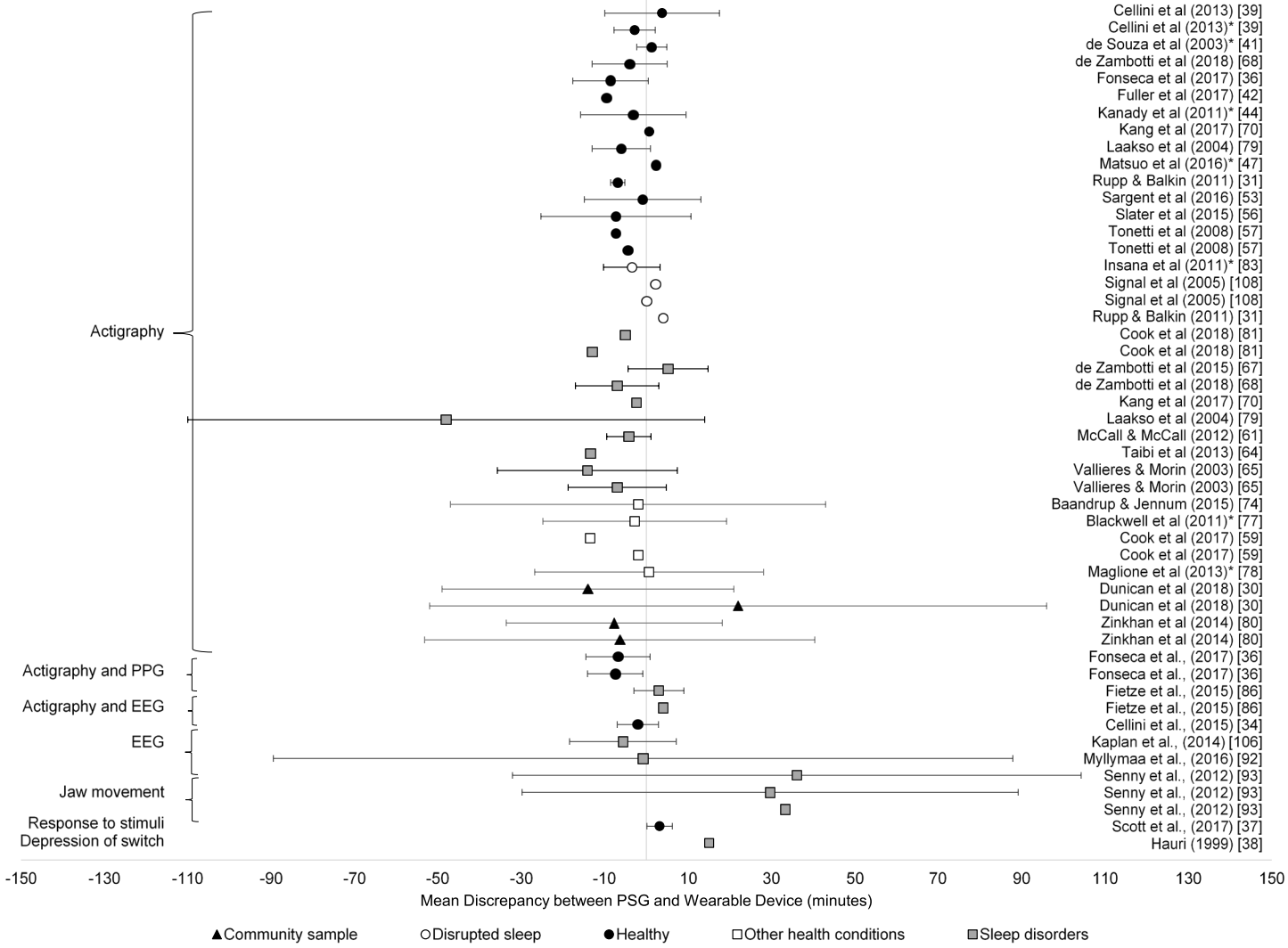


Figure 2-3. Representation of the mean discrepancies between PSG and each device, separated by type of sleep measurement device. The type of marker indicates sample characteristics. Negative values on the x axis indicate underestimations of device-derived SOL, positive values indicate overestimations of device-derived SOL compared to PSG. Error bars represent one standard deviation, where available. If a study tested the accuracy of multiple algorithms on the same dataset, only the most accurate algorithm was chosen for inclusion in this figure. * indicates studies that tested multiple algorithms.

Discussion

This review summarised the literature about the accuracy of practical wearable devices estimating SOL. Several devices were identified that measure physiological and behavioural processes to estimate sleep. Actigraphy-based devices were the most common type of device and the included articles indicated that these devices err on the side of underestimating SOL. This is not surprising considering that these devices base SOL estimation on a reduction in movement, and these reductions tend to occur before the onset of sleep (Pollak, Tryon, Nagaraja, & Dzwonczyk, 2001). However, this rationale assumes that a reduction in movement before sleep onset is systematic and occurs similarly across individuals (Tryon, 2004). If this were true, then actigraphy algorithms could account for this systematic bias to provide more accurate estimations of sleep onset. The findings of this review suggest that the reduction in movement is not systematic across the population as variability in accuracy between individuals was high, particularly for individuals with sleep-disorders or other health conditions. This was further highlighted in the findings of studies conducted with community samples (Dunican et al., 2018; Zinkhan et al., 2014). The large variability in discrepancy with PSG indicates an element of unpredictability in the actigraphy data that has not yet been explained. Until individual differences can be accounted for in actigraphy algorithms, the accuracy of actigraphy devices and their algorithms will vary considerably across individuals.

Presumably in an effort to reduce this variability and measure sleep more accurately, some actigraphy devices have incorporated additional physiological signals. The rationale behind this decision is that incorporating additional signals, which are more sensitive to changes in sleep depth, will result in more accurate

estimations of sleep and wakefulness. Some devices have combined actigraphy with EEG, PAT or PPG: techniques that have existed for decades yet have not been practical to administer until the more recent developments of small, simple sensors. Evidence that is publicly available to support the accuracy of these devices is growing and appears to be promising (Fonseca et al., 2017). Staying at the fore-front of consumer sleep product development is a goal for sleep medicine, which can be achieved by working with companies in the development of this technology, and importantly, by disseminating evidence of the accuracy of these devices to the awaiting populations of researchers, clinicians and tech-savvy consumers.

Some devices identified in this review are essentially simplified, portable PSG devices, and as such, estimation of SOL from these devices closely aligned with PSG-SOL (Cellini et al., 2015; Kaplan et al., 2014; Tonetti et al., 2013). Importantly, variability in the accuracy of these devices is low between individuals. However, not all physiology-based devices produced accurate estimates of SOL (Senny et al., 2012), and these devices and their associated consumables are considerably more expensive than other wearables devices. Nonetheless, the limited number of research studies included in this review indicate that physiology-based devices may accurately estimate sleep onset, but more research is required.

Some devices rely on behavioural indices of sleep onset. These methods are not new. In fact, researchers used these methods as the basis for developing the scoring criteria for PSG (Dement & Kleitman, 1957; Loomis et al., 1935). As a consequence, studies have consistently found a high degree of correspondence between these behavioural devices and PSG-SOL, with small discrepancies in the order of 2-3 minutes (Connelly, 2004; Mair, 1994; Scott et al., 2018). Behavioural devices consistently overestimate SOL because lowered awareness and ability to

perform simple motor actions, either by depressing a switch or responding to a stimulus, does not cease until late-N1 or early N2 sleep (Ogilvie & Wilkinson, 1988; Ogilvie et al., 1989). Therefore, these behavioural methods tend to slightly overestimate N1-sleep onset, but inter-individual variability is low.

Actigraphy devices have the most empirical evidence publicly available across diverse populations, however, these devices are not suitable for the administration of ISR, power napping, and daytime sleepiness tests. To administer these applications, sleep onset needs to be determined shortly after it occurs so that the individual can be woken at the appropriate moment. This is particularly important for the administration of ISR, because the treatment requires that the individual maintains high homeostatic sleep drive (Lack et al., 2017). Since a sleep duration of 5-10 minutes can reduce sleep pressure (Brooks & Lack, 2006), sleep onset must be detected and the individual woken after <5-10 minutes of sleep during ISR. Similarly, an additional 5-10 minutes of sleep during a power nap can reduce the benefits to daytime functioning immediately post-waking (Brooks & Lack, 2006; Hayashi et al., 2005), or increase their SOL on subsequent tests during MSLTs, potentially impacting diagnosis. Actigraphy devices are therefore not suitable because they typically do not achieve this required degree of accuracy across all individuals, particularly for individuals with sleep disorders or serious health conditions: the populations that may benefit most from the applications dependent on accurate estimation of sleep onset. Furthermore, actigraphy devices typically score sleep onset as the first epoch of a sustained period of immobility (often 5 or 10 immobile minutes) and therefore cannot detect sleep onset immediately but require the elapsed confirmatory immobility period (Chae et al., 2009).

Similarly, EEG-based devices using manual scoring or automatic scoring

algorithms rarely produce sleep data in real-time, as this largely occurs retrospectively. Furthermore, these devices are expensive and relatively difficult to operate, limiting their usefulness. Whilst these methods and devices may be appropriate for passively tracking sleep for some purposes, they are not appropriate methods for administering ISR, power naps and daytime sleepiness tests.

Behavioural devices could be appropriate for these purposes outside the laboratory setting. Of particular advantage is that these devices provide an almost instantaneous detection of sleep onset. Hauri (1999) tested a device that displayed SOL on the display of the device immediately after sleep onset occurred. Similarly, the smartphone application tested by Scott et al. (2018) required two consecutive missed responses to stimuli to occur before determining sleep onset. As stimuli occurred approximately 30 seconds apart, sleep onset was scored quickly after it occurred. These devices are also relatively inexpensive and simple to operate. The rapid, accurate detection of sleep onset means that these simple behavioural devices may be suitable for the administration of ISR, power naps, and daytime sleepiness tests.

One potential challenge is that these devices may alter SOL: the presence of the stimulus and the required movement response may promote prolonged wakefulness. If the stimuli can be calibrated to a just perceptible level, and the degree of movement required as a response can be reduced to a minimal level, then a device may be able to overcome this limitation. More research is required to resolve this challenge and to make strong conclusions about the accuracy of these devices.

As with the other devices discussed, these behavioural devices also tend to overestimate sleep onset. This is more desirable for the administration of ISR, power

naps and daytime sleepiness tests than underestimating sleep onset. If an individual experiences 2-3 minutes of light sleep before being woken up during ISR, then sleep pressure is unlikely to significantly reduce. Similarly, an additional 2-3 minutes of sleep may not greatly impact subsequent daytime functioning for a power nap (Lovato & Lack, 2010a), or the diagnosis for a daytime sleepiness test. Whereas, if a device detected sleep onset before an individual had fallen asleep on the vast majority of occasions (e.g. actigraphy devices), then the device may intervene too early. This would be particularly problematic for the administration of ISR because an individual may be disturbed before they have fallen asleep, wasting a sleep onset trial and presumably reducing the efficacy of the treatment. The overestimation of SOL by behavioural devices seems relatively systematic with little variability between individuals (Scott et al., 2018). More research is required to test these devices with individuals with various sleep disorders to determine whether accuracy is impacted by sleep characteristics. This could lead to the development of a device capable of performing these functions reliably and effectively outside the laboratory setting.

Whilst the accurate estimation of sleep onset is essential to develop a device for use outside the laboratory, it is not the only important consideration. For researchers, clinicians and consumers to make use of sleep wearable devices, the device must be practical and user-friendly. Consumer wearable sleep devices have set the expectation that these devices are simple to setup and operate, inexpensive, comfortable to wear, have a long battery life, a user-friendly interface on a smartphone application or computer software, and do not require specialised knowledge to understand and make use of the results. The more complex and expensive a device, the less practical it becomes for lay people to use. But, the less accurate the device, the less useful it is for consumers, researchers and clinicians.

This has been a somewhat overlooked requirement in the development and marketing of some devices, particularly those targeted towards consumers. There is an opportunity for the sleep research community to work alongside industry to develop technology for specific applications with a reach beyond what may be achievable by the community alone, to the benefit of the general population.

Synthesising the findings of 71 articles was challenging because different articles have reported different statistics, and many studies only reported statistics regarding correspondence and not agreement between the PSG and the wearable device. Correlations test an association (correspondence) between two variables. Whereas, agreement indicates the degree of concordance between two assessments that are measuring the same variable: in this case, SOL. A high correlation, or association, does not always equate to high agreement. For example, Figure 2-4 is of two Bland-Altman plots illustrating a hypothetical scenario where correspondence is equal for both plots (r = .94). However, agreement is higher in plot (b) compared to plot (a), as indicated by the bias being closer to zero and the narrower levels of agreement (± 1.96 SDs). Some studies included in this review reported Bland-Altman plots, but the majority of studies reported correspondence statistics, which limited our ability to determine and compare the accuracy of the wearable devices. Relatedly, many studies did not report device specifications which are known to impact the accuracy of devices, such as the algorithm and scoring parameters (de Zambotti, Cellini, Goldstone, Colrain, & Baker, 2019). It is also important to note that only practical wearable devices were included in this review and as such, non-wearable devices may exist that have more robust evidence to support their accuracy.

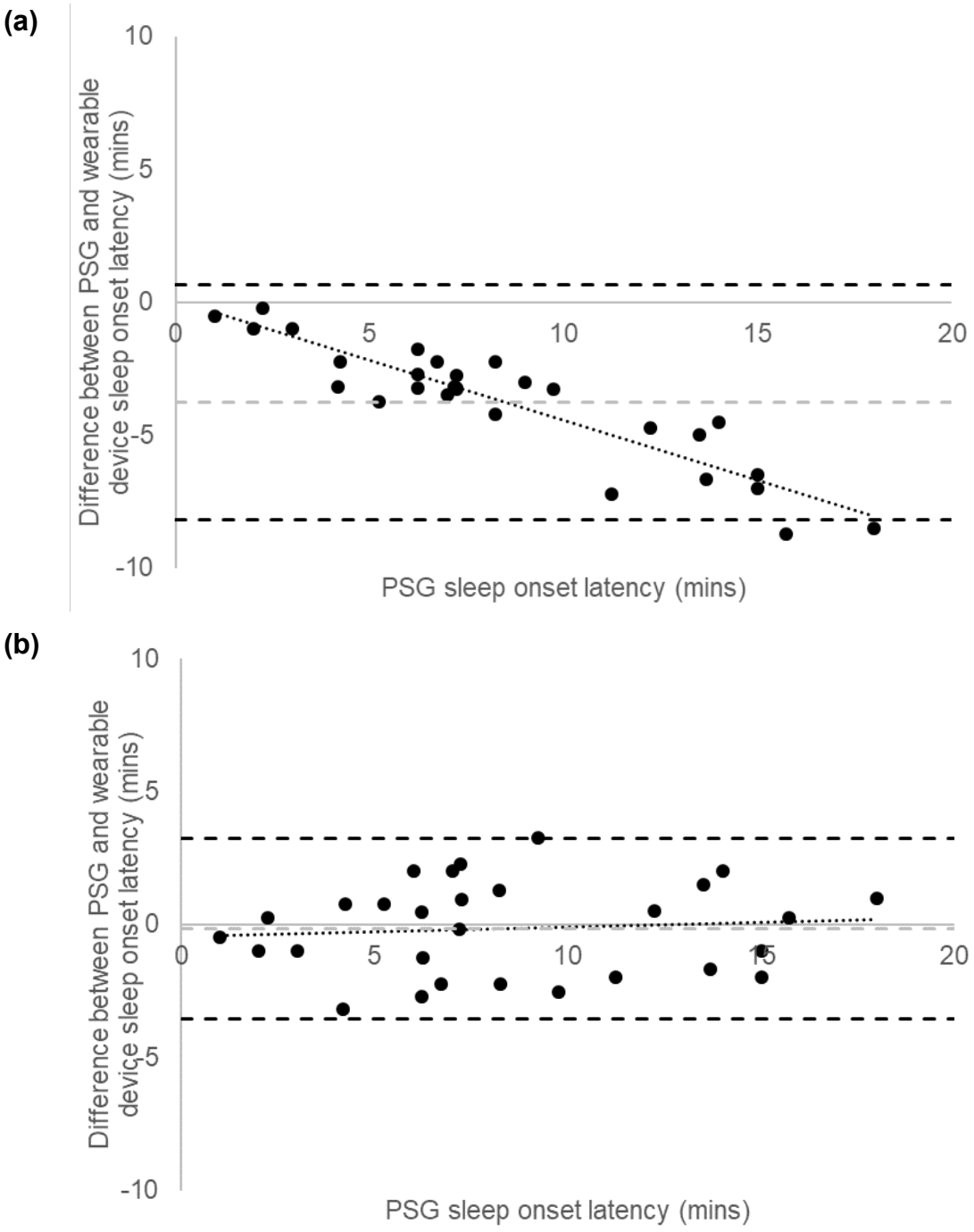


Figure 2-4. Bland-Altman plots of two different scenarios indicating the level of agreement between PSG and a wearable device.
The grey dashed line indicates bias between the two measurements. The black dashed lines indicate the upper and lower levels of agreement. The dotted black line is the trend line.

Conclusion

This review used a systematic approach to investigate the accuracy of wearable devices for measuring sleep onset compared to PSG for the first time.

Several types of devices were identified, including the most widely used method, actigraphy, as well as physiological and other behavioural measurement devices. For administering ISR, power naps and daytime sleepiness tests outside the laboratory setting, behavioural devices were identified as the most suitable type of device. These devices provide accurate estimations of SOL without the limitations of the other types of devices, namely the need for manual and/or retrospective scoring, or large variability in accuracy across individuals. As we begin to understand the ramifications of consumer sleep devices providing inaccurate data to consumers (Gavriloff et al., 2018), it is imperative that consumer sleep devices are refined to be as accurate as possible and provide valuable feedback to consumer, clinicians and researchers. The sleep research community are encouraged to collaborate with industry to refine current technologies, the goal being to develop accurate sleep wearable devices that can not only provide useful feedback but can also help poor sleepers improve their sleep. Clinicians, researchers and health-conscious consumers look forward to the development of a simple, accurate device to translate the findings of sleep research from the laboratory to the home environment.

Practice Points

(1) The accuracy of wearable devices for estimating sleep onset varies by the objective measurement method employed by the device, participant characteristics, and algorithm specifications.

(2) Behaviour-based wearable devices may be most suitable for the purposes of administering Intensive Sleep Retraining, power naps, and daytime sleepiness tests.

Research Agenda

In the future, we need to work with industry to develop and refine practical wearable devices which:

(1) Provide accurate estimations of sleep onset with little variability across individuals.

(2) Administer Intensive Sleep Retraining, power naps, and daytime sleepiness tests practically outside the laboratory setting.

(3) Are accurate enough for these purposes with the specific populations they are designed to be used with, e.g. test with an insomnia sample for administering Intensive Sleep Retraining.

Chapter 3: The Accuracy of the THIM Wearable Device for Estimating Sleep Onset Latency.

Hannah Scott[1], Ashwin Whitelaw[1], Alex Canty[1], Nicole Lovato[2] and Leon Lack[1,2]

[1] College of Education, Psychology and Social Work, Flinders University.

[2] College of Medicine and Public Health, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, Flinders University.

Author contributions: HS contributed to study design, data collection, analysis and interpretation of the data, and manuscript preparation. AW contributed to data collection, analysis and interpretation of the data, and manuscript drafting. AC contributed to data collection and manuscript drafting. LL and NL contributed to conception and study design, interpretation of the data and drafting of the manuscript.

Conflict of interests: Research costs were partially funded by Re-Time Pty Ltd, the company that sell THIM. None of the study authors were financially supported by Re-Time for this project. LL is a shareholder of Re-Time Pty Ltd. LL and HS have a patent pending regarding the THIM device. AW, AC and NL have no conflicts of interest to declare.

Citation: Scott, H., Whitelaw, A., Canty, A., Lovato, N., & Lack, L. (2020). The accuracy of the THIM wearable device for estimating sleep onset latency. under review.

Abstract

THIM is a wearable device that is designed to administer a behavioural insomnia treatment called Intensive Sleep Retraining. To achieve this, it is imperative that THIM can accurately estimate sleep onset. This article presents two studies that aimed to develop the THIM sleep onset detection algorithm compared to polysomnography (PSG). Twelve (Study 1) and twenty (Study 2) individuals slept overnight in the sleep laboratory on two nights, one week apart. On both nights, participants underwent THIM-administered sleep onset trials for four hours with simultaneous PSG recording. During these trials, participants attempted to fall asleep whilst using THIM. Once THIM determined sleep onset, the device woke them up. In Study 1, there was no significant difference between PSG (M = 1.94 min, SD = 1.32) and THIM-sleep onset latency (M = 2.05, SD = 1.38) on the first night, p = .40, with similar findings on the second night, p = .07. On 23.74% of trials, PSG-sleep onset could not be determined before THIM ended the trial. With a revised THIM algorithm in Study 2, there was no significant difference between PSG (M = 3.41, SD = 2.21) and THIM-sleep onset latency (M = 3.65, SD = 2.18), p = .25. There were significantly less trials where PSG-sleep onset had not occurred (10.24%), p = .04. The revised THIM algorithm was accurate at estimating sleep onset latency in average sleepers. Future research will investigate whether THIM is similarly accurate for individuals with insomnia to determine whether the device can administer Intensive Sleep Retraining appropriately.

Keywords: sleep onset latency; Intensive Sleep Retraining; wearable technology; consumer sleep technology; polysomnography; actigraphy.

Accurate objective measurement of sleep onset latency (SOL) is required for a variety of research and clinical purposes. For instance, Intensive Sleep Retraining (ISR) is a behavioural treatment for insomnia that involves repeatedly falling asleep and waking up shortly thereafter over the course of one overnight session (Harris et al., 2012; Harris et al., 2007). Additionally, brief daytime sleeps such as power naps or sleep diagnostic tests like the Multiple Sleep Latency Test (MSLT) involve achieving a precise amount of sleep (Carskadon, 1986; Lovato & Lack, 2010b). These three purposes require the accurate detection of sleep onset so that the individual can be awoken after the appropriate duration of sleep. Yet, the accurate estimation of sleep onset in the home environment is difficult, with the accuracy of popular actigraphy-based wearable devices varying widely across individuals (Scott, Lack, & Lovato, 2019). This limits the translation of these purposes beyond the sleep laboratory. The current article investigated the accuracy of a new wearable device for estimating SOL, which may be used to implement these purposes outside the laboratory setting.

THIM is a new consumer sleep device worn like a ring (Re-Time, 2016). To estimate SOL, THIM administers brief, low intensity vibrations at intervals averaging 30 seconds apart. The individual is required to respond to the vibrations by tapping their finger. When the individual does not respond to two consecutive vibrations, the device infers that they have fallen asleep. Thus, the device can estimate sleep onset in real time shortly after it occurs. THIM can also be programmed to wake the individual after a pre-specified duration of sleep. This means that THIM is capable of administering ISR, power naps, and daytime diagnostic tests (e.g. the MSLT) outside of the laboratory and without the need for expensive equipment or trained individuals to setup, administer or score the data. However, the accuracy of THIM for estimating

sleep onset is currently unknown and must be tested to ensure that it can conduct these applications appropriately.

THIM uses the stimulus-response method to estimate sleep onset. The scoring criteria for polysomnography (PSG) was developed in part by examining electroencephalography (EEG) changes that occur with the cessation of behavioural responses to external stimuli (Dement & Kleitman, 1957; Loomis et al., 1935). Hence, this behavioural method of estimating sleep onset corresponds highly with PSG-defined sleep onset, with responses to stimuli typically ceasing between late-N1 sleep and N2-sleep onset (Ogilvie, 2001; Ogilvie et al., 1989). There is often little individual variability in the accuracy of SOL estimations from these devices because the ability to perceive a stimulus and produce a simple motor movement in response does not vary considerably across individuals, especially when the stimulus is delivered at an individually-tailored minimal intensity (Mair, 1994; Ogilvie et al., 1989; Scott et al., 2018). Therefore, THIM is predicted to estimate SOL accurately compared to PSG, with an expected discrepancy of 2-3 minutes in line with similar devices.

Whilst similar devices using the stimulus-response method are accurate for estimating SOL, THIM differs from previously tested devices in ways that may affect its accuracy. Devices tested in previous research have typically administered auditory stimuli perceived through the auditory perception pathway (Cohen, Bennur, Christison-Lagay, Gifford, & Tsunada, 2016), whereas vibratory stimuli emitted from THIM are perceived through the somatosensory system (Abraira & Ginty, 2013; Kaas, 2012). Whether these pathways show similar inhibition across the sleep onset period is currently unknown.

MacLean, Arnedt, Biedermann, and Knowles (1992) tested the discrepancy

between PSG-sleep onset and behavioural responses (depression of a switch) to a hand-held device that administered vibratory stimuli. The authors found no significant differences between PSG and the hand-held device for estimating SOL. However, the vibratory stimuli were not calibrated to a minimally perceptible level: the vibrations were delivered at five standard deviations above participant's waking threshold. Therefore, responsiveness to minimal intensity tactile stimuli - as utilised by THIM - during the sleep onset period is yet to be tested. Furthermore, THIM requires a simple finger tap response to the vibratory stimuli whereas other devices require larger hand/wrist movements (Mair, 1994; Ogilvie et al., 1989; Scott et al., 2018). Individuals may be able to exert finger taps into deeper stages of sleep than more onerous movements that are inhibited earlier by loss of muscle tension. Other slight variations in the algorithms between THIM and other devices may exist that could impact its accuracy, such as the stimulus intensity, the stimulus duration, and the interval between stimuli. It is therefore important to test the accuracy of THIM for estimating sleep onset as its accuracy may differ from similar devices.

A potential, currently untested limitation of devices that use the stimulus-response method is the effect of learning on the device's accuracy. When using THIM, finger tap responses are elicited frequently in response to vibratory stimuli. Over repeated use, the finger taps may become an automatic response to stimuli that the individual could produce without conscious awareness of the stimuli occurring. Under classical conditioning theory, the finger tap response would become a conditioned response to the vibratory stimuli after many paired repetitions over time. This would be problematic if the conditioned finger tap response could occur during deeper stages of sleep, potentially causing THIM to increasingly overestimate SOL with repeated use. Therefore, it is important to investigate whether

the accuracy of THIM reduces after repeated use.

The current article summarises the development of the THIM device for estimating SOL in comparison to the gold standard objective measure of sleep, PSG. Two studies will be presented. The aim of the first study was to test the accuracy of the initial THIM algorithm for estimating SOL with healthy individuals. The findings informed modifications to the algorithm, with the aim of the second study to assess the accuracy of the revised THIM algorithm with a larger independent sample. Rather than discussing the proprietary THIM algorithms, we aim to describe the research that informed the refinement of the algorithm to the point where THIM could reliably provide accurate estimations of SOL. We also conducted secondary analyses to determine whether the accuracy of THIM is affected by previous use - indicative of potential learning effects. Additionally, we examined whether the accuracy of THIM varies between individuals with good or poor sleep, with a sample that represented the variability of sleep patterns found in the general population.

Study 1: Method

Participants

Ethics approval was obtained from the Flinders University Social and Behavioural Research Ethics Committee, South Australia. Potential participants were recruited via advertisements on community noticeboards and social media. Eligibility criteria was as follows:

1. Self-reported average habitual bedtime between 22:00-00:00 and wake up time between 06:00-08:00;
2. Fluent in English;
3. No self-reported diagnosis of a physical or mental health condition;
4. No active nicotine or illicit substance use, or alcohol (>10 standard drinks

p/wk) or caffeine (>250 mg p/day) dependence;

5. No consumption of medications known to interfere with sleep;

6. No overnight shift work or trans-meridian travel within the last two months;

7. Not pregnant or lactating.

The screening questionnaires comprised of the Insomnia Severity Index ([ISI], Morin, Belleville, Bélanger, & Ivers, 2011) and the Pittsburgh Sleep Quality Index (PSQI), Buysse, Reynolds, Monk, Berman, & Kupfer, 1989) to assess sleep schedules and insomnia symptomology, as well as a health and lifestyle questionnaire to assess physical and mental health conditions, medication use, caffeine/alcohol/nicotine consumption, and recent overseas travel.

Thirteen healthy individuals met eligibility criteria, but one participant withdrew after participating in Night 1. The final sample comprised on twelve individuals, see Table 3-1 for participant characteristic information. Scores on the ISI indicated that five participants had subthreshold levels of insomnia and were categorised as poor sleepers (ISI score ≥ 7), and seven were good sleepers (ISI score < 7).

Table 3-1. *Descriptive characteristics for participants in Study 1.*

Characteristic	**Result (N = 12)**	
Age, mean (SD), y	24.92 (6.05)	
Sex, No. (%)		
Men	3 (25)	
Women	9 (75)	
Weekly alcohol consumption, No. (SD)	0.75 (0.97)	
Daily caffeine consumption, No. (SD)	1.29 (1.05)	
Sleep characteristics	**Good sleepers (N = 7)**	**Poor sleepers (N = 5)**
ISI, mean (SD)	2.14 (1.57)	11.00 (3.39)
PSQI, mean (SD)	3.26 (1.50)	7.40 (3.29)
Habitual Bedtime, mean (SD), min	22:38 (28.44)	22:36 (31.64)
Habitual Wake Up Time, mean (SD), min	07:10 (24.41)	07:30 (20.42)
Habitual TST, mean (SD) hrs	8.11 (1.02)	7.10 (1.52)

ISI = Insomnia Severity Index, N = sample size, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

Design

This study employed a within-groups quasi-experimental design. All participants slept overnight in the sleep laboratory on three nights as part of a larger THIM project. The current study only concerns data from Night 2 and Night 3. On both nights, participants underwent sleep onset trials administered by THIM with simultaneous PSG recording. The predictor variable was PSG-SOL during each trial and the primary outcome variable was THIM-derived SOL.

Materials

Polysomnography

PSG was recorded using Compumedics Grael 4K PSG:EEG devices (Compumedics, Victoria, Australia). Six EEG (F3-M2, F4-M1, C3-M2, C4-M1, O1-

M2, O2-M1), reference and ground, right and left electrooculography (EOG), chin electromyography (EMG), and electrocardiography (ECG) sites were sampled at 256Hz. PSG data was scored using Profusion Compumedics software (v4.0) by a qualified, independent sleep technician. In accordance with AASM scoring criteria (American Academy of Sleep Medicine, 2018), PSG-SOL was defined as the time between the start of the attempt to sleep (beginning of the sleep onset trial) and the first epoch of any stage of sleep during the trial.

THIM

THIM (firmware v1.0.3) is a small, ring-like device worn on the index finger of the dominant hand. To setup THIM, the device was connected via Bluetooth to the accompanying smartphone application (v1.0.1) using an Apple iPhone 5s model (iOS 8.0). Participants started a sleep onset trial by tapping their index finger on which THIM was placed onto their thumb, twice in quick succession (see Figure 3-1). During the trials, the device emitted low intensity, short duration vibratory stimuli at non-regular intervals (averaging 30 seconds apart). The intensity of the vibrations was individually calibrated to the minimum level that the participant could consistently respond to whilst awake using the threshold hunting procedure outlined in the THIM smartphone application. Participants were required to respond to the vibratory stimuli by tapping their index finger once onto their thumb, with responses detected by the device's accelerometer. If participants failed to respond to two consecutive vibratory stimuli, the device inferred that sleep onset had occurred and it emitted a high intensity alarm vibration to wake them up, signalling the end of the trial. Shortly afterwards (approximately 1-2 minutes later), participants attempted another trial. THIM's estimations of SOL is the time from the beginning of the trial to slightly before the time of the first of the two consecutively-missed vibratory stimuli.

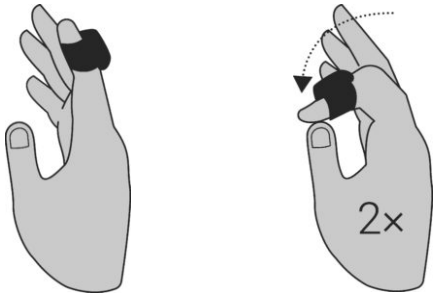


Figure 3-1. Illustration of the finger tap motion with the THIM device.

To monitor THIM, we mounted a small piezo-electric sensor to the side of the THIM device using adhesive tape. This sensor was inputted into a channel on the PSG device. From this sensor, we observed four events of interest: vibrations emitted from THIM, finger taps as responses to the vibrations, as well as the beginning (the double-tap motion) and end (the high-intensity alarm vibration) of each trial. These four events were scored manually on the Profusion Compumedics software by two scorers (HS and AW). If the events of interest on the sensor data were obscured by body movements, the trial was removed from analysis. The sensor data allowed the PSG and THIM data to be precisely time-locked, reducing error of measurement. The interrater reliability on 10 randomly selected nights of data exceeded 95% agreement between the two scorers.

Procedure

Home Testing

Participants completed a sleep diary based on the Consensus Sleep Diary (Carney et al., 2012) and wore an actigraphy device (Actiwatch-2, Philips Respironics) every day for one week to monitor their sleep pattern prior to the first laboratory night. Participants' average bedtimes and wake up times were calculated from the sleep diary to inform the timing of the study protocol. The actigraphy data corroborated the bedtimes and wake up times reported in the sleep diaries.

Laboratory Night 1

The first night was an adaptation night to help participants become accustomed to sleeping in the laboratory environment with the sleep monitoring equipment. Participants went to bed at their typical bedtime and slept overnight whilst monitored by PSG and THIM. They were awoken at their typical wake up time when both devices were removed, and participants left the sleep laboratory. Participants continued to wear the Actiwatch-2 device during the subsequent day to confirm that they did not nap prior to Night 2.

Laboratory Night 2

Participants arrived at the sleep laboratory at approximately 20:00 and were setup for overnight PSG recording. The THIM device was placed on the participant's index finger on their dominant hand along with a piezo-electric sensor secured to the side of the device. After setting the vibratory stimulus intensity, participants received instructions from research assistants on how to operate THIM. See Appendix 5 for this procedure.

An hour before participants' bedtime, they began THIM-administered sleep onset trials that continued for four hours, three hours past their habitual bedtime. Compliance was confirmed by qualified research assistants observing participants via video recording and the THIM sensor data in real-time. Once THIM determined sleep onset during the final trial, instead of emitting a high intensity alarm vibration, the device let them sleep uninterrupted until they spontaneously awoke in the morning. All devices except the Actiwatch-2 device were removed and participants returned home.

Home Testing

Between Night 2 and Night 3, participants completed sleep diaries and wore the Actiwatch-2 device every day for another week.

Laboratory Night 3

Participants returned to the sleep laboratory to undergo the same testing protocol as experienced on Laboratory Night 2.

Data Analysis

The mean PSG and THIM estimations of SOL was calculated in minutes from each trial, separately for each individual, before averaging together for Nights 2 and 3. Paired samples t-tests were then conducted to test whether THIM significantly underestimated or overestimated SOL compared to PSG, separately for both laboratory nights. The mean discrepancies between PSG and THIM were calculated for each individual separately. Then, these individual means were averaged together for each night so that each individual contributed equal weighting to the overall mean. Positive mean discrepancy values meant that THIM overestimated SOL, whereas negative values indicated that THIM underestimated SOL compared to PSG.

The level of agreement between PSG and THIM was assessed with Bland-Altman plots, which shows the discrepancy between PSG and THIM-SOL (y axis) against PSG N1 SOL (x axis) across all trials on each (Bland & Altman, 1986). This involved calculating the mean difference (bias) and the limits of agreement (± 1.96 SD of the mean difference) between these measures. The r squared value for the linear regression line and coefficient p value are reported in the figures. Some datapoints represent many overlapping values.

To examine differences in the accuracy of THIM after repeated use which may indicate a learning effect, a paired samples t-test was conducted to compare the discrepancies between PSG and THIM-SOL on Night 2 versus Night 3. Additionally, paired samples t-tests were conducted to compare differences in the discrepancy

between PSG and THIM-SOL on Night 2 versus Night 3 for each trial (e.g. on the first, second, third trial, etc.). To examine the impact of participants' sleep quality on the accuracy of THIM, an independent samples t-test was conducted to determine whether the discrepancy between PSG and THIM differed between good or poor sleepers, separately for Night 2 and Night 3.

Study 1: Results

First Sleep Onset Trial Night

On laboratory Night 2, there was no significant difference between the mean PSG-SOL (M = 1.94 min, SD = 1.32) and mean THIM-SOL (M = 2.05 min, SD = 1.38), $t(11) = -0.88$, $p = .40$, $d = .08$. The mean discrepancy between PSG and THIM-SOL on this night was low, M = 0.08 min, SD = 0.49. The level of agreement between PSG and THIM-SOL on Night 2 is illustrated in Figure 3-2. As shown by the narrow levels of agreement, there is little variability in the discrepancy between PSG and THIM-SOL across the 411 trials. Furthermore, the discrepancy between PSG and THIM is consistent across trials with increasing latency duration, as indicated by the blue trendline. Of note, are data points above the upper limit of agreement that seem to depict trials where participants were responding to THIM's vibratory stimuli for 5+ mins into PSG-sleep. Closer inspection of these trials revealed that participants did not remain asleep after the first epoch of PSG-sleep in these trials: participants were fluctuating between wake and N1 sleep during this time.

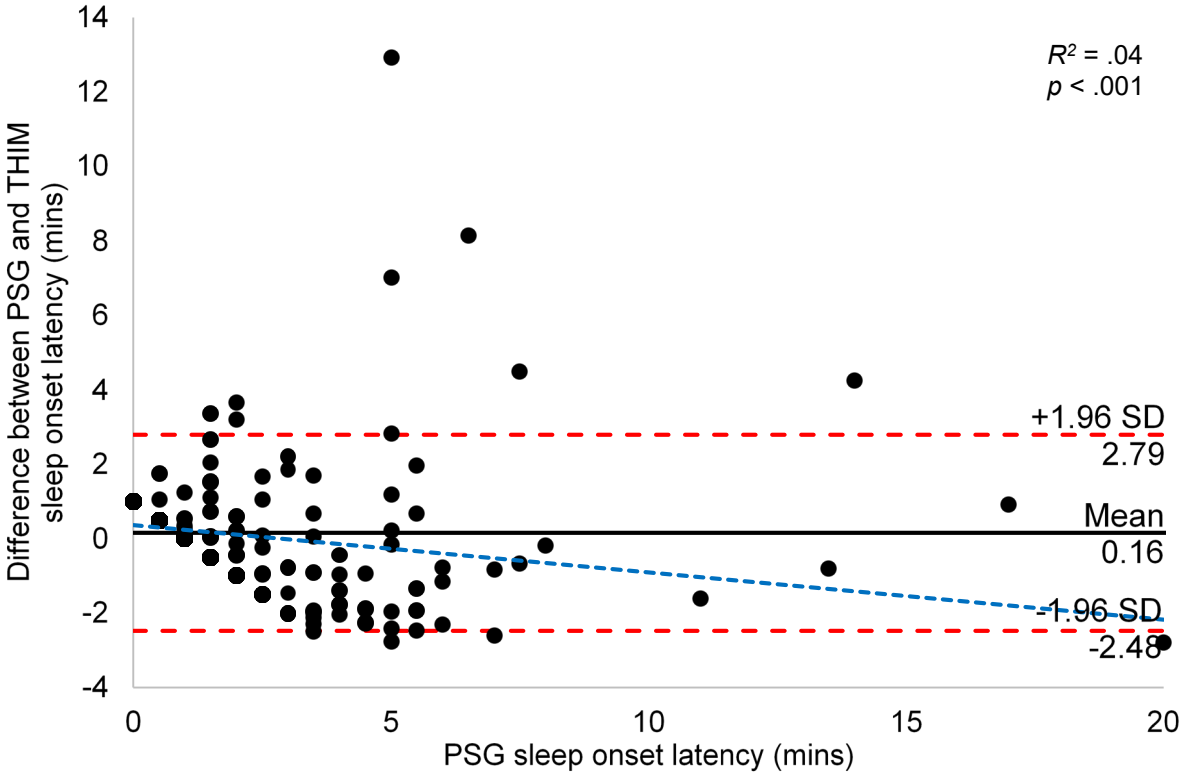


Figure 3-2. Bland-Altman plot indicating agreement between PSG and THIM-sleep onset latency on Night 2.
The solid black line indicates the mean difference, the dotted red lines indicate the upper and lower limits of agreement and the dotted blue line is the linear trendline.

Second Sleep Onset Trial Night

There was no significant difference between mean PSG-SOL (*M* = 1.40 min, *SD* = 0.64) and mean THIM-SOL (*M* = 2.12 min, *SD* = 1.71) on laboratory Night 3, *t*(11) = -2.02, *p* = .07. Despite a medium effect size, *d* = 0.56, the mean discrepancy between PSG and THIM-SOL on this night was still relatively low, *M* = 0.57 min, *SD* = 1.10. Figure 3-3 is a Bland-Altman plot illustrating the level of agreement between PSG and THIM-SOL across all Night 3 trials. Similar to Figure 3-2, the variability in the discrepancy between PSG and THIM-SOL across 527 trials is low. Figure 3-3 also shows trials where participants were responding to THIM's vibratory stimuli whilst fluctuating between wake and N1 sleep (points above the upper limit of agreement).

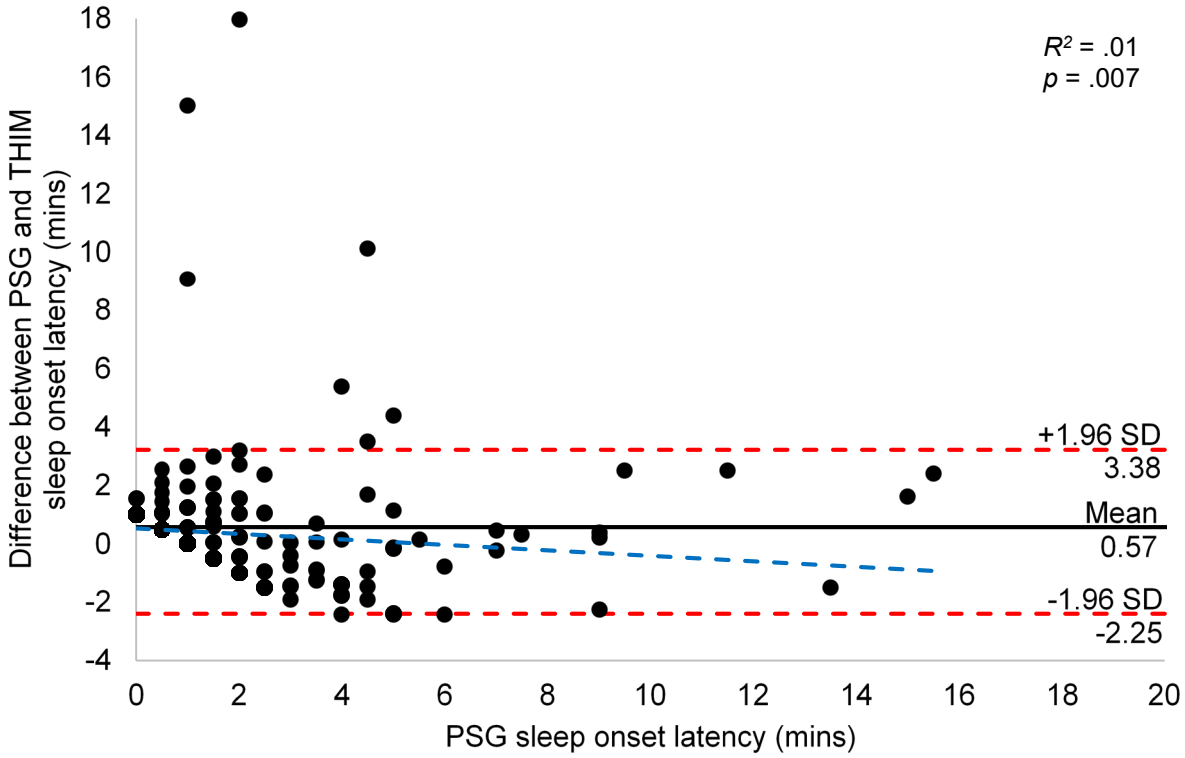


Figure 3-3. Bland-Altman plot indicating agreement between PSG and THIM-sleep onset latency on Night 3.
The solid black line indicates the mean difference, the dotted red lines indicate the upper and lower limits of agreement and the dotted blue line is the linear trendline.

Learning Effects

A paired samples t-test indicated that there was no significant difference in the mean discrepancy between PSG and THIM-SOL on Night 2 compared to Night 3, $t(11) = -1.90$, $p = .08$. There was a medium effect size, $d = 0.57$. Paired samples t-tests revealed no significant differences in the discrepancy between PSG and THIM on Night 2 versus Night 3 for any trial (e.g. on the first, second, third trial, etc.), $p > .10$. The accuracy of THIM compared to PSG appears to remain high and does not significantly decrease, even after repeated use.

Good and Poor Sleeper Comparison

An independent samples t-test revealed that there was no significant difference in the mean discrepancy between PSG and THIM-SOL on Night 2 for good sleepers ($M = 0.06$ min, $SD = 0.44$) compared to poor sleepers ($M = 0.09$ min,

SD = 0.60), *t*(10) = -0.11, *p* = .92, *d* = 0.08. Similarly, there was no significant difference in the mean discrepancy on Night 3 between good sleepers (*M* = 0.34 min, *SD* = 0.21) and poor sleepers (*M* = 0.88 min, *SD* = 1.75), *t*(4.08) = -0.68, *p* = .53, although there was a medium effect size, *d* = 0.48. Therefore, the accuracy of THIM does not appear to differ between good and poor sleepers.

THIM False Positive Trials

Due to a slight delay between THIM-sleep onset and the end of the trial, there were some occasions where THIM underestimated sleep onset but PSG-sleep onset was reached before THIM ended the trial, as shown in Figures 3-2 and 3-3. However, it became apparent that there was a considerable proportion of sleep onset trials during which PSG-sleep onset had not occurred before THIM estimated sleep onset and ended the trial. Because a PSG-SOL datapoint was unavailable for those trials, and it could not be predicted, they were excluded from the above analyses. On average, PSG-sleep onset had not occurred in an average of 15.42 (*SD* = 16.22, 31.04% of Night 2 trials) of Night 2 trials per participants where THIM had detected sleep onset. Similarly, there was an average of 8.92 'false positive' trials (*SD* = 9.82, 16.88%) per participant on Night 3. There was no significant difference between Nights 2 and 3 on the number of false positive trials, *t*(11) = 1.47, *p* = .17, *d* = 0.49.

There are several possible reasons for the THIM determination of sleep onset when participants were still awake according to PSG. One potential explanation is that participants did not respond to the vibratory stimulus because they did not perceive it. However, this was not the case for the majority of these false positive trials. Participants did not respond to either of the last two consecutive vibratory stimuli for 28.42% of these false positive trials on Night 2 and 42.00% of these trials

on Night 3. In other words, participants had indeed responded to one or both of the last two consecutive vibratory stimuli before the trial ended, but the device had not registered the response. This was true for the majority of false positive trials on both Night 2 (71.58%) and Night 3 (58.00%).

To register as a legitimate response to vibratory stimuli, finger tap responses had to meet timing and intensity criteria. In order to exclude any spontaneous, random finger twitches, a time window following the stimulus was established during which the response had to occur to meet the valid response criterion. THIM failed to detect 42.02% on Night 2 and 48.77% on Night 3 of responses that occurred just beyond the time window. Therefore, a majority of the finger tap responses on Night 2 and Night 3 occurred within the required time window yet were not registered by THIM. This is presumably because the finger taps were not vigorous enough to exceed the accelerometer threshold criterion required to register as a legitimate response.

Study 1: Discussion

The aim of Study 1 was to test the accuracy of THIM for estimating SOL against PSG. Overall, there was strong agreement between THIM and PSG, regardless of sleeper type (good or poor sleeper status) and repeated use (Night 2 versus Night 3). Having said this, THIM had estimated sleep onset and prematurely ended the trial before PSG-sleep onset criteria were met on a considerable number of trials. This is an issue for two reasons. Firstly, we needed to exclude these trials from analysis: 23.74% of trials across Night 2 and Night 3. This undermined our ability to make strong conclusions about the accuracy of THIM. Secondly, this issue is problematic for the administration of many functions, including ISR. If THIM determined that the patient had fallen asleep and ended the trial when they were still

awake, then the trial would be a wasted retraining opportunity as presumably, sleep onset must occur during the trial to obtain therapeutic benefit. Of greater concern, the patient may perceive that they are unsuccessful at the treatment if they correctly perceive that they were awake during these trials and may subsequently doubt their ability to asleep. Whilst ISR is a behavioural treatment for insomnia, the experience of rapidly falling asleep on multiple occasions may be therapeutic in a cognitive sense by reassuring the patient that their sleep mechanism is not 'broken', and they retain the ability to sleep.

Consequently, we made recommendations to the manufacturers of THIM, Re-Time Pty. Ltd., about potential modifications to the THIM algorithm. The recommendations included reducing the threshold accelerometer intensity required for a legitimate finger twitch and expanding the time window during which such a response could occur to include the full distribution of reaction times to the vibratory stimuli observed in Study 1. The company incorporated these modifications into a revised algorithm, which we tested in the second study to determine whether the issue had been resolved.

Study 2: Method

The study design, materials, study protocol, and data analysis plan of the second study were identical to the first study, except that we tested the revised version of THIM (firmware v1.0.4) with a larger, independent sample.

Participants

Participants of the second study were required to meet the same eligibility criteria as participants in the first study. Twenty healthy individuals met eligibility criteria and consented to participate. ISI scores at screening indicated that ten participants had subthreshold levels of insomnia and were categorised as poor

sleepers (ISI score ≥ 7), and ten were good sleepers (ISI score < 7). See Table 3-2 for participant characteristic information and a comparison between the Study 1 and Study 2 samples. There were no significant differences on the participant characteristics between the two samples.

Table 3-2. *Participant characteristics for Study 2 compared to Study 1.*

Characteristic	**Study 1 (N = 12)**	**Study 2 (N = 20)**	**Comparison between Studies**
Age, mean (SD), y	24.92 (6.05)	23.58 (4.89)	$t(30) = 0.68$, $p = .50$
Sex, No. (%)			
Men	3 (25)	7 (35)	$\chi(1) = 1.66$, $p = .20$
Women	9 (75)	13 (65)	
Weekly alcohol consumption, No. (SD)	0.75 (0.97)	1.60 (1.79)	$t(29.80) = -1.51$, $p = .14$
Daily caffeine consumption, No. (SD)	1.29 (1.05)	1.89 (1.47)	$t(30) = -1.20$, $p = .24$
Sleep characteristics	**Good sleeper (N = 7)**	**Poor sleeper (N = 5)**	**Good sleeper (N = 10)**	**Poor sleeper (N = 10)**	
ISI, mean (SD)	2.14 (1.57)	11.00 (3.39)	2.00 (1.15)	11.70 (3.86)	$t(30) = -0.51$, $p = .62$
PSQI, mean (SD)	3.26 (1.50)	7.40 (3.29)	3.10 (1.73)	8.30 (3.09)	$t(30) = -0.56$, $p = .58$
Habitual Bedtime, mean (SD), min	22:38 (28.44)	22:36 (31.64)	22:45 (64.58)	23:02 (68.41)	$t(28.93) = -1.01$, $p = .32$
Habitual Wake Time, mean (SD), min	07:10 (24.41)	07:30 (20.42)	07:27 (61.27)	07:56 (72.23)	$t(26.93) = -1.47$, $p = .15$
Habitual TST, mean (SD) hrs	8.11 (1.02)	7.10 (1.52)	8.05 (0.83)	7.10 (1.58)	$t(30) = 0.24$, $p = .82$

ISI = Insomnia Severity Index, N = sample size, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

Study 2: Results

First Sleep Onset Trial Night

One PSG recording failed due to technical error and thus, this night's data is only based upon 19 participants. With the revised THIM algorithm, there was still no

significant difference between PSG (*M* = 3.41 min, *SD* = 2.21) and THIM estimations of mean SOL (*M* = 3.65 min, *SD* = 2.18) on laboratory Night 2, *t*(18) = -1.18, *p* = .25, *d* = 0.11. There was a small mean discrepancy between the two measures, *M* = 0.24 min, *SD* = 0.90. As shown in Figure 3-4, there was strong agreement between PSG and THIM-SOL across 535 trials.

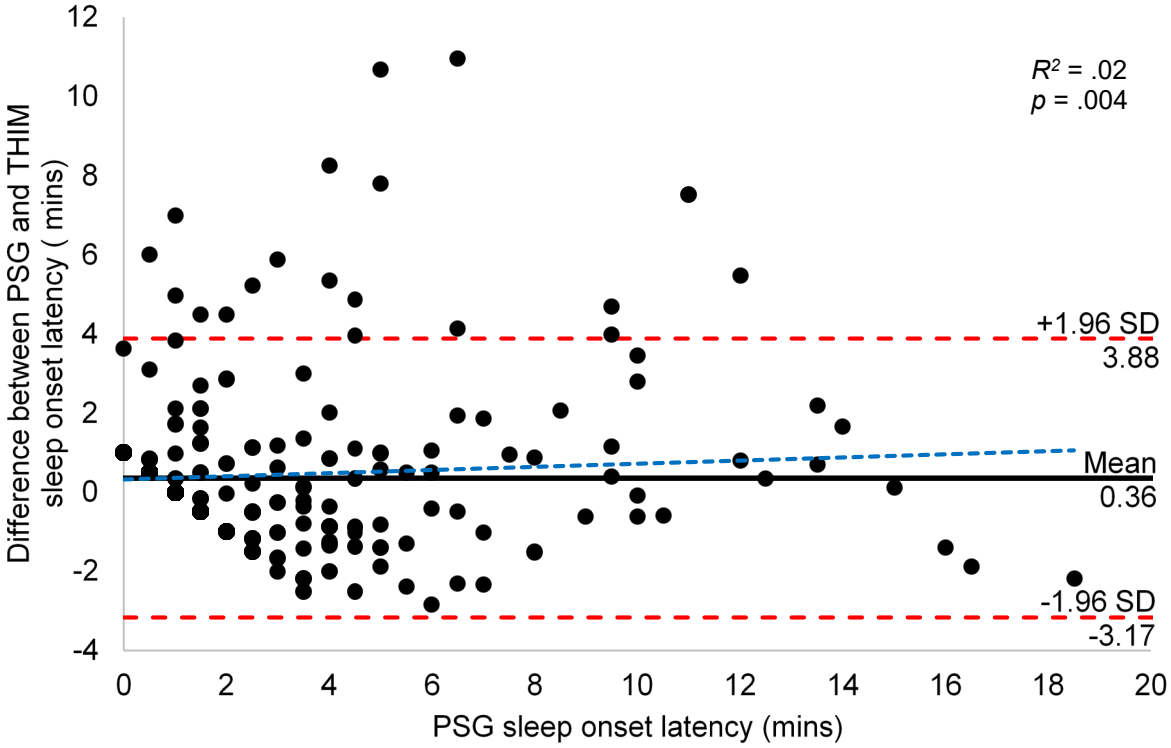


Figure 3-4. Bland-Altman plot indicating agreement between PSG and THIM-SOL on Night 2 for Study 2 data.
The solid black line indicates the mean difference, the dotted red lines indicate the upper and lower limits of agreement and the dotted blue line is the linear trendline.

Second Sleep Onset Trial Night

Unlike Night 2, on Night 3 there was a significant difference between PSG (*M* = 3.93 min, *SD* = 3.32) and THIM-SOL (*M* = 4.75 min, *SD* = 3.85). THIM significantly overestimated SOL compared to PSG, *t*(19) = -2.78, *p* = .01, *d* = 0.23. However, the effect size and mean discrepancy between PSG and THIM was still low, *M* = 0.82 min, *SD* = 1.31. Figure 3-5 continued to show strong agreement between PSG and THIM across 578 trials, as evident by the narrow levels of agreement.

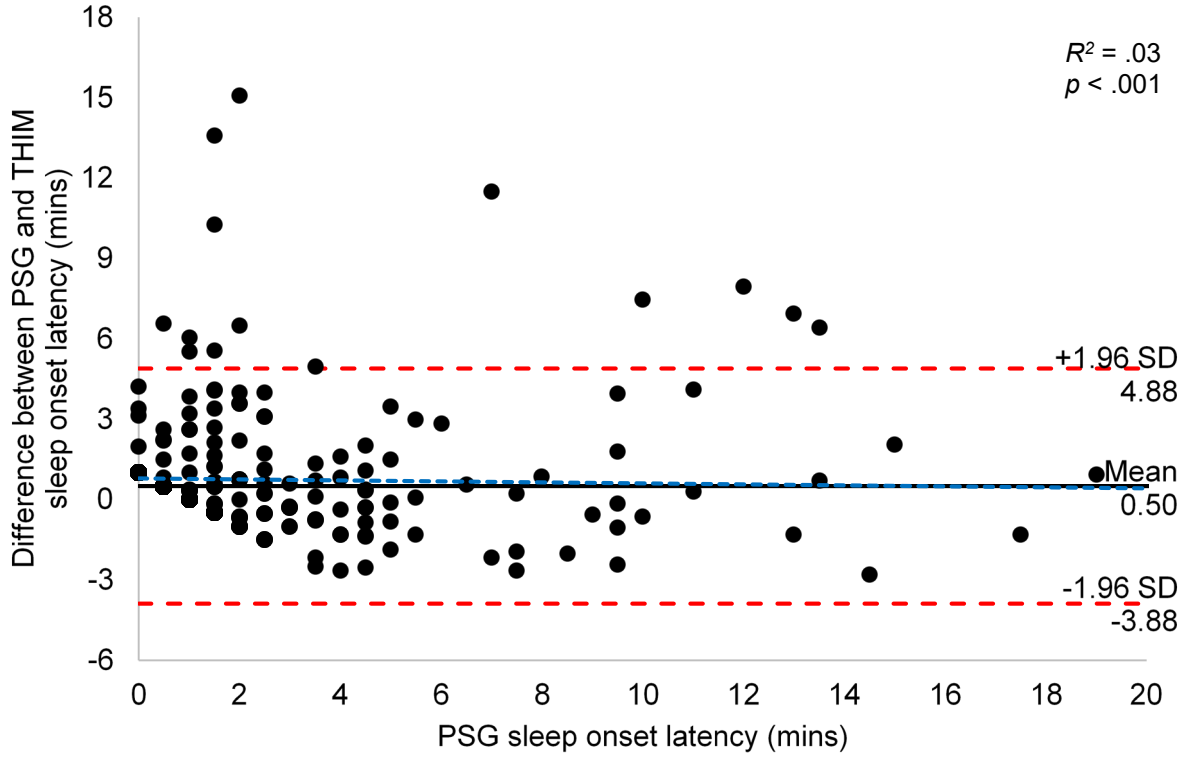


Figure 3-5. Bland-Altman plot indicating agreement between PSG and THIM-SOL on Night 3 for Study 2 data.
The solid black line indicates the mean difference, the dotted red lines indicate the upper and lower limits of agreement and the dotted blue line is the linear trendline.

Comparison between THIM algorithms

The goal of revising the THIM algorithm was to reduce the number of THIM false positive trials. With the revised algorithm, there was a mean of 4.05 false positive trials (*SD* = 3.76) per participant on Night 2 and 2.53 trials (*SD* = 2.09) per participant on Night 3, or 10.24% of trials overall.

We conducted an independent samples t-test to determine whether the issue occurred in less trials with the revised THIM algorithm compared to the original algorithm. There was a significantly lower number of false positive trials with the revised algorithm compared to the original algorithm on Night 2, $t(11.75) = 2.39$, $p = .04$, and Night 3, $t(11.57) = 2.24$, $p = .046$. The effect sizes were large for Night 2, $d = 1.09$, and Night 3, $d = 1.04$. Considering that the issue occurred in a smaller minority of trials in Study 2, it appears that the modifications made to the THIM

algorithm largely resolved this issue without substantially increasing the mean discrepancy between THIM and PSG.

Learning Effects

As in Study 1, there was no significant difference between the mean discrepancy of PSG and THIM-SOL on Night 2 compared to Night 3, $t(18) = -1.84$, $p = .08$, although there was a medium effect size, $d = 0.51$. Additional paired samples t-tests revealed no significant differences in the discrepancy between PSG and THIM on Night 2 versus Night 3 on any trial, $p > .13$. Therefore, the accuracy of THIM does not appear to significantly reduce after repeated use.

Good and Poor Sleeper Comparison

An independent samples t-test showed no significant difference in the mean discrepancy between PSG and THIM-SOL on Night 2 for good sleepers ($M = 0.45$ min, $SD = 0.88$) compared to poor sleepers ($M = 0.55$ min, $SD = 0.68$), $t(17) = -0.28$, $p = .78$, $d = 0.13$. Similarly, there was no significant difference in the mean discrepancy on Night 3 between good sleepers ($M = 0.89$ min, $SD = 1.65$) and poor sleepers ($M = 0.87$ min, $SD = 1.06$), $t(18) = 0.03$, $p = .98$, $d = 0.01$. This is further evidence to suggest that sleeper type does not affect the accuracy of THIM.

Study 2: Discussion

The aims of both studies were to assess the accuracy of THIM for estimating SOL compared to PSG. Study 1 tested the original THIM algorithm and Study 2 tested a THIM algorithm that was modified based on the findings from Study 1. THIM-SOL showed strong agreement with PSG N1-SOL, for both good and poor sleepers and even after repeated use (Night 2 compared to Night 3). The revised THIM algorithm also largely resolved an issue found in Study 1 where THIM estimated that sleep onset had occurred in trials where PSG-sleep onset criteria

were not yet met. Therefore, the revised algorithm was an improvement upon the original algorithm and the device appeared to remain accurate for estimating SOL.

THIM had considerably closer agreement with PSG-N1 sleep onset compared to other wearable devices (Cellini et al., 2013; Chae et al., 2009). The next generation of actigraphy devices that incorporate information from additional sensors, such as heart rate variability, appear to have greater accuracy compared to standard actigraphy devices (de Zambotti et al., 2018; Fonseca et al., 2017). However, THIM shows greater agreement with PSG for estimating SOL than these multi-sensor devices. In fact, THIM produced comparable accuracy to simplified EEG-based devices (Cellini et al., 2015; Kaplan et al., 2014; Markwald, Bessman, Reini, & Drummond, 2016a).

THIM also showed closer agreement with PSG N1 sleep onset than similar devices that also use the stimulus-response method of sleep onset estimation (Mair, 1994; Ogilvie et al., 1989). This may be due to differences in the stimulus type. THIM uses vibratory stimuli, which is perceived via a different sensory processing pathway compared to the auditory stimuli utilised by similar devices (Cohen et al., 2016; Kaas, 2012). It is more likely that the difference in accuracy between THIM and similar devices is due to the stimulus type rather than due to other variations between devices. It was evident from the piezo-electric motion sensor data collected during Study 2 that once participants entered PSG N1 sleep, they ceased responding to the vibratory stimuli. This suggests that participants either a) did not perceive the vibratory stimulus and remained totally asleep, or b) the individual stirred from sleep slightly, but the vibratory stimulus was not salient enough to arouse the individual enough to produce a finger tap response. A quantified EEG analysis comparing brainwave activity before and after a vibratory stimulus would

shed light on whether participants aroused at all to vibratory stimuli during N1 sleep. Either way, it appears that the type of stimulus to which participants respond may impact the accuracy of stimulus-response devices. Future research could directly compare the use of different types of low intensity stimuli to determine when each sensory system is inhibited during the sleep onset period.

An important limitation to consider is that the PSG data was scored by only one qualified sleep technician in the current study. The interrater reliability of N1-sleep onset in particular is relatively low (Rosenberg & Van Hout, 2013). This adds to the error of measurement in the gold-standard measure that should be considered when interpreting the findings of the current study.

Investigating the accuracy of THIM for individuals with insomnia is particularly important for the administration of ISR because the device may be less accurate with this population. In line with the neurocognitive model of insomnia (Perlis, Giles, Mendelson, Bootzin, & Wyatt, 1997), individuals with insomnia may have abnormally sensitive/acute sensory and information processing during the sleep onset period. Increased sensory responsivity may mean that people with insomnia perceive vibratory stimuli beyond N1-sleep onset more so than average sleepers. Consequently, THIM may overestimate SOL to a greater extent for those with insomnia compared to good sleepers. The current studies did not include individuals with insomnia, but there was no significant difference in the accuracy of THIM between good and poor sleepers. However, neither of the two studies presented were adequately powered to detect small differences between groups that may be relevant. Furthermore, insomnia-related arousal may not be present for those identified as having poor sleep: this conditioned arousal is theorised to develop over time (Perlis, Smith, & Pigeon, 2005), whereas poor sleep in general may be episodic

in nature (Perlis et al., 2019).Therefore, the accuracy of THIM should be investigated with individuals with insomnia in future research.

Conclusion

This article showcased the development of the THIM algorithm for estimating SOL in comparison to PSG. Study 1 showed that there was high agreement between THIM and PSG. However, on a considerable percentage of trials, THIM determined that the participant had fallen asleep when they were still awake according to PSG scoring criteria. This led to modifications to the THIM algorithm and Study 2 showed that the revised algorithm had similarly high agreement with PSG but with a considerably lower percentage of false positive trials. Additionally, repeated use and sleeper type (good or poor sleeper) did not impact the accuracy of THIM in either study. More research is needed to investigate whether other individual characteristics affect the accuracy of THIM, particularly a diagnosis of insomnia. Collaborating with industry resulted in the development of an accurate device that may allow for the widespread implementation of many research and clinical applications - to the benefit of patients, researchers, and clinicians wanting to improve sleep.

Chapter 4: Correspondence between Physiological and Behavioural Responses to Vibratory Stimuli during the Sleep Onset Period: A Quantitative Electroencephalography Analysis.

Hannah Scott[1], Bastien Lechat[2], Nicole Lovato[3] and Leon Lack[1,3]

[1] College of Education, Psychology and Social Work, Flinders University.

[2] College of Science and Engineering, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, Flinders University.

[3] College of Medicine and Public Health, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, Flinders University.

Author contributions: HS contributed to study design, data collection, analysis and interpretation of the data, and manuscript preparation. BL contributed to data analysis, interpretation of the data, and drafting of the manuscript. LL and NL contributed to conception and study design, interpretation of the data and drafting of the manuscript.

Conflict of interests: The research costs were partially funded by Re-Time Pty Ltd, the company that sell the THIM device. None of the study authors were financially supported by Re-Time for this project. LL is a shareholder of Re-Time Pty Ltd. LL and HS have a patent pending regarding the THIM device. BL and NL have no conflicts of interest to declare.

Abstract

Behavioural responses to auditory stimuli cease in late N1-sleep or early N2 sleep. Yet, responsiveness to minimal intensity tactile stimuli and the correspondence with sleep microstructure during the sleep onset period is unknown. The aim of the current study was to investigate sleep microstructure when participant behaviourally responded to minimal intensity vibratory stimuli compared to when participants did not respond to stimuli during the sleep onset period.

Eighteen participants wore a device that emitted vibratory stimuli to which individuals responded by tapping their index finger. A Fast-Fourier Transform using multitaper-based estimation was applied to the electroencephalography signal in 5-second epochs. Participants exhibited increases in higher frequencies five seconds before and immediately after the stimulus presentation when they responded to the stimulus compared to when they did not respond during all sleep stages. They also had greater delta power after stimulus onset when they did not respond to stimuli presented in N1- and N2-sleep compared to when they did respond. Participants responded to a significantly greater proportion of stimuli in wake compared to in N1-sleep, $p < .001$, $d = 2.38$, which was also significantly greater than the proportion of responses in N2-sleep, $p < .001$, $d = 1.12$.

Participants showed wake-like sleep microstructure when they responded to stimuli, and sleep-like microstructure when they did not respond during any sleep stage. This study adds to the body of evidence that characterises N1 sleep as a transitional period between sleep and wake containing rapid fluctuations between these two states.

Keywords: wearable technology; consumer sleep technology; sleep onset period; electroencephalography; sleep stages; quantitative EEG.

Sleep onset is considered to be not a specific point in time, but rather a transitional period between wakefulness and sleep. Whilst there is debate about when the sleep onset period begins and ends, much of the literature suggests that the onset of N1-and N2-sleep are the start and end points of this transition (Carskadon & Dement, 2017; Ogilvie, 2001). During this period, a variety of physiological, behavioural and psychological changes occur, not necessarily at the same time. However, this complexity is not reflected in traditional electroencephalography (EEG) sleep scoring criteria (American Academy of Sleep Medicine, 2018). According to this criteria, N1-sleep onset is defined as a reduction in alpha to less than 50% of the 30-second epoch. Thus, standardised criteria do not capture many of the physiological, behavioural or psychological changes that characterise the process of falling asleep.

The correspondence between behavioural indications of sleep and changes in physiology were utilised to develop the EEG scoring criteria (Dement & Kleitman, 1957; Loomis et al., 1935). As such, there is strong agreement between these two measures. For instance, reduced responsiveness to minimal intensity external stimuli typically occurs in late-EEG N1 sleep or early N2 sleep (Harsh, Voss, Hull, Schrepfer, & Badia, 1994; Ogilvie et al., 1989; Scott et al., 2018), at which point, individuals have reduced awareness of the external environment (Bonnet, 1986; Oniz, Inanc, Guducu, & Ozgoren, 2016). Reduced awareness is likely due to an increase in the threshold needed to perceive a stimulus around N2 sleep (Bonnet & Moore, 1982; Rechtschaffen, Hauri, & Zeitlin, 1966) although an intense or salient enough stimulus may still elicit a response after this point (Blume et al., 2017; Perrin, García-Larrea, Mauguière, & Bastuji, 1999). Lauerma, Kaartinen, Polo, Sallinen, and Lyytinen (1994) found that the probability of behaviourally responding to an auditory

stimulus was 12.2% in N2 sleep and 61.7% in N1 sleep compared to 84.8% in wake. Czisch et al. (2002) found that there was reduced activation in the auditory cortex during N1 and N2 sleep. Together, these findings suggest that responses to auditory stimuli are less likely but still possible after N1 sleep onset, with responses after this point potentially coinciding with brief awakenings or arousals (Ogilvie, Simons, Kuderian, MacDonald, & Rustenburg, 1991).

Previous research investigating behavioural responsiveness to external stimuli during the sleep onset period have largely been conducted with minimal intensity auditory stimuli (Kuderian et al., 1991; Ogilvie et al., 1989; Scott et al., 2018). Only one previous study has examined behavioural responsiveness to vibratory stimuli during the sleep onset period (MacLean et al., 1992). The authors found that individuals were much less likely to respond in N1 and N2 sleep than during wake. However, the vibratory stimuli were presented at an intensity of five standard deviations above waking threshold. Nonetheless, we expected similar findings when testing correspondence between EEG-sleep onset and a behavioural device that used behavioural responses to minimal intensity vibratory stimuli. Our research produced an unexpected finding: the cessation of responses to vibrations aligned much more closely with EEG N1-sleep onset than responses to auditory stimuli. We found an average discrepancy of 0.24 minutes (*SD* = 0.90) between N1-sleep onset and behavioural sleep onset on the first night of testing and a discrepancy of 0.82 minutes (*SD* = 1.31) on the second night (see Chapter 3).

There are three possible reasons for this surprising finding with vibratory stimuli. Firstly, participants may have had a lack of awareness of stimuli presented after N1-sleep onset. Secondly, participants may have been marginally aware of stimuli after N1-sleep onset, but the vibrations were not salient enough to arouse the

individual to produce a finger tap response. Thirdly, participants may have been aware of the stimuli, but they could not produce the required behavioural response due to the inhibition of muscle activity associated with N1-sleep onset (Fogel et al., 2005; Mezzanotte, Tangel, & White, 1996).

In this study, EEG was scored in 30-second epochs in accordance with standardised criteria that allows a significant minority of the epoch to be in another state. With such an imprecise determination of sleep/wake state, the participants' immediate state at the moment that a stimulus occurred could not be determined. To overcome this limitation, the current study extends the analysis of EEG data to shed light on the correspondence between responsiveness to minimal intensity vibratory stimuli and sleep microstructure during the sleep onset period. QEEG analysis was conducted on 5-second windows to characterise sleep microstructure through a finer-grained analysis than traditional EEG sleep staging. The aim was to compare sleep microstructure when the participant produced a behavioural response versus when they did not behaviourally respond to minimal intensity vibratory stimuli that were administered around the sleep onset period. In line with previous research into the correspondence between sleep microstructure and behavioural responsiveness (Prerau et al., 2014), we predicted that alpha activity would have attenuated, and lower frequency activity will have intensified when people failed to respond to the vibratory stimulus.

Method

Participants

The study protocol was described previously in Study 2, Chapter 3. Briefly, twenty individuals met eligibility criteria and completed this study. However, two participants were excluded from this analysis due to low quality polysomnography

(PSG) recordings on greater than 25% of at least one of the two laboratory nights of interest (Nights 2 and 3). Therefore, the findings of this study pertain to 18 individuals (see participants characteristics in Table 4-1). These healthy individuals had good or poor sleep (N = 9 for both groups) as defined by scores on the Insomnia Severity Index ([ISI], good sleepers ISI < 7, poor sleepers ISI ≥ 7).

Table 4-1. *Descriptive statistics for participants characteristics.*

Characteristic	**Good sleepers (N = 9)**	**Poor sleepers (N = 9)**	**Total Sample (N = 18)**
Age, mean (SD), y	24.54 (5.47)	22.57 (4.46)	25.56 (4.95)
Sex, No. (%)			
Men	4 (44)	3 (33)	7 (39)
Women	5 (56)	6 (67)	11 (61)
BMI, mean (SD)	24.34 (2.21)	24.13 (4.14)	24.23 (3.22)
Lifestyle characteristics			
Weekly alcohol consumption, No. (SD)	1.56 (1.74)	1.33 (1.94)	1.44 (1.79)
Daily caffeine consumption, No. (SD)	2.00 (1.58)	2.00 (1.50)	2.00 (1.50)
Sleep characteristics			
ISI, mean (SD)	2.22 (0.97)	11.89 (4.04)	7.06 (5.73)
PSQI, mean (SD)	3.22 (1.79)	8.22 (3.27)	5.72 (3.63)
Habitual Bedtime, mean time (SD, min)	22:53 (61.60)	23:16 (53.35)	23:04 (57.37)
Habitual Wake Up Time, mean time (SD, min)	07:27 (64.97)	07:58 (75.92)	07:43 (70.76)
Habitual total sleep time, mean hrs (SD)	7.89 (0.70)	7.00 (1.64)	7.44 (1.30)

BMI = body mass index, ISI = Insomnia Severity Index, N = sample size, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

Design and Study Procedure

This study employed a within-groups quasi-experimental design. To summarise, participants slept in the sleep laboratory on three nights after screening and consent. The current study only concerns data from Night 2 and Night 3. On both nights, participants arrived at the sleep laboratory at approximately 20:00 and were setup for overnight PSG recording. The THIM device was placed on the participant's index finger of their dominant hand along with a piezo-electric sensor. After setting the vibratory stimulus intensity to the lowest perceptible level, participants were taught how to operate the THIM device (see Appendix 5 for this procedure). An hour before participants' typical bedtime, they began THIM-administered sleep onset trials that continued for four hours. During these trials, THIM administered brief (≈ 500ms), minimal intensity vibratory stimuli. Participants were asked to respond to these stimuli by tapping their index finger against their thumb. Qualified research assistants observed participants via video recording and monitored the THIM sensor data in real-time to confirm compliance. After four hours of trials, participants slept uninterrupted until the morning when all devices were removed, and participants returned home.

The predictor variable for the current study was whether participants responded or did not respond to the vibratory stimulus during each sleep stage and the primary outcome variables were percentage change in EEG power (μV^2) from baseline in the delta, theta, alpha, sigma, and beta frequency bands before and after the onset of the THIM vibratory stimulus.

Materials

THIM

As described in Chapter 3, THIM (firmware v1.0.4) is a device worn on the index finger of the dominant hand. To commence a sleep onset trial, participants

tapped their finger onto their thumb twice in quick succession. THIM then emitted low intensity, short duration vibratory stimuli, averaging 30 seconds apart. The intensity of the vibratory stimulus was calibrated to the minimum level that the participant could consistently respond to whilst awake using the threshold hunting procedure in the THIM smartphone application. Participants responded to the vibratory stimulus by tapping their index finger once onto their thumb. If participants failed to respond to two consecutive stimuli, the device inferred that sleep onset had occurred and subsequently emitted a high intensity alarm vibration to wake them up, signalling the end of the trial. Shortly afterwards (approximately 1-2 minutes later), participants initiated another trial.

From a small piezo-electric sensor attached to the side of THIM, four events of interest were manually scored: vibrations emitted from THIM, finger taps as responses to the vibrations, the beginning (the double-tap motion) and the end (the high-intensity alarm vibration) of each trial. If the sensor data were obscured by body movements, the trial was removed from analysis. Agreement between the two scorers (HS and AW) exceeded 95% on 10 randomly selected nights of data.

Polysomnography

PSG was recorded using Compumedics Grael 4K PSG:EEG devices (Compumedics, Victoria, Australia). Six EEG sites (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), reference and ground, and right and left electrooculography (EOG) were sampled at 512Hz, whilst chin electromyography (EMG), and electrocardiography (ECG) sites were sampled at 256Hz. PSG data was scored in accordance with AASM scoring criteria (American Academy of Sleep Medicine, 2018) using Profusion Compumedics software (v 4.0) by a qualified, independent sleep technician blind to the THIM sensor data.

Quantitative EEG Analysis

All PSG data was exported into European Data Format (EDF) for analysis along with scored sleep stage files using Compumedics Profusion software (v 4.0). This algorithm applied a Fast-Fourier Transform FFT using multitaper-based estimation (Prerau, Brown, Bianchi, Ellenbogen, & Purdon, 2017), on non-overlapping 5-second epochs on the C3-M1 signal. These epochs began 15 seconds before the onset of the stimulus, with zero representing stimulus onset (-15 seconds, -10 seconds, -5 seconds, 0 seconds, 5 seconds, 10 seconds and 15 seconds). Absolute power in five frequency bands – delta (0.5-4.5Hz), theta (4.5-8Hz), alpha (8-12Hz), sigma (12-15Hz), and beta (15-32Hz) – were calculated in the five-second epochs.

The algorithm automatically identified noisy epochs using previously validated methods (D'Rozario et al., 2015), which were subsequently removed from analysis. See Table 4-2 for the number and percentage of epochs removed from analysis for each sleep stage. With the remaining epochs, the power in each frequency band at baseline (epochs before stimulus onset) was averaged across every trial, separately for each individual. The absolute power in each frequency band during the 5-second epochs were then divided by this baseline for each individual to evaluate the change in power from baseline. This step standardised the power density across individuals. Next, the average spectral power in the frequency bands during each five-second epoch was calculated separately for wake, N1 and N2 sleep epochs scored using AASM criteria for analysis.

Table 4-2. *Sum and percentage of noisy epochs removed from analysis.*

Nights	Sleep stage, no. (%)			Total
	Wake	N1	N2	
Night 2	145 (3.78)	10 (1.12)	11 (1.22)	166 (2.95)
Night 3	162 (3.80)	20 (2.09)	11 (1.18)	193 (3.14)
Total	307 (3.79)	30 (1.62)	22 (1.20)	359 (3.05)

N1 = Non-rapid eye movement Stage 1, N2 = Non-rapid eye movement Stage 2.

Data Analysis

Linear Mixed Modelling (LMM) were conducted to investigate the spectral power in each frequency band before and after the onset of the THIM vibratory stimuli when participants responded versus when they did not respond to the stimulus. LMM analyses used a first-order autoregressive covariance structure. Fixed factors were the laboratory night (Night 2 and Night 3), sleep stage (wake, N1 or N2), time (the five-second epochs), and whether participants behaviourally responded or not to the stimulus. Participant ID was identified as the random intercept. Spectral power data was log10 transformed to meet assumptions of normality. Where appropriate, post hoc comparisons were conducted with the Bonferroni correction and Cohen's d was calculated and interpreted according to standard criteria (Cohen, 2013). All analyses were conducted in IBM SPSS (v 23).

A repeated measures Analysis of Variance (ANOVA) was conducted to determine whether the proportion of stimuli that participants responded to significantly differed across sleep stages. To test this, the proportion of stimuli that participants responded to out of all stimuli presentations in that sleep stage were calculated separately for each participant and averaged together.

Results

The following findings are based on 11,777 stimuli presentations across both

laboratory nights. See Table 4-3 for the number of stimuli presented for each participant, separated by sleep stage and whether they did or did not respond to the stimulus.

Table 4-3. *Number of stimuli that participants responded and did not respond to during each sleep stage across both laboratory nights.*

Participant	Responded to stimulus				Did not respond to stimulus			
	Wake	N1	N2	Total	Wake	N1	N2	Total
1	633	37	0	670	64	40	0	104
2	387	88	42	517	51	62	70	183
3	388	47	45	480	14	26	63	103
4	538	19	6	563	8	18	5	31
5	532	47	28	607	32	43	37	112
6	390	35	19	444	27	42	81	150
7	392	60	17	469	39	89	81	209
8	480	118	38	636	29	46	33	108
9	306	47	27	380	51	87	113	251
10	354	50	101	505	16	19	69	104
11	452	64	37	553	38	77	52	167
12	329	78	101	508	32	31	99	162
13	545	49	53	647	31	29	88	148
14	521	53	46	620	14	16	63	93
15	220	75	108	403	35	37	91	163
16	178	19	2	199	111	74	44	229
17	656	73	10	739	18	22	21	61
18	119	41	47	207	69	91	92	252
Total	**7420**	**1000**	**727**	**9147**	**679**	**849**	**1102**	**2630**
	91.62%	54.08%	39.75%		8.38%	45.92%	60.25%	

N1 = Non-rapid eye movement Stage 1, N2 = Non-rapid eye movement Stage 2. Figures represent data from both laboratory nights combined as the four-way interactions were not significant.

Sleep microstructure and behavioural responsiveness

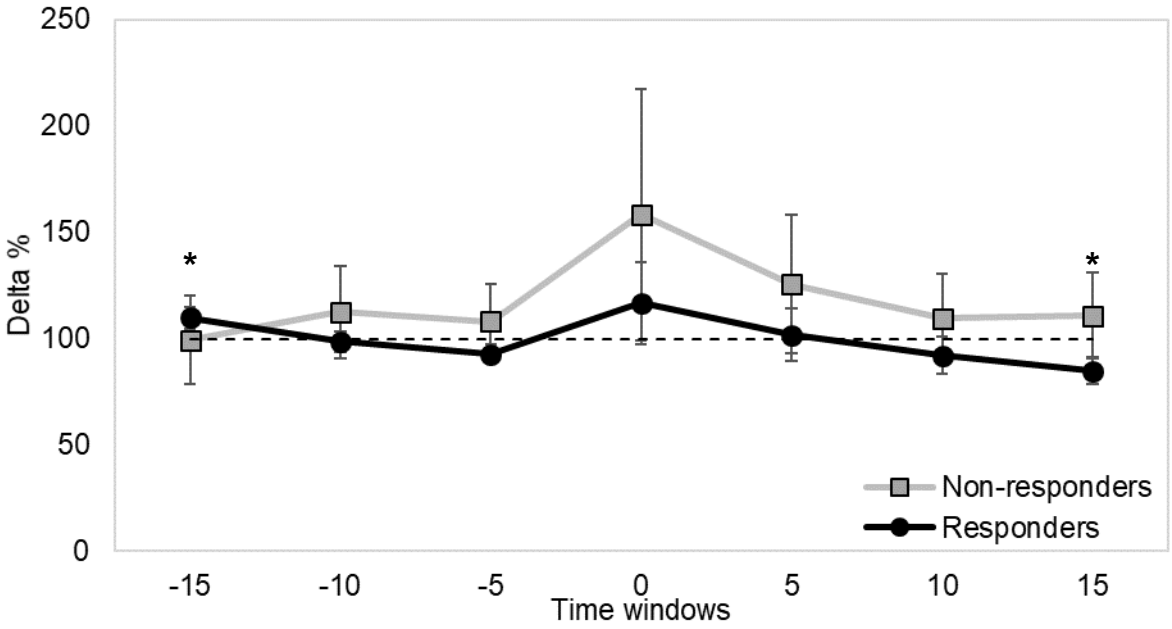
Four-way interactions between laboratory night, sleep stage, time, and behavioural responsiveness to the stimuli were investigated with LMMs, separately on each frequency band. The four-way interactions were not significant across any frequency band, $p > .26$. The three-way interactions of time, sleep stage and responsiveness to the stimulus were significant on delta, $p = .005$, alpha, $p < .001$, sigma, $p < .001$, and beta, $p < .001$, but not significant on theta, $p = .32$. For completeness, the results pertaining to the theta band will be presented, but not discussed in detail. These three-way interactions indicate that, collapsed across both nights, the interaction between time (5-second epochs) and behavioural responses versus non-responses to vibratory stimuli on the power in these frequency bands significantly differs by sleep stage. As such, the sleep microstructure when participants responded versus did not respond to vibratory stimuli are discussed separately for each sleep stage below.

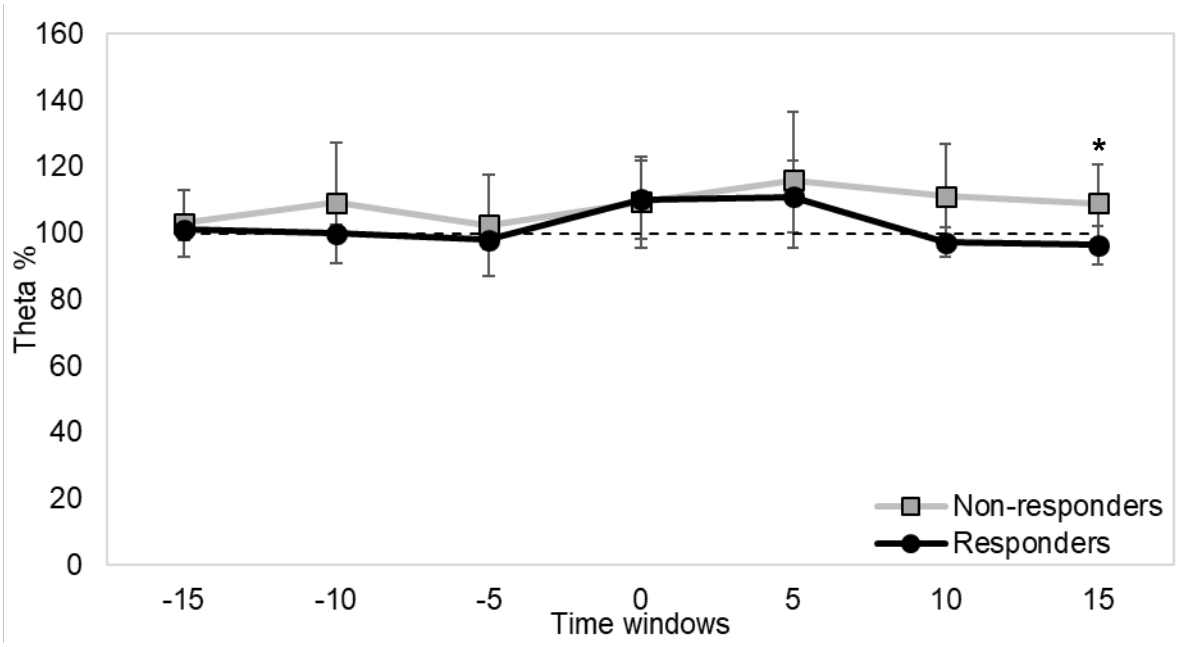
Responsiveness to Stimuli during Wake

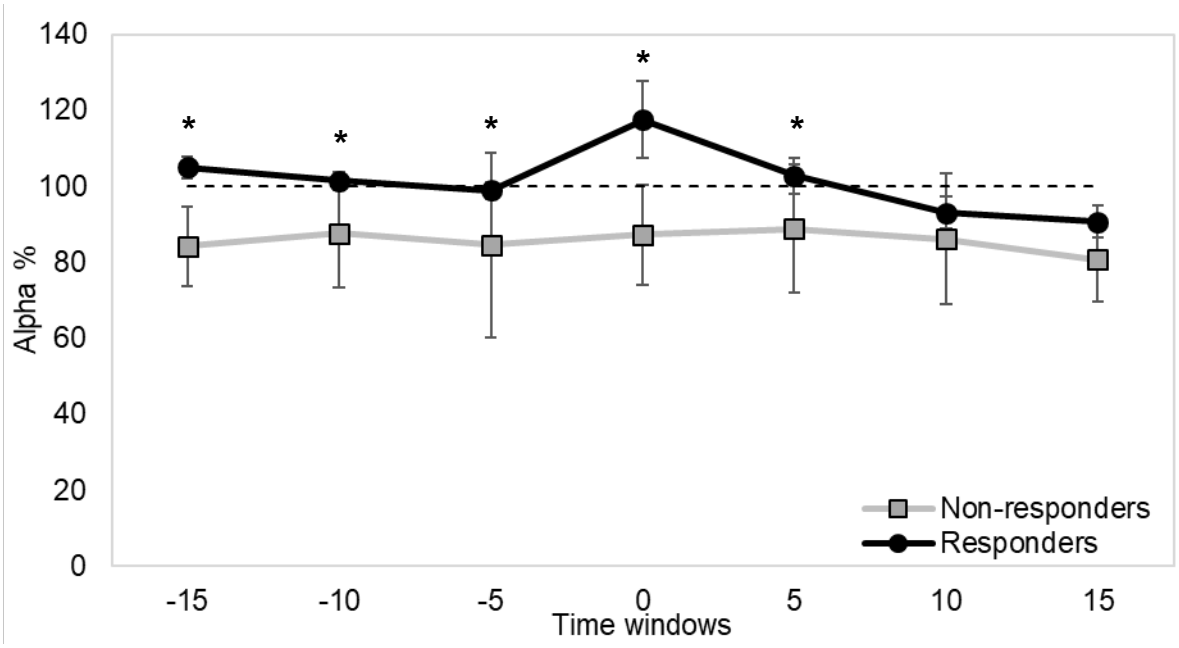
Figure 4-1 below shows the differences in power in each frequency band when participants responded versus when they did not respond to the vibratory stimulus during 30-seconds epochs scored as wake. Pairwise comparisons indicated that participants had significantly higher alpha power in all three epochs before stimulus onset when they responded to the stimulus versus when they did not respond, $p < .01$, $d > 0.38$. This would suggest that participants were 'more awake' when they responded to the stimulus compared to when they did not respond during wakefulness. Participants also had significantly higher delta, $p = .01$, $d = 0.31$, and sigma power, $p = .04$, $d = 0.14$, in the 15-second epoch before stimulus onset when they responded to the stimulus compared to when they did not respond during wake.

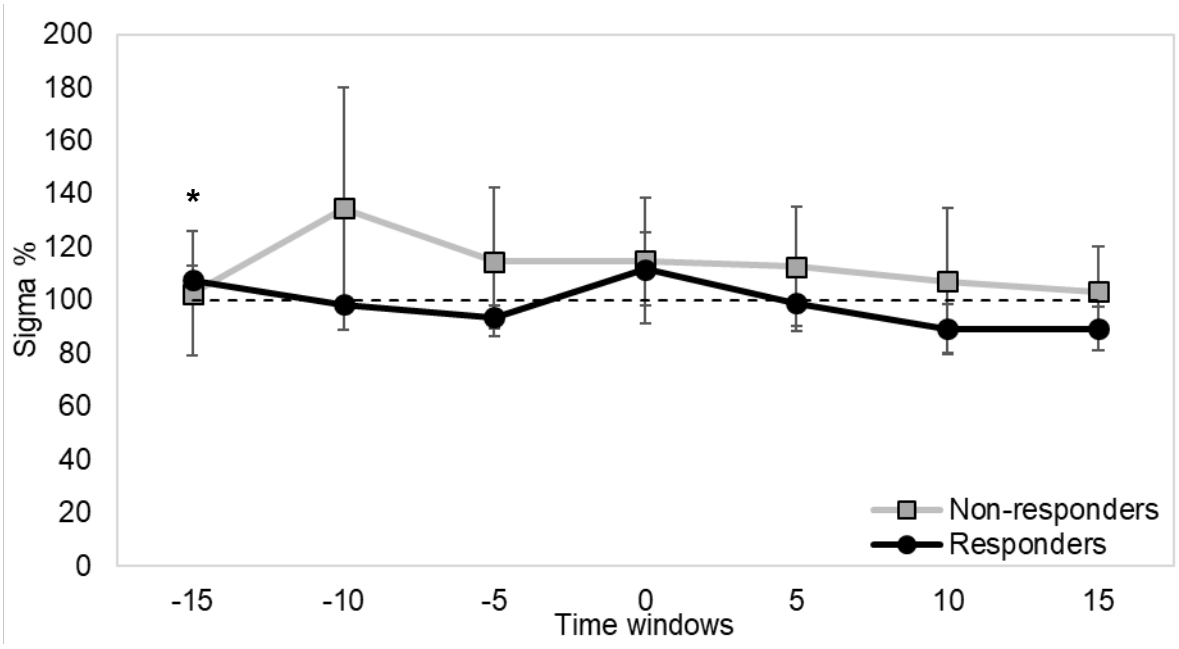
However, these differences were not significant in the 10-second and 5-second epochs before stimulus onset.

At stimulus onset (0-second epoch), there was greater alpha frequency power when participants responded to the stimulus compared to occasions where they did not respond, $p < .001$, $d = 1.20$. This increase in power persisted into the 5-second epoch, $p = .006$, $d = 0.52$, but was not significantly different at the 10-second epoch after stimulus onset, $p = .08$. At 15 seconds after stimulus onset, there was significantly greater delta, $p = .004$, $d = 0.81$, and theta power, $p = .04$, $d = 0.63$, on occasions where participants did not respond compared to occasions where they did respond to stimuli during the 30-second epochs scored as wake. Participants may have begun to enter sleep on the occasions where they did not respond to the vibratory stimulus during the 30-second "wake" epochs, suggesting that their lack of behavioural and physiological response to the stimulus may have been due to acutely reduced wakefulness.




160
140
120
100
80
60
40
20
0
Theta %
*
Non-responders
Responders
-15
-10
-5
0
5
10
15
Time windows


140
120
100
80
60
40
20
0
Alpha %
*
*
*
*
*
Non-responders
Responders
-15
-10
-5
0
5
10
15
Time windows



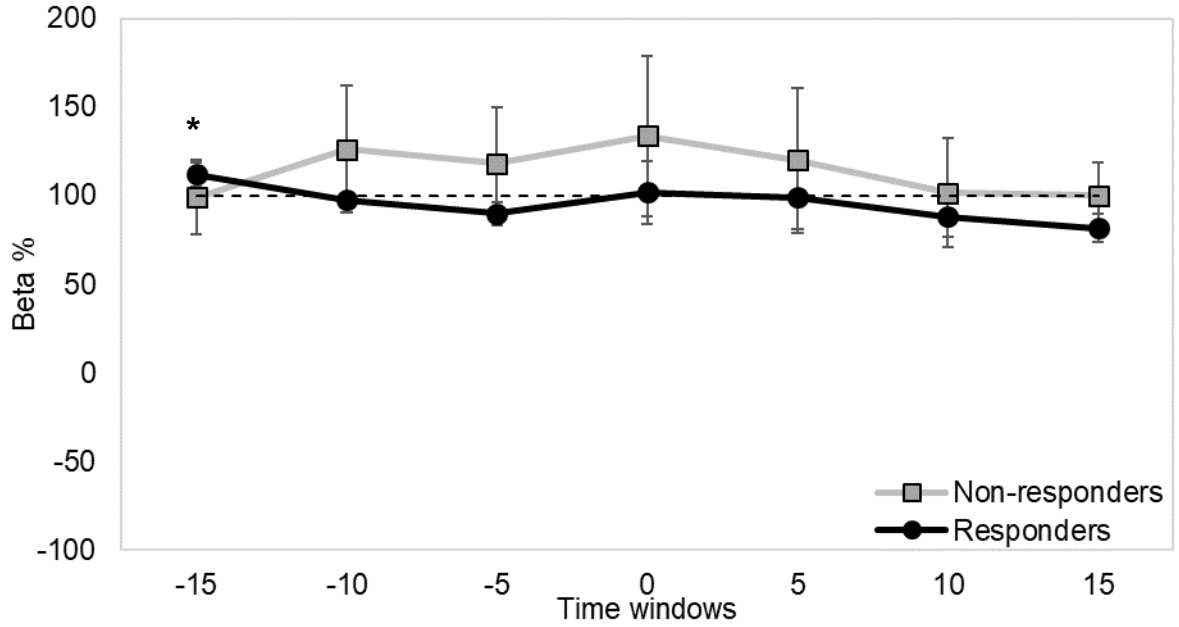


Figure 4-1. Power in delta, theta, alpha, sigma and beta frequency bands when participants responded (black line) versus when they did not respond (grey line) to the vibratory stimulus during wake.

* indicates $p < .05$. Bars indicate 95% confidence intervals. The dotted line indicates baseline power. Figures represent data from both laboratory nights combined as the four-way interactions were not significant.

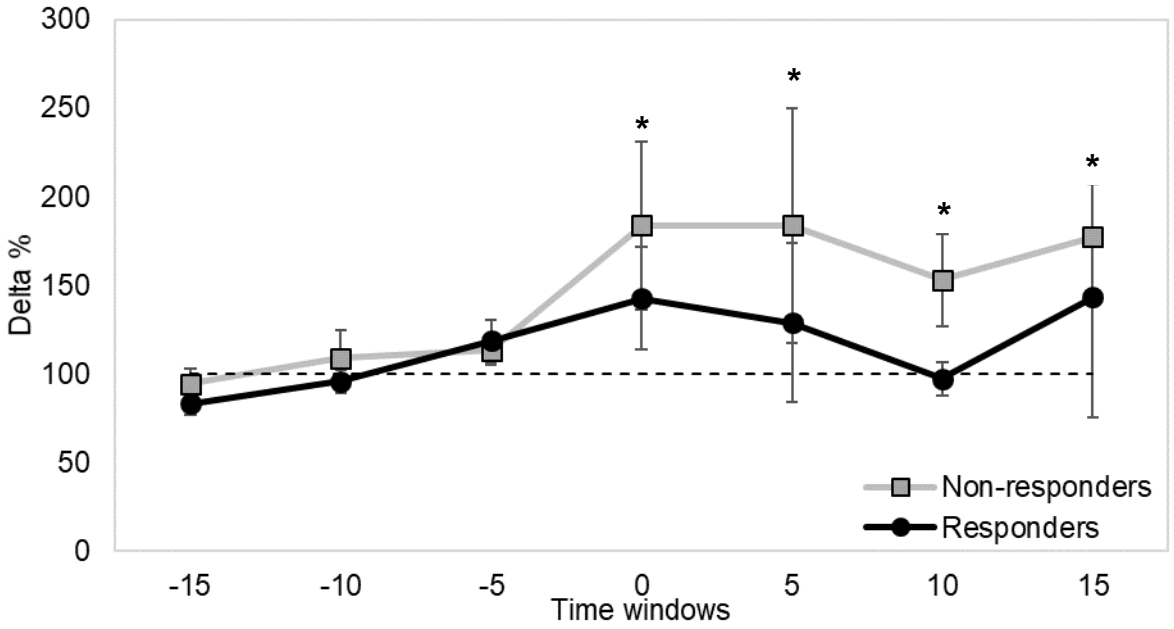
Responsiveness to Stimuli during N1 sleep

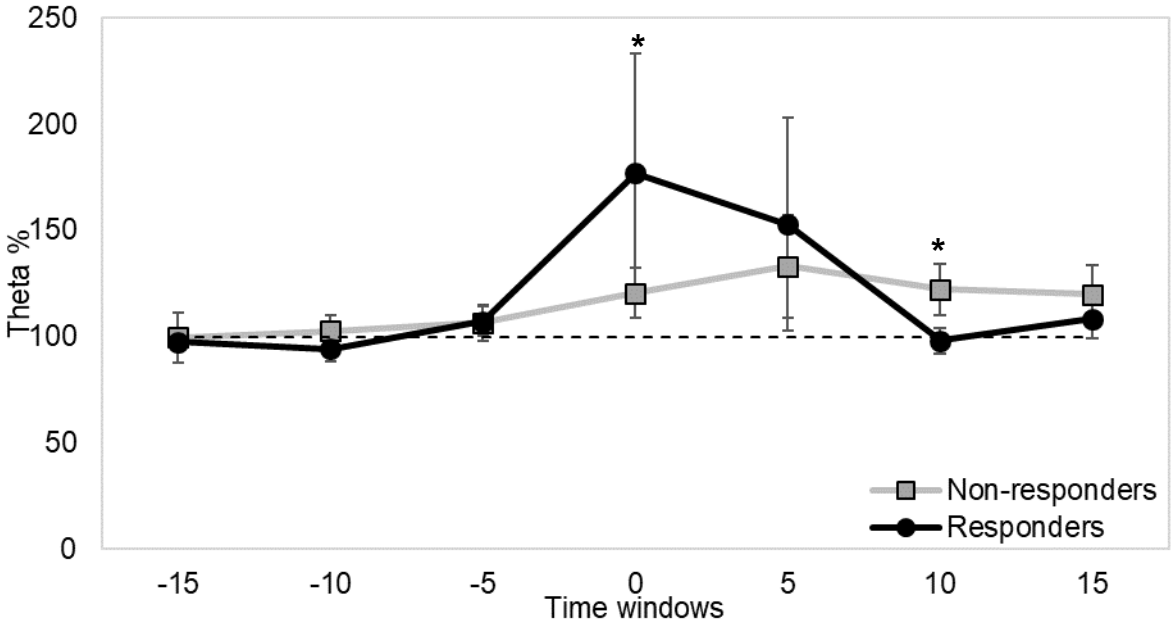
Figure 4-3 below shows the differences in power in each frequency band when participants responded versus when they did not respond to the vibratory stimulus during N1 sleep. Pairwise comparisons indicated that participants had significantly higher alpha, $p < .001$, $d = 2.17$, and beta power, $p < .001$, $d = 2.33$, in the five seconds before stimulus onset on occasions when they responded to the stimulus compared to when they did not respond during N1 sleep. This suggests that participants were 'more awake' before the stimulus occurred when they responded to the stimulus versus when they did not respond in N1 sleep.

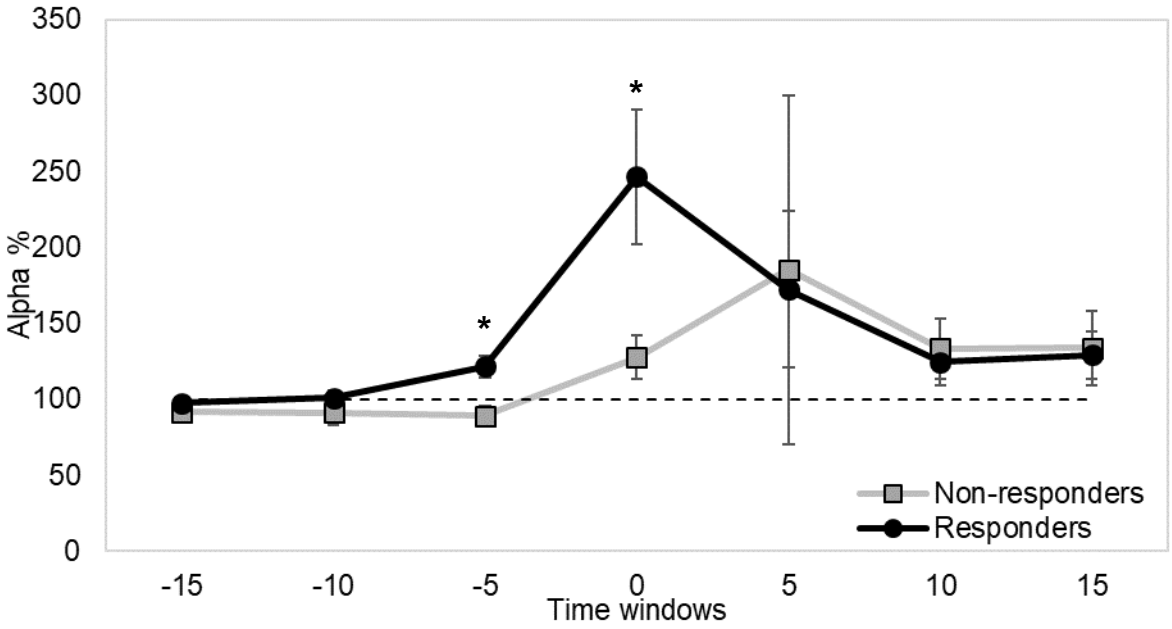
At stimulus onset, participants had higher theta, $p = .002$, $d = 0.64$, alpha, $p < .001$, $d = 1.67$, sigma, $p < .001$, $d = 1.52$, and beta power, $p < .001$, $d = 2.25$, when they responded to the stimulus compared to when they did not respond during epochs scored as N1 sleep. This increase suggests that a shift to wakefulness or an arousal/movement occurred at or in the five seconds before stimulus onset. At 5, 10, and 15 seconds after stimulus onset, participants had significantly lower sigma power when they responded to the stimulus compared to when they did not respond, but this was a small effect, $p < .003$, $d > 0.32$. Participants also had significantly lower theta, $p = .002$, $d = 1.18$, and beta power, $p = .02$, $d = 0.31$, at the 10-second epoch after stimulus onset when they responded to the stimulus compared to occasions where they did not respond. There were no significant differences in alpha power at 5, 10, and 15 seconds after stimulus onset when participants responded versus when they did not respond to the stimulus, $p > .08$.

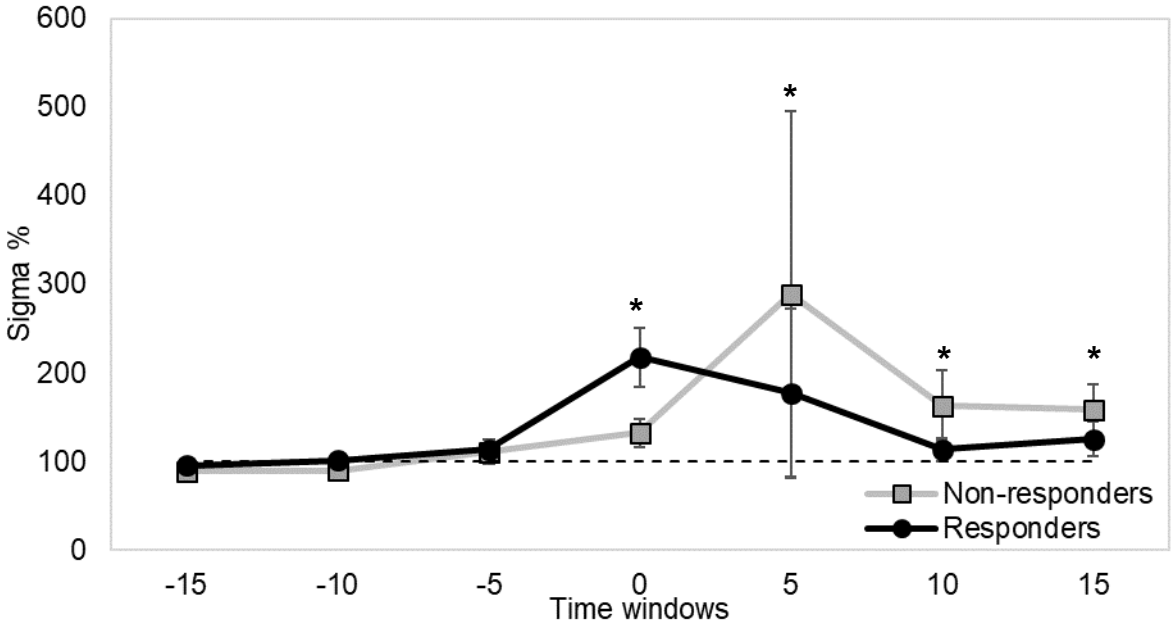
Participants had higher delta power, $p = .002$, $d = 0.48$, when they did not respond to the stimulus versus when they did respond during epochs scored as N1 sleep. This higher power in delta continued in the 5-, 10-, and 15-second epochs after stimulus onset, $p < .001$, $d > 0.30$. As in N2 sleep, this pattern suggests that

participants did not respond to the stimulus in N1 sleep because they were in deeper sleep, or at least at a lower level of arousal.






350
300
250
200
150
100
50
0
Alpha %
*
*
-15
-10
-5
0
5
10
15
Time windows
Non-responders
Responders


600
500
400
300
200
100
0
Sigma %
*
*
*
*
-15
-10
-5
0
5
10
15
Time windows
Non-responders
Responders

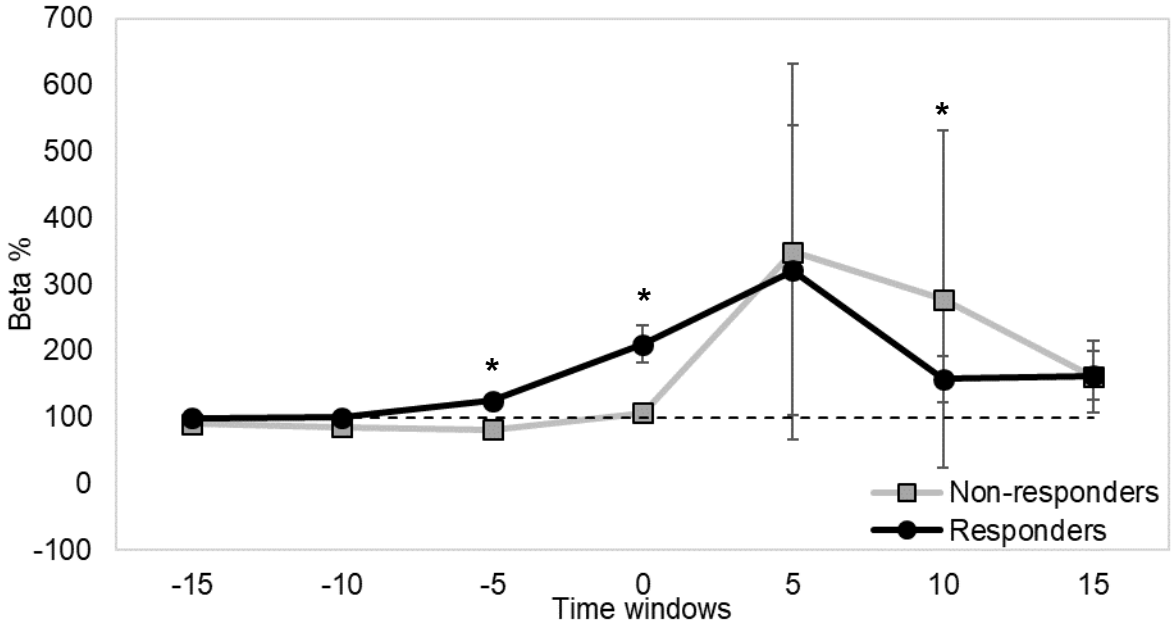


Figure 4-3. Power in delta, theta, alpha, sigma and beta frequency bands when participants responded (black line) versus when they did not respond (grey line) to the vibratory stimulus during N1 sleep.
* indicates p < .05. Bars indicate 95% confidence intervals. The dotted line indicates baseline power. Figures represent data from both laboratory nights combined as the four-way interactions were not significant.

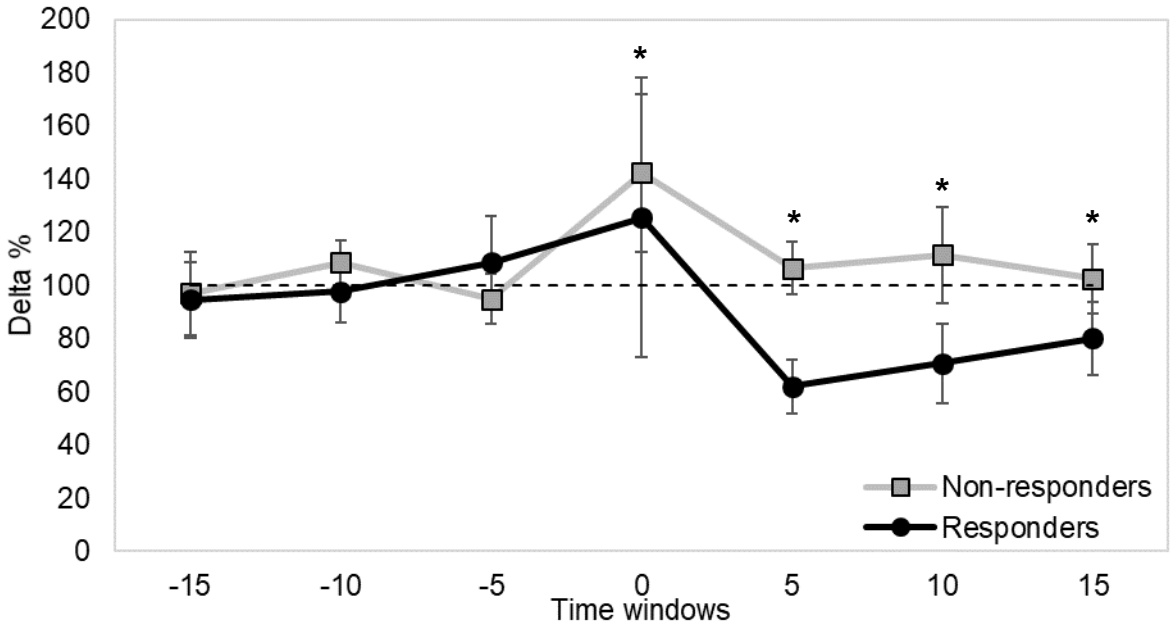
Responsiveness to Stimuli during N2 sleep

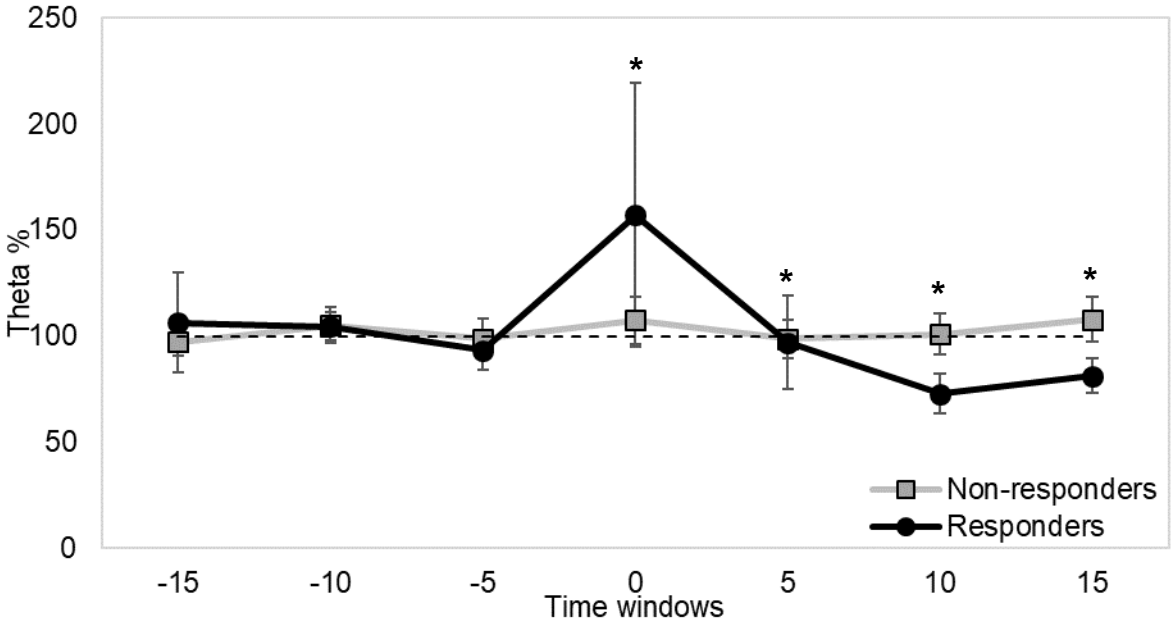
Figure 4-2 below shows the differences in power in each frequency band when participants responded versus when they did not respond to the vibratory stimulus during epochs scored as N2 sleep. Pairwise comparisons indicated no significant differences between responses versus non-responses to vibratory stimuli in power in the epochs before stimulus onset (-15, -10 and -5 seconds), $p > .06$. The exception was higher alpha power in the 5-second epoch before stimulus onset when participants responded to the stimulus versus when they did not respond during N2 sleep, $p = .006$, $d = 1.14$.

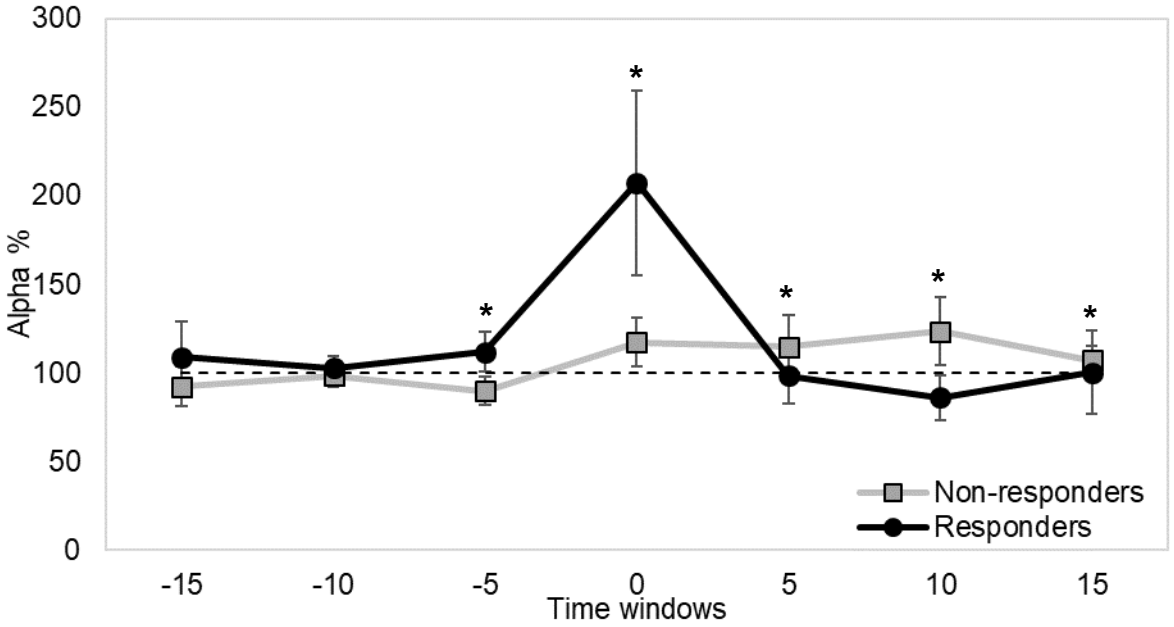
At stimulus onset during N2 sleep, when participants responded to stimuli they had higher theta, $p = .009$, $d = 0.55$, alpha, $p < .001$, $d = 1.16$, sigma, $p = .005$, $d = 1.02$, and beta power, $p < .001$, $d = 1.22$, compared to when they did not respond. As in N1 sleep, this increase in higher frequencies suggests a shift to wake

or an arousal/movement coinciding with the presentation of the stimulus and response. This pattern reverses at 5, 10 and 15-seconds after stimulus onset. Participants had significantly lower theta, $p < .02$, $d > 0.05$, alpha, $p < .03$, $d > 0.20$, sigma, $p < .001$, $d > 0.53$, and beta power, $p < .002$, $d > 0.20$, during these epochs when they responded to the stimulus compared to when they did not respond. However, these effect sizes were relatively small compared to those found for changes immediately before and at the onset of the stimulus.

At stimulus onset, participants had higher delta power when they did not respond to the stimulus versus when they did respond, $p = .004$, $d = 0.19$. This pattern continued through the 5-second, $p < .001$, $d = 2.18$, 10-second, $p < .001$, $d = 1.21$, and 15-second epochs after stimulus onset, $p = .001$, $d = 0.81$. This finding would suggest that participants sustained sleep after the stimulus was presented, which is consistent with their failure to respond.




250
200
150
100
50
0
Theta %
*
*
*
*
Non-responders
Responders
-15
-10
-5
0
5
10
15
Time windows


300
250
200
150
100
50
0
Alpha %
*
*
*
*
Non-responders
Responders
-15
-10
-5
0
5
10
15
Time windows

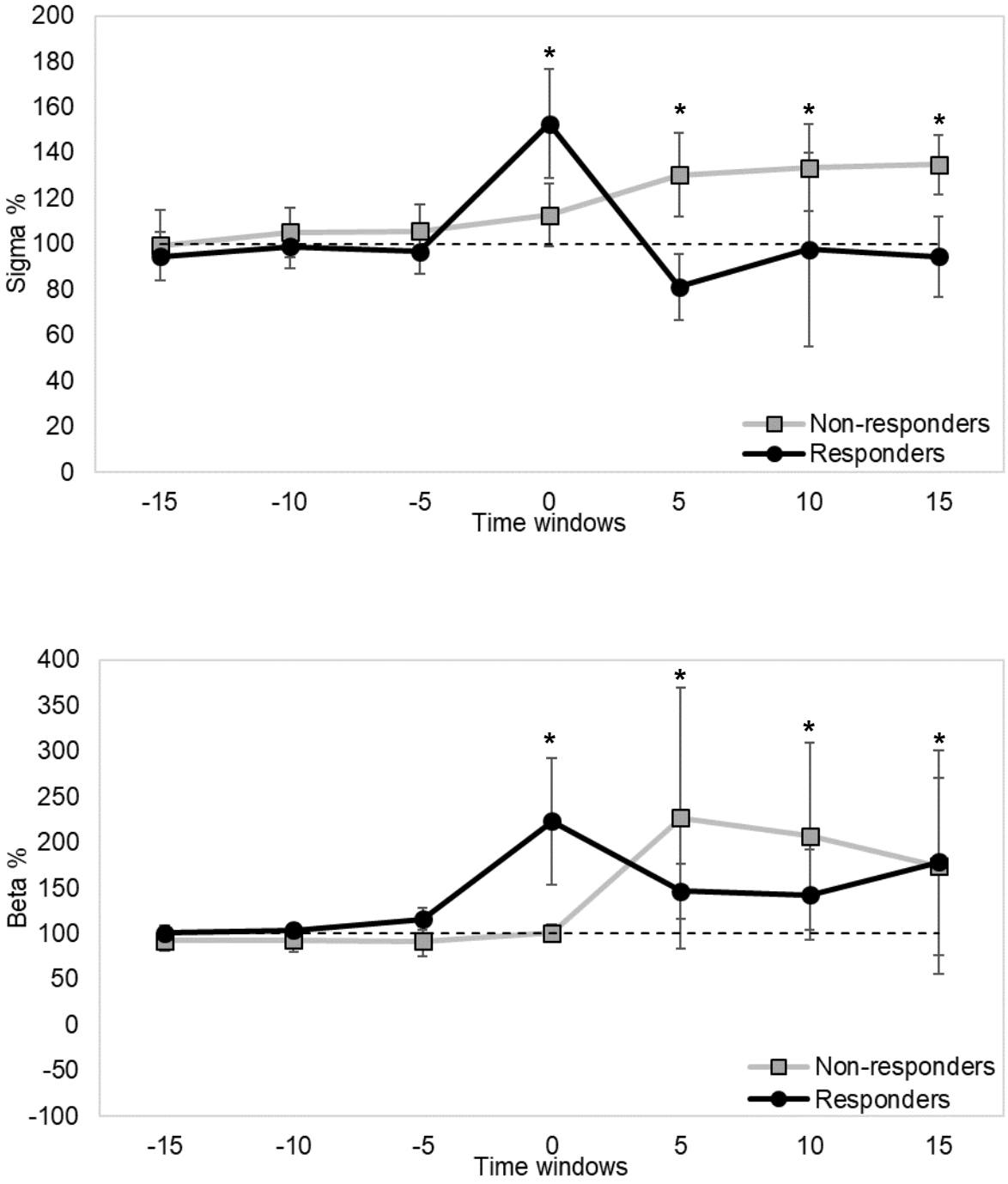


Figure 4-2. Power in delta, theta, alpha, sigma and beta frequency bands when participants responded (black line) versus when they did not respond (grey line) to the vibratory stimulus during N2 sleep.
* indicates p < .05. Bars indicate 95% confidence intervals. The dotted line indicates baseline power. Figures represent data from both laboratory nights combined as the four-way interactions were not significant.

Comparison of Behavioural Responsiveness between Sleep Stages

A repeated measures ANOVA revealed that mean proportion of responses to vibratory stimuli differed significantly between sleep stages, $F(2, 32) = 149.11$, $p <$

.001. Post hoc tests using the Bonferroni correction indicated that the mean proportion of responses to stimuli was significantly higher in wake (M =0.90, SD = 0.11) compared to the mean proportion of responses in N1-sleep (M =0.56, SD = 0.17), $p < .001$, $d = 2.38$, and N2-sleep (M =0.38, SD = 0.15), $p < .001$, $d = 3.95$. The mean proportion of responses was also significantly higher in N1-sleep compared to N2-sleep, $p < .001$, $d = 1.12$.

Discussion

This study was the first to investigate the correspondence between sleep microstructure and behavioural responsiveness to minimal intensity vibratory stimuli presented during the sleep onset period. During epochs scored as wake, participants behaviourally responded to the majority of stimuli presentations. This was coupled with increased alpha before and immediately after stimulus onset compared to when participants did not respond to the stimulus during epochs scored as wake. Fifteen seconds after stimulus onset, greater delta and theta power was observed when participants did not respond to the stimulus. Together these findings suggest that their lack of response was due to participants being closer to initiating sleep on the few occasions where they did not respond to stimuli during epochs scored as wake compared to occasions when they did respond.

Participants responded to much less stimuli during N2. The difference between whether participants behaviourally responded to the vibratory stimulus or not may be whether they had increases in alpha and higher frequencies at stimulus onset. A response to the stimulus was typically precipitated by signs of wakefulness before stimulus onset during N2 sleep. Therefore, participants appeared to be ‘more awake’ before stimulus presentation. The increase in theta and sigma power may have been due to increases in alpha brainwaves that were captured in these

frequency bands instead of the alpha band by the quantitative electroencephalography (qEEG) analysis. If this shift to higher frequencies did not occur (as was the case for the majority of stimuli presentations during N2 sleep), participants did not respond to the stimulus. Instead, they experienced higher delta activity. This may indicate the presence of a k-complex or other phenomenon, but our analysis is unable to confirm this and would require greater precision of measurement to test appropriately than was obtained in the current study. Thus, a behavioural response to a vibratory stimulus during wake or N2 sleep coincided with wake-like brain qEEG and a lack of response to the stimulus coincided with more sleep-like qEEG.

The findings for N1 sleep support the conceptualisation of this sleep stage as a transitional period between sleep and wake (De Gennaro, Ferrara, & Bertini, 2001; Gorgoni, D'Atri, Scarpelli, Ferrara, & De Gennaro, 2019; Ogilvie, 2001; Prerau et al., 2014). During N1 sleep, participants showed a similar pattern in the frequency bands compared to N2 sleep. If participants did not respond to the stimulus, they exhibited higher delta activity, signalling that they were 'more asleep', which may be why they did not respond to the stimulus. If participants responded to the stimulus, they had increases in higher frequencies immediately before and after stimulus onset. This signals that they were 'more awake', which may explain why they responded to the stimulus. Yet, the repeated measures ANOVA revealed that this occurred much more frequently in epochs scored as N1 sleep compared to those scored as N2 sleep. Together, these findings illustrate that participants regularly exhibited both wake and sleep-like physiological (power spectral analyses) and behavioural responsiveness to vibratory stimuli during N1 sleep. This adds to the body of evidence suggesting that N1 sleep is dynamic and transitional between - and

exhibits signs of both – wake and sleep.

This study has implications for the use of behavioural responsiveness to minimal intensity vibratory stimuli to estimate sleep onset. When participants responded to the stimulus during sleep, their brain activity indicated the occurrence of an arousal, suggesting that the arousal precipitating the stimulus was necessary for the individual to respond. A criticism of using responses to external stimuli to estimate sleep onset is that the presentation of stimuli may disrupt the process of falling asleep and prolong wakefulness (Casagrande, De Gennaro, Violani, Braibanti, & Bertini, 1997; Ogilvie, 2001). Yet, these findings indicate that disruption to minimal intensity vibratory stimuli rarely occurs once participants enter sustained sleep (N2 sleep), and participants that responded to the stimulus had typically experienced a brief arousal before the stimulus presentation. Furthermore, when participants responded to the stimulus during 'sleep', their frequency power largely returned to baseline by the 10-second or 15-second epoch after stimulus onset. This indicates that, if the stimulus arouses the individual, their sleep resumes within a short period of time. It is therefore unlikely that the presentation of minimal intensity vibratory stimuli would significantly disrupt the process of falling asleep.

Our previous analyses found strong correspondence between PSG-sleep onset and behavioural responsiveness to THIM (see Chapter 3). The current study suggests that these findings may have been due to participants being unaware of the stimulus after PSG-sleep onset. The current study also supports the findings that behavioural responses to vibratory stimuli are closely associated with EEG activity. In this study, behavioural responses occurred during brief arousals in N1- and N2-sleep that had been overlooked when the EEG was scored according to the AASM scoring criteria (Hertig-Godeschalk et al., 2019). Therefore, THIM detected signs of

wakefulness during these brief arousals that the current PSG scoring criteria does not, rendering the device highly accurate for estimating wakefulness/brief arousals from sleep.

Conclusion

This article was the first to examine responsiveness to vibratory stimuli and sleep microstructure during the sleep onset period. Across all sleep stages, responses to vibratory stimuli corresponded with increases in higher frequencies indicative of wakefulness. Whereas, a lack of response to vibratory stimuli coincided with higher delta activity, signalling greater sleep depth. The difference between states was the percentage of response to stimuli: responses were frequent in wake, common in N1 sleep, and uncommon in N2-sleep. Both wake and sleep-like behavioural and brain responses regularly occurred in N1-sleep. This further supports the theory that N1 sleep is a fluctuating and transitional state between wake and sustained sleep.

Chapter 5: The Development and Accuracy of the THIM Wearable Device for Monitoring Sleep and Wakefulness.

Hannah Scott[1], Nicole Lovato[2] and Leon Lack[1,2]

[1] College of Education, Psychology and Social Work, Flinders University.

[2] College of Medicine and Public Health, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, Flinders University.

Author contributions: HS contributed to study design, data collection, analysis and interpretation of the data, and manuscript preparation. LL and NL contributed to conception and study design, interpretation of the data and drafting of the manuscript.

Conflict of interests: Research costs were partially funded by Re-Time Pty Ltd, the company that sell THIM. None of the study authors were financially supported by Re-Time for this project. LL is a shareholder of Re-Time Pty Ltd. LL and HS have a patent pending regarding the THIM device. NL has no conflicts of interest to declare.

Citation: Scott, H., Lovato, N., & Lack, L. (2020). The development and accuracy of the THIM wearable device for estimating sleep and wakefulness. *under review*.

Abstract

THIM is a new wearable device worn on the finger that can passively monitor sleep and wakefulness overnight using actigraphy. This article showcases the development of the THIM sleep tracking algorithm (Study 1), and the test of its accuracy against polysomnography (PSG) with a small independent sample of good and poor sleepers (Study 2). The accuracy of THIM was also compared to two popular actigraphy devices, Fitbit and Actiwatch devices.

Twenty-five (Study 1) and twenty (Study 2) individuals slept overnight in the sleep laboratory on one night. Participants slept from their typical bedtime to their typical wake up time with simultaneous recording from PSG and the THIM, Fitbit and Actiwatch actigraphy devices.

In both studies, THIM had lower sensitivity (M = 0.89, SD = 0.06 in Study 2) compared to the Actiwatch (M = 0.95, SD = 0.04) and Fitbit devices (M = 0.96, SD = 0.04), $p < .001$, $d > 1.18$, yet THIM had higher specificity (M = 0.59, SD = 0.18). There were no significant differences between PSG and THIM in either study for sleep onset latency, total sleep time, wake after sleep onset, or sleep efficiency, $p > .06$. Yet, there was high variability in the accuracy of all three actigraphy devices between individuals that was not explained by sleep quality.

Together, these studies suggest that THIM is capable of accurately monitoring sleep and wake overnight in good and poor sleepers. Future research will examine the accuracy of THIM for monitoring sleep in people with insomnia.

Keywords: wearable technology; consumer sleep technology; polysomnography; actigraphy; sleep parameters; validation.

There are many uses for a sleep wearable device that can accurately estimate objective sleep in the home environment. For researchers, an accurate sleep tracker would enable research studies that are currently resource-heavy to be conducted more practically, including observational studies to explore sleep health with big data (Kuula et al., 2019; Ong, Tandi, Patanaik, Lo, & Chee, 2019). For clinicians, an accurate sleep tracker may represent a substantially cheaper alternative to PSG for the monitoring of certain sleep disorders. For consumers, an accurate sleep tracker may allow individuals to monitor their sleep and benefit from individualised programs that incorporate their sleep tracker data and make recommendations to improve their sleep health. For individuals with insomnia, an accurate sleep tracker could assess the degree to which patients adhere with prescribed time in bed in the case of monitoring adherence to behavioural treatments. It is imperative that sleep trackers are accurate to ensure that any decisions made based on this data are appropriate, which is of even greater importance for diagnostic and treatment-related purposes. This article describes the development and accuracy of a new consumer sleep tracker, called THIM, for estimating sleep and wakefulness while in bed across the intended sleep period.

THIM passively estimates sleep and wakefulness using actigraphy (Re-Time, 2016), which is a method employed in many research and consumer sleep trackers. These typically wrist-worn devices contain an in-built accelerometer that quantifies wrist movement (Sadeh & Acebo, 2002). Individuals tend to remain relatively immobile when sleeping and move their limbs to a greater extent when awake. However, individuals also tend to lie still in bed whilst awake, particularly when they are close to initiating sleep (Pollak et al., 2001). It is therefore unsurprising that these devices tend to overestimate sleep and underestimate wakefulness in most

individuals (Van den Water et al., 2011). This can be particularly true for individuals that typically spend considerable durations of time in bed awake but inactive, such as those with insomnia (Silvertsen et al., 2006; Vallières & Morin, 2003). For this reason, the accuracy of THIM will be examined with both good and poor sleepers to identify significant differences that may exist between groups.

Despite their limitations, actigraphy devices have their advantages. Firstly, these devices are simple to operate – the individual simply wears the device whilst in bed. Secondly, the devices are relatively non-intrusive and therefore unlikely to alter sleep parameters by their use, whereas PSG is known to produce a disturbance of sleep on at least the first night of use (Agnew, Webb, & Williams, 1966). Thirdly, the devices can be used over multiple nights in the home environment and, fourthly, without the need for trained individuals to setup the device or manually score the data. Fifthly, they are less expensive than PSG devices. Therefore, actigraphy devices are a more practical alternative to PSG for objective sleep monitoring. Improving the accuracy of these devices would minimise their only disadvantage and make them more suitable for a variety of research and clinical purposes.

Whilst the actigraphy method does have limited sensitivity for estimating wakefulness, placing the device in a different body location to the traditional wrist placement may improve accuracy. The wrist was selected when actigraphy was developed in the 1980s because it could accommodate the relatively bulky devices at the time. Once algorithms were developed for scoring sleep from wrist movements, the wrist location perpetuated despite recent technology allowing miniaturisation of actigraphy to a smaller location, such as the finger. Wrist actigraphy devices typically only detect significant body movements involving the forearm. An actigraphy device placed on the finger may be able to detect much

smaller movements from the hand and finger that occur during wakefulness and light stages of sleep, such as finger twitches (Reiter, Roach, Sargent, & Lack, 2020). THIM differs from most common sleep trackers because it is worn on the index finger as opposed to the wrist and may consequently be more accurate than wrist actigraphy devices. Research with similar finger-worn actigraphy device have found promising results for estimating sleep and wake (de Zambotti, Rosas, Colrain, & Baker, 2017). We therefore anticipate that THIM will be more accurate than wrist-based actigraphy devices.

Furthermore, the primary function of THIM is to treat insomnia by administering a brief but effective behavioural treatment called Intensive Sleep Retraining ([ISR], Harris et al., 2012; Lack et al., 2019). It would be advantageous to use one simple device to administer a treatment for insomnia and track sleep to tailor treatment instructions and monitor sleep improvements over time. If THIM was more accurate than current actigraphy devices, it may be useful for not only treating and monitoring sleep but also for obtaining a better representation of sleep in the home environment (Withrow, Roth, Koshorek, & Roehrs, 2019).

This article described the development and accuracy of the THIM device for tracking sleep and wakefulness overnight compared to PSG. There will be two studies presented. Study 1 aimed to 1) develop the algorithm that THIM uses to track sleep and wakefulness, 2) test whether it performs comparably to other actigraphy devices, and 3) assess the impact of insomnia symptoms on device performance. This was investigated by dichotomising participants into good and poor sleeper groups based on their scores on the Insomnia Severity Index (ISI). The potential uses of THIM for these two groups differ, and therefore, it is more meaningful for the interpretation of the study findings to consider good and poor sleepers separately.

Study 2 provides preliminary evidence about the accuracy of THIM in an independent sample of healthy individuals with self-reported good or poor sleep.

Study 1: Method

Participants

Ethics approval was obtained from the Flinders University Social and Behavioural Research Ethics Committee, South Australia. Participants were recruited via print and online advertisements and completed a battery of screening questionnaires to assess their eligibility. The screening questionnaires comprised of the ISI (Morin et al., 2011), and the PSQI (Buysse et al., 1989), to assess sleep patterns and symptoms of insomnia. A health and lifestyle questionnaire was administered to assess physical and mental health conditions, as well as lifestyle factors, such as medication use, caffeine/alcohol/nicotine consumption, and recent trans-meridian travel. Both good sleepers (ISI score < 7) and those with subthreshold clinical insomnia symptoms (ISI score 8-15), termed 'poor sleepers', were recruited for this study to develop the sleep tracking algorithm in a sample with varied sleep quality. Specific eligibility criteria were as follows:

1. Self-reported average habitual bedtime between 22:00-00:00 and wake up time between 06:00-08:00;
2. Fluent in English;
3. No self-reported diagnosis of a physical or mental health condition;
4. No active nicotine or illicit substance use, or alcohol (>10 standard drinks p/wk) or caffeine (>250 mg p/day) dependence;
5. No consumption of medications known to interfere with sleep;
6. No overnight shift work or trans-meridian travel within the last two months;
7. Not pregnant or lactating.

After screening, 25 healthy individuals met the eligibility criteria. Twelve individuals participated in this study in June 2017 and 13 individuals participated from April-July 2018 as part of a larger laboratory study (described in Study 1, Chapter 3). See Table 5-1 for participant characteristic information.

Table 5-1. *Descriptive characteristics for participants in Study 1.*

Characteristic	**Good sleepers (N = 19)**	**Poor sleepers (N = 6)**	**Total Sample (N = 25)**
Age, mean (SD), y	25.20 (6.60)	25.92 (6.23)	25.38 (6.39)
Sex, No. (%)			
Men	10 (53)	0 (0)	10 (40)
Women	9 (47)	6 (100)	15 (60)
BMI, mean (SD)	23.36 (3.10)	24.10 (4.50)	23.54 (3.40)
Lifestyle characteristics			
Weekly alcohol consumption, No. (SD)	1.84 (2.52)	1.67 (1.97)	1.80 (2.36)
Daily caffeine consumption, No. (SD)	1.45 (1.21)	1.67 (1.21)	1.50 (1.19)
Sleep characteristics			
ISI, mean (SD)	1.74 (1.19)	10.67 (3.14)	3.88 (4.28)
PSQI, mean (SD)	3.53 (1.54)	7.00 (3.16)	4.36 (2.48)
Habitual Bedtime, mean time (SD, min)	22:47 (50.67)	22:35 (31.75)	22:44 (46.57)
Habitual Wake Up Time, mean time (SD, min)	07:24 (54.08)	07:25 (24.31)	07:25 (48.14)
Habitual total sleep time, mean hrs (SD)	7.96 (0.88)	7.17 (1.37)	7.77 (1.05)

BMI = body mass index, ISI = Insomnia Severity Index, N = sample size, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

Study Design

This was a within-groups quasi-experimental study. All participants slept

overnight in the sleep laboratory with polysomnography (PSG), THIM and two additional actigraphy devices, the Fitbit Flex and the Actiwatch devices, recording simultaneously. The degree of agreement was assessed between the three actigraphy devices and PSG.

Materials

Polysomnography

PSG was recorded using Compumedics Grael 4K PSG:EEG devices (Compumedics, Victoria, Australia). Six electroencephalography (EEG) sites (F3-M2, F4-M1, C3-M2, C4-M1, O1-M2, O2-M1), reference and ground, right and left electrooculography (EOG), chin electromyography (EMG), and electrocardiography (ECG) sites were recorded in accordance with the 10-20 EEG placement system. An independent registered sleep technician blind to the output from the actigraphy devices scored the PSG data using Profusion Compumedics software (v 4.0) according to standardised AASM PSG scoring criteria (American Academy of Sleep Medicine, 2018).

THIM

THIM (firmware v 1.0.3) is a ring-like device worn on the middle phalanx of the index finger of the dominant hand. The device contains an in-built tri-axial accelerometer which measures acceleration (change of velocity). The device pre-processes the raw acceleration values and stores an average value for each 30 second epoch. To retrieve this data, the device is connected via Bluetooth to the accompanying THIM smartphone application (app). Data is sent to cloud-based servers for further processing, during which a sleep tracking algorithm is applied to score every 30-second epoch into sleep or wake. This information is subsequently displayed on the THIM app as key sleep parameters –total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), and sleep efficiency - and as a

visual sleep hypnogram.

For this study, the THIM smartphone app (v 1.0.1) was operated on an Apple iPhone 5s model (iOS 8.0 operating system) to send the 30-second epoch data to the cloud-based servers. At present, the epoch data is not accessible for download. The manufacturers of THIM, Re-Time Pty. Ltd., retrieved and forwarded the data to us for the purpose of this study. We then developed the THIM sleep tracking algorithm on this data, which is applied to the THIM data in the cloud-based servers (from firmware v 1.0.4).

To create the algorithm, we first developed a smoothing function applied to pre-processed data (data after high and low pass filter processing) by iteratively adjusting the number of included previous and subsequent epochs and their weightings in the smoothing function until the algorithm reached high agreement with PSG for estimating sleep and wake periods. Secondly, a threshold applied to the epoch data to distinguish between sleep and wake epochs was identified by iteratively adjusting the threshold until acceptable sensitivity and specificity was reached across the whole sample. Thirdly, specific scoring criteria regarding the number of ‘wake’ epochs required to determine SOL and subsequent awakenings were iteratively adjusted until high correspondence was obtained between PSG and THIM across this sample. The algorithm cannot be discussed in greater detail as it is proprietary.

Actiwatch Devices

Developed by Philips Respironics, this device uses an internal tri-axial accelerometer to identify participants’ wrist movements in three-dimensional space. The Actiwatch Spectrum model was used to collect data in 2017 and the Actiwatch-2 model for 2019, however these models perform equivalently (Respironics, 2009).

The devices were set up and the data was retrieved in 30-second epochs using the Actiware Sleep software (v 6.0.0, Philips Respironics, Bend, OR). The times in and out of bed were manually entered from the lights out/lights on times recorded on the laboratory night. The default software algorithm automatically scored the epochs by applying the medium threshold criterion and '10 immobile minutes' scoring parameters. The sleep/wake epoch data was exported into Microsoft Excel for analysis.

Fitbit Flex

Similar to the Actiwatch device, the Fitbit Flex uses accelerometry to measure wrist movement. The device was operated using the Fitbit Flex smartphone app (v 3.3.1) on the same Apple iPhone 5s model phone used to operate THIM. Participants' age, gender, height and weight were entered into the Fitbit app before the laboratory nights commenced as it is unknown whether this information is incorporated into the proprietary Fitbit algorithm to estimate sleep and wakefulness. The sleep recording periods were manually initiated and terminated by tapping on the Fitbit device when the participants got in/out of bed. After the laboratory night, the 'normal' Fitbit algorithm setting was applied to score the data into 60-second epochs. The sleep/wake epoch data was retrieved via Squash Leagues (www.squashleagues.org/): a website independent of Fitbit that retrieves the epoch data from the Fitbit account, which was downloaded in a CSV format for analysis.

Procedure

Home Testing

Participants completed an online sleep diary every morning for one week via Qualtrics software. This online diary is based on the Consensus Sleep Diary (Carney et al., 2012). Participants also wore the Actiwatch device every day during this week to corroborate the sleep diary information.

Laboratory Night

Participants arrived at the Flinders University Sleep Research Laboratory at approximately 20:00. Participants were setup for overnight PSG recording. THIM was attached to the index finger on their self-reported dominant hand. The Fitbit Flex and Actiwatch devices were attached to the wrist of their non-dominant hand. Participants went to bed at their typical bedtime and woke up at their typical wake up time, as calculated from the previous week of sleep diaries.

Statistical Analysis

The accuracy of the three actigraphy devices (THIM, Fitbit Flex and Actiwatch) compared to PSG was analysed in accordance with recommended guidelines for device validation studies (de Zambotti et al., 2019; Depner et al., 2019). Epoch-by-epoch analyses were conducted by calculating the sensitivity (proportion of epochs that the device scored as *sleep* when the individual was asleep according to PSG), specificity (proportion of epochs that the device scored as *wake* when the individual was awake according to PSG) and accuracy (proportion of correctly scored epochs) separately for each participant and averaging these values together for each actigraphy device. Linear Mixed Modelling (LMM) was performed to examine whether there were any significant differences between the actigraphy devices (the fixed effect) on sensitivity, specificity and accuracy (IBM SPSS, v 23). All LMM analyses used a first-order autoregressive covariance structure with device as a fixed effect. Where appropriate, post hoc comparisons were conducted with the Bonferroni correction.

The limit of agreement between PSG and each actigraphy device was also analysed using Bland-Altman plots (Bland & Altman, 1986; Giavarina, 2015). These plots show the discrepancy between PSG and the device for each participant (y axis)

against PSG (x axis) on separate plots for each sleep parameter. The plots also display the overall mean difference (also known as the bias), standard deviation, and the lower and upper limits of agreement (± 1.96 SD of the mean difference).

Estimations of the common sleep parameters were compared between each actigraphy device and PSG. For the actigraphy devices, TST was calculated from the sum of epochs that do not exceed the sensitivity threshold (i.e. epochs defined as sleep). SOL was calculated from the sum of epochs that exceeded the sensitivity threshold (i.e. wake epochs) between lights out and the first sleep epoch. WASO was calculated from the sum of wake epochs between the first epoch of sleep and lights on. Sleep efficiency was calculated by dividing TST by the total time spent in bed and multiplied by 100. PSG sleep parameters were defined according to established guidelines (American Academy of Sleep Medicine, 2018). LMM analyses examined whether there were any significant differences between actigraphy devices (the fixed effect) for estimating each sleep parameter (SOL, TST, WASO, and sleep efficiency). A statistically significant main effect for device was further examined using Bonferroni adjusted pairwise comparisons.

Additional analyses included examining whether the type of sleeper (good or poor sleeper) impacted the accuracy of the actigraphy devices. Sleeper type was entered as a factor in all LMM analyses discussed above. Where the interaction between device and sleeper type was statistically significant, Bonferroni-adjusted pairwise comparisons were conducted to further investigate the effect.

Study 1: Results

Missing Data

Four nights of Actiwatch data were missing due to battery difficulties. All nights of data were obtained with the THIM and Fitbit devices.

Epoch by Epoch Analysis

The sensitivity, specificity and accuracy of each actigraphy device are presented in Table 5-2. As shown, all three actigraphy devices had high sensitivity. A LMM indicated that the sensitivities differed between devices, $F(2, 68) = 21.16$, $p < .001$. Post hoc tests showed that THIM had significantly lower sensitivity than the Actiwatch, $p = .001$, and the Fitbit Flex devices, $p < .001$. According to Cohen's d criteria, the difference was large between THIM and the Actiwatch Spectrum, $d = 0.88$, as well as between THIM and the Fitbit Flex device, $d = 1.70$. There was no significant difference between the Actiwatch and Fitbit Flex mean sensitivities, $p = .07$.

Specificities also differed between devices, $F(2, 68) = 12.11$, $p < .001$. Post hoc tests indicated that THIM had a significantly higher specificity than the Actiwatch Spectrum device, p = .001, and the Fitbit Flex devices, p < .001. The effect sizes were large between THIM and the Actiwatch Spectrum, $d = 1.23$, as well as between the THIM and the Fitbit Flex devices, $d = 1.26$. However, there was no significant difference between the specificities for the Actiwatch and Fitbit Flex devices, $p = .99$. There were no significant differences in accuracy between devices, $F(2, 68) = 0.49$, $p = .61$.

Table 5-2. *Sensitivity, specificity and accuracy for each actigraphy device in Study 1.*

Device	Sensitivity (SD)	Specificity (SD)	Accuracy (SD)
THIM	0.91 (0.05)	0.59 (0.21)	0.85 (0.07)
Actiwatch	0.95 (0.04) *	0.35 (0.18) *	0.85 (0.10)
Fitbit Flex	0.98 (0.03) *	0.32 (0.22) *	0.87 (0.09)

* p < .05 between THIM and device. SD = standard deviation.

Sleep Parameter Estimations

Table 5-3 presents the descriptive statistics on estimations of each sleep

parameter for each device. A LMM determined there were significant differences between devices for estimations of SOL, $F(3, 92) = 6.39$, $p = .001$. Post hoc comparisons indicated there were large significant differences between PSG and the Actiwatch device, $p < .001$, $d = 1.31$, and Fitbit Flex devices, $p = .04$, $d = 0.77$. There was no significant difference between PSG and THIM estimations of SOL, $p = .99$.

Table 5-3. *Sleep parameter descriptive statistics for PSG and the actigraphy devices in Study 1.*

Sleep Parameter	Actigraphy Device			
	PSG	**THIM**	**Actiwatch**	**Fitbit Flex**
SOL, mean (SD), min	24.22 (19.98)	20.40 (21.46)	5.36 (3.77) *	11.64 (11.54) *
TST, mean (SD), min	400.96 (67.02)	394.46 (48.72)	423.10 (51.00)	440.60 (45.06)
WASO, mean (SD), min	48.06 (36.34)	58.38 (22.55)	38.00 (26.53)	19.60 (18.37) *
Sleep efficiency, mean (SD), %	84.58 (11.10)	83.50 (7.56)	90.73 (6.04)	93.52 (5.10) *

* $p < .05$ between PSG and actigraphy device. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

There were significant differences for estimations of TST, $F(3, 92) = 3.84$, $p = .01$. However, post hoc comparisons indicated no significant differences between PSG and any of the actigraphy devices, $p > .06$.

There were significant differences for WASO, $F(3, 92) = 9.48$, $p < .001$. Post hoc tests indicated a large significant difference between PSG and the Fitbit Flex device, $p = .002$, $d = 0.99$, but no significant differences between PSG and THIM, $p = .99$, or PSG and the Actiwatch device, $p = .99$.

There were significant differences for sleep efficiency, $F(3,92) = 9.27$, $p < .001$. Post hoc comparisons indicated large significant differences between PSG and

the Fitbit Flex, p = .004, d = 1.04. There were no significant differences between PSG and THIM, p = .99, or the Actiwatch for estimations of sleep efficiency, p = .06.

Bland-Altman Plots

Figure 5-1 presents Bland-Altman plots for each actigraphy device on key sleep parameters. THIM had a mean bias closer to perfect agreement with PSG compared to the Actiwatch and Fitbit Flex devices. This suggests that THIM agrees more closely with PSG for estimating SOL compared to the Actiwatch and Fitbit Flex. Furthermore, THIM had a less steep slope for the line of best fit compared to the Actiwatch and Fitbit Flex, suggesting that the device was less likely to underestimate SOL to a greater extent as PSG-SOL increased. The limits of agreement range between 60.07 minutes for the Fitbit Flex to 72.56 minutes for the Actiwatch.

The TST plots show that the mean biases for the Actiwatch and Fitbit Flex devices trends towards overestimation. Yet, considering the findings of the LMM analyses above, there is no significant difference between devices. All three devices have lines of best fit with steep negative slopes. The limits of agreement range between 139.34 minutes for the Actiwatch to 188.50 minutes for the Fitbit Flex.

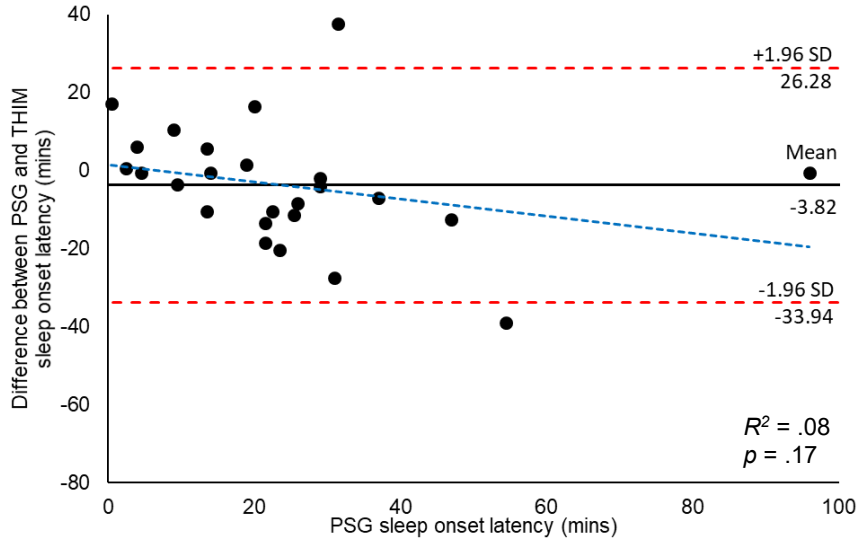
The WASO plots illustrate that THIM tends to overestimate WASO whilst the other devices tend to underestimate WASO. Yet, the LMM analyses indicate that only the Fitbit Flex produced significantly different estimates of WASO compared to PSG. All devices have lines of best fit with steep negative slopes. The limits of agreement range between 122.43 minutes for the Actiwatch device to 153.85 minutes for the Fitbit Flex, potentially due to the presence of an outlier.

The sleep efficiency plots illustrate that THIM had a small bias towards underestimating sleep efficiency, while the other devices overestimated sleep efficiency. Yet, the Fitbit Flex was the only device to produce significantly different

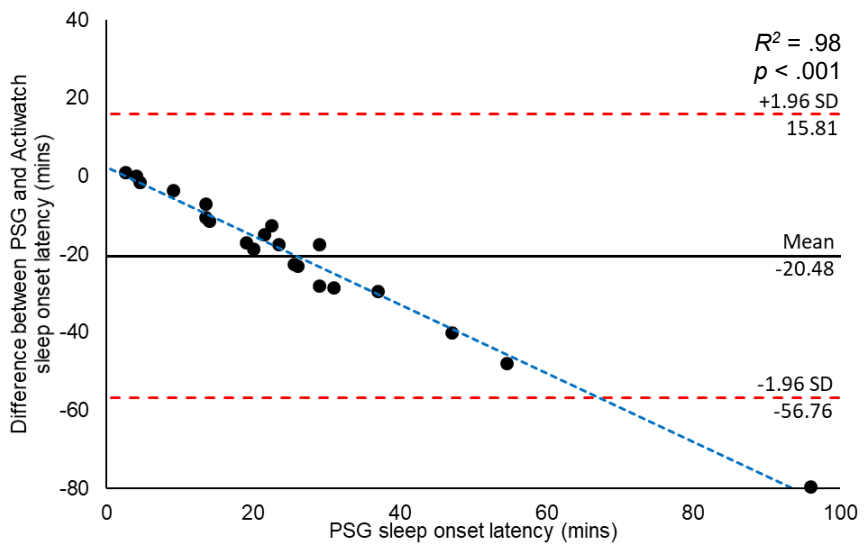
estimates of sleep efficiency compared to PSG. The limits of agreement range between 33.25% for the Actiwatch device to 43.19% for the Fitbit Flex.

Sleep Onset Latency

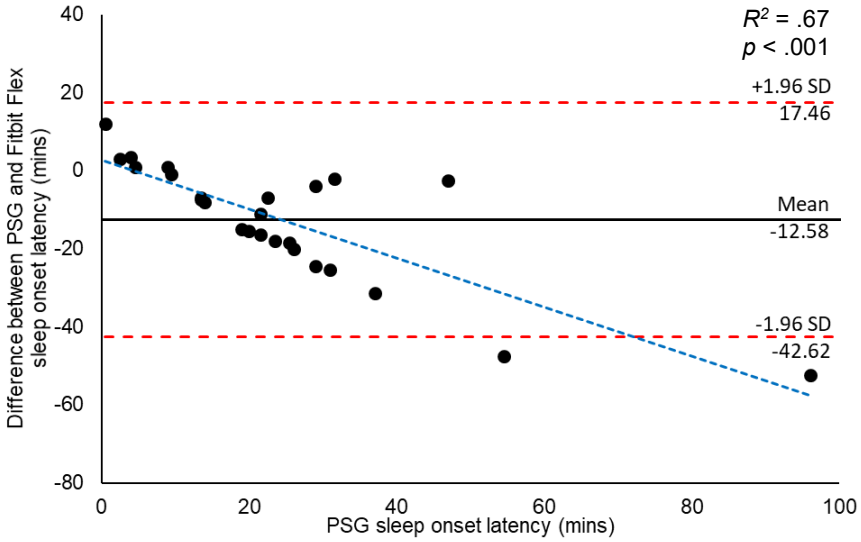
(a) THIM



(b) Actiwatch

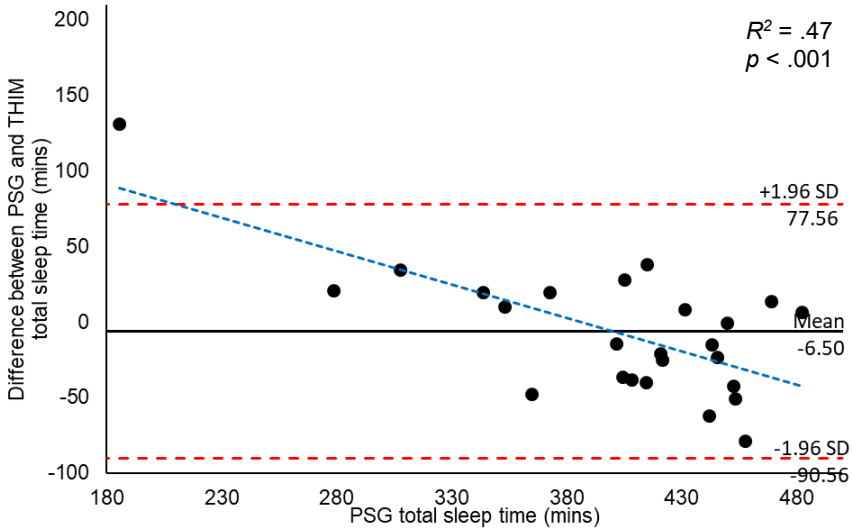


(c) Fitbit Flex

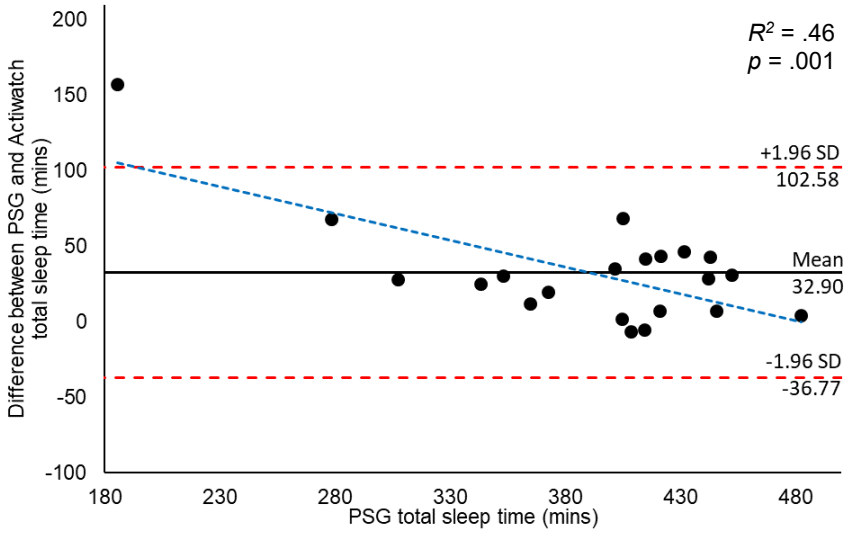


Total Sleep Time

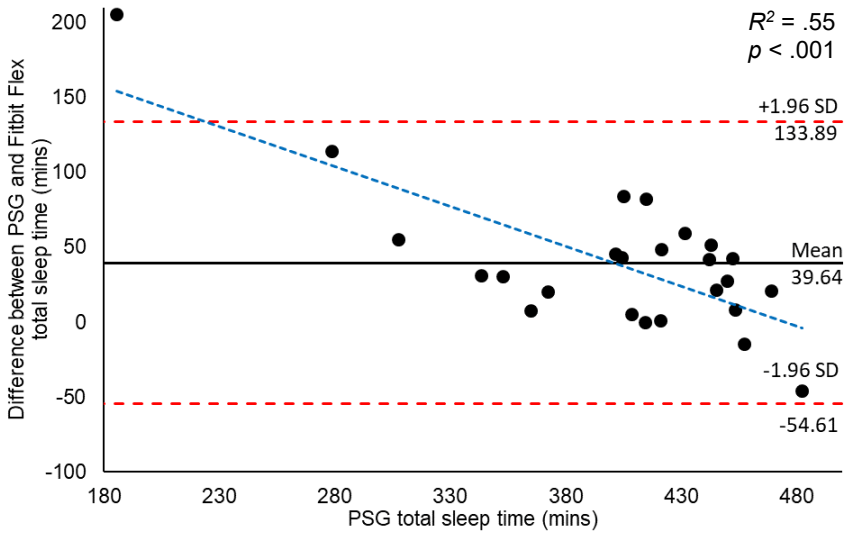
(a) THIM



(b) Actiwatch

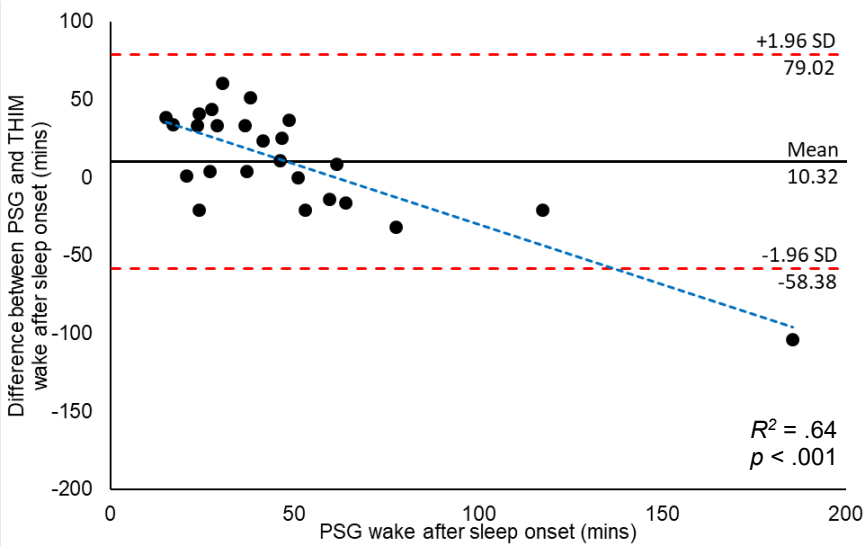


(c) Fitbit Flex

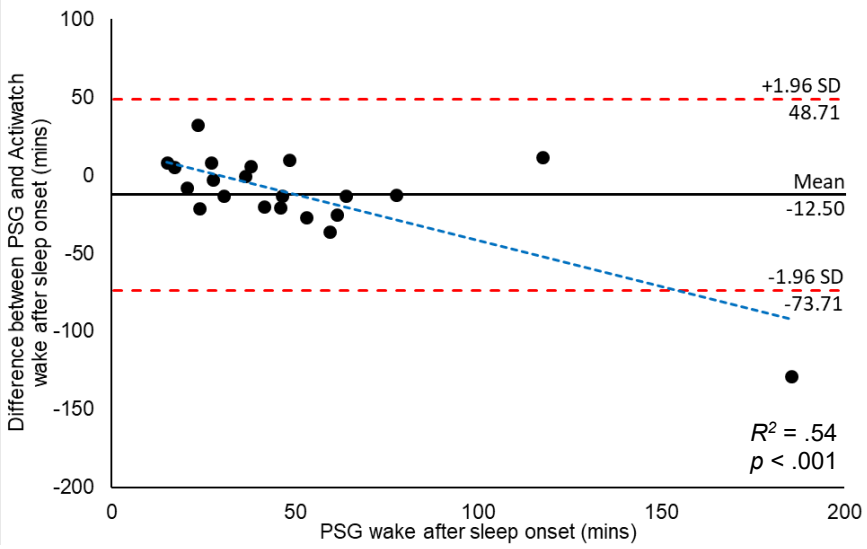


Wake After Sleep Onset

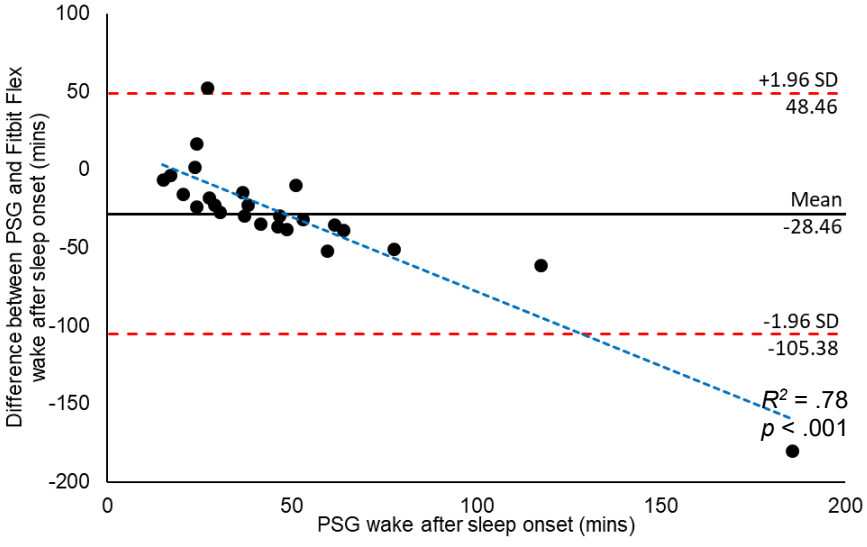
(a) THIM



(b) Actiwatch

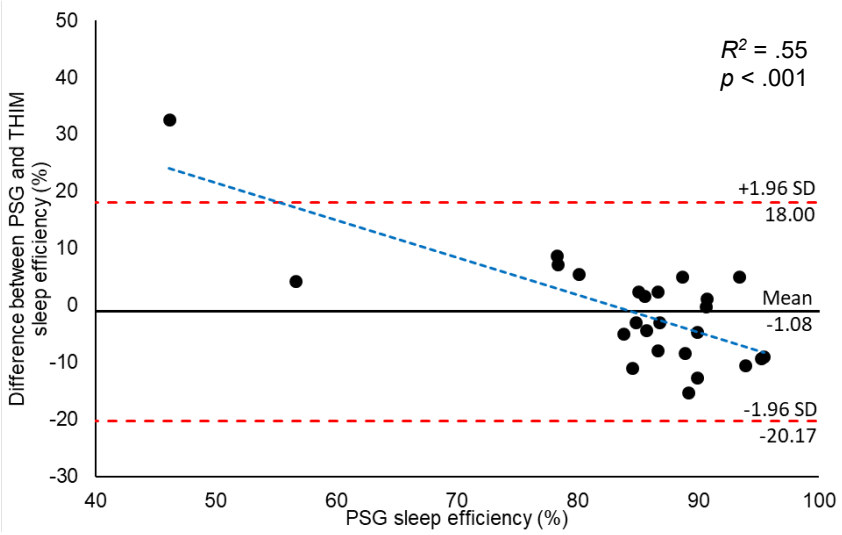


(c) Fitbit Flex

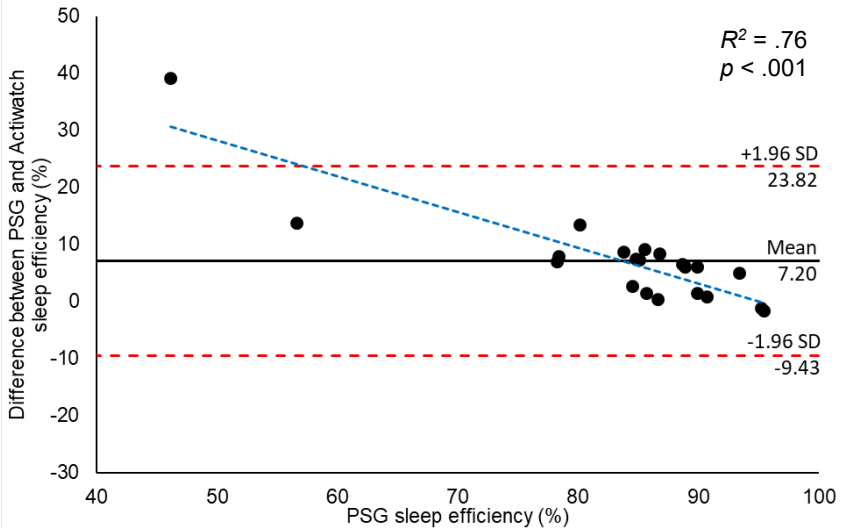


Sleep Efficiency

(a) THIM



(b) Actiwatch



(c) Fitbit Flex

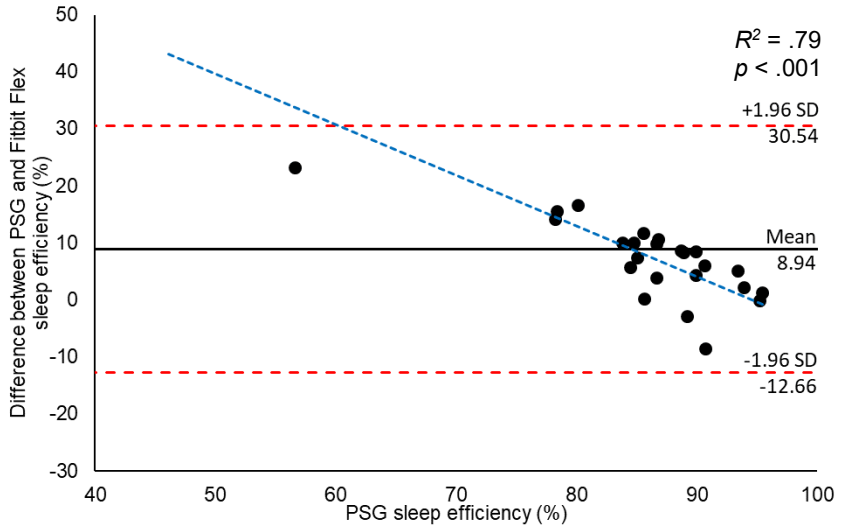


Figure 5-1. Bland-Altman plots showing the agreement between PSG and THIM, Actiwatch and Fitbit Flex devices separately on sleep onset latency, total sleep time, wake after sleep onset and sleep efficiency for Study 1.
The solid black lines indicate the bias. The dashed red horizontal lines indicate the upper and lower limits of agreement, and the dotted blue lines are the lines of best fit.

Good and Poor Sleeper Comparison.

Table 5-4 contains the descriptive statistics for these secondary analyses. For sensitivity there was a statistically significant interaction between device and the type of sleeper, $F(2,43.81) = 8.66$, $p = .001$. Post hoc analyses revealed that sensitivity was significantly higher for good sleepers compared to poor sleepers for the THIM device, $p < .001$. This was a large effect, $d = 1.50$. There were no significant differences between good and poor sleepers for the sensitivity of the Actiwatch, $p = .07$, or the Fitbit Flex devices, $p = .93$. A LMM examining the interaction between device and sleeper type on specificity was not significant, $p = .77$. There was a statistically significant interaction between device and the type of sleeper on accuracy, $F(2,42.55) = 6.44$, $p = .004$. However, post-hoc comparisons between groups within devices were not significant, $p > .12$.

LMM analyses determined whether there were any significant differences between good and poor sleepers on the mean discrepancies of the actigraphy devices for each sleep parameter. The descriptive statistics for these analyses are presented in Table 5-4. These LMM analyses found no significant interactions between device and the type of sleeper on SOL, $p = .66$, TST, $p = .06$, WASO, $p = .08$, or sleep efficiency, $p = .08$.

Table 5-4. *Epoch-by-epoch and sleep parameter descriptive statistics comparing good and poor sleepers in Study 1.*

Variable	Actigraphy Device		
	THIM	**Actiwatch**	**Fitbit Flex**
Epoch-by-epoch analyses			
Sensitivity, mean (SD)			
Good sleepers	0.92 (0.04)	0.94 (0.04)	0.98 (0.03)
Poor sleepers	0.86 (0.04) *	0.98 (0.01)	0.98 (0.02)
Specificity, mean (SD)			
Good sleepers	0.55 (0.22)	0.34 (0.15)	0.31 (0.19)
Poor sleepers	0.68 (0.16)	0.45 (0.27)	0.37 (0.29)
Accuracy, mean (SD)			
Good sleepers	0.86 (0.07)	0.84 (0.11)	0.86 (0.10)
Poor sleepers	0.84 (0.04)	0.91 (0.04)	0.92 (0.03)
Sleep parameters			
SOL discrepancy, mean (SD), min			
Good sleepers	-7.24 (13.12)	-21.31 (19.98)	-14.45 (15.81)
Poor sleepers	7.00 (18.13)	-9.90 (10.77)	-6.66 (13.12)
TST discrepancy, mean (SD), min			
Good sleepers	5.13 (40.91)	33.61 (38.46)	45.13 (52.31)
Poor sleepers	-43.33 (25.73)	15.43 (21.20)	22.25 (27.75)
WASO discrepancy, mean (SD), min			
Good sleepers	2.11 (35.93)	-12.38 (34.59)	-32.16 (43.45)
Poor sleepers	36.33 (13.43)	-7.25 (10.73)	-16.75 (19.48)
Sleep Efficiency discrepancy, mean (SD), %			
Good sleepers	1.39 (9.62)	7.37 (9.23)	10.25 (12.05)
Poor sleepers	-8.91 (5.06)	3.19 (4.36)	4.77 (5.74)

* $p < .05$ between good and poor sleepers with this device. SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Study 1: Discussion

Study 1 aimed to develop the THIM sleep tracking algorithm and compare its accuracy to two popular actigraphy devices. The epoch-by-epoch analysis revealed that THIM was less sensitive for detecting sleep compared to the Fitbit Flex and Actiwatch devices but had higher specificity and comparable overall accuracy. Analysis of the sleep parameter estimations further demonstrated that THIM aligned closely with PSG, with no significant differences between THIM and PSG for any sleep parameter. In comparison, the Fitbit Flex and Actiwatch Spectrum devices' estimations of SOL were significantly lower than PSG, and the Fitbit Flex's estimations of WASO and sleep efficiency differed to PSG. Overall, these findings suggest that THIM has comparable accuracy to the Actiwatch and Fitbit Flex devices, with perhaps greater agreement with PSG for estimating key sleep parameters. However, the THIM sleep tracking algorithm was developed and optimised for estimating sleep and wake with this sample. As such, the device may not be as accurate with a different sample of healthy individuals. To draw stronger conclusions about the accuracy of THIM, the device needed to be tested with a separate sample. This was addressed in Study 2.

This study examined whether this high variability across individuals evident on the Bland-Altman plots could be due to the type of sleeper (good or poor sleeper). There was only one significant difference between good and poor sleepers across the dependent variables for the actigraphy devices – THIM showed significantly lower sensitivity for poor sleepers. Nonetheless, sleeper characteristics is an important factor that has impacted the accuracy of actigraphy devices in previous research (Hedner et al., 2004; Van den Water et al., 2011), although some studies have found no significant differences between good sleepers and those with

insomnia (Kang et al., 2017; Sanchez-Ortuno et al., 2010). Further investigation is warranted to understand the suitability of THIM for monitoring the sleep of individuals with good or poor sleep.

The aims of Study 2 were three-fold: to 1) test the accuracy of the THIM algorithm developed in Study 1 with an independent sample, 2) determine whether the device is more accurate than other actigraphy devices, and 3) investigate whether the accuracy of the actigraphy devices differ between good and poor sleepers.

Study 2: Method

This study tested the accuracy of the actigraphy devices with an independent sample. The Actiwatch Spectrum and Fitbit Flex devices were substituted with the updated Actiwatch-2 and Fitbit Alta devices. Other aspects of the study method are identical to the first study.

Participants

Participants met the same eligibility criteria as participants in the first study. Twenty-one healthy individuals participated in this study. However, one recording failed due to technician error with the PSG. As such, these findings are based on 20 participants. See participant characteristic information in Table 5-5.

Table 5-5. *Descriptive statistics for participant characteristics collected at screening in Study 2.*

Characteristic	Good sleepers (N = 10)	Poor sleepers (N = 10)	Total Sample (N = 20)
Age, mean (SD), y	24.86 (5.60)	21.88 (4.29)	23.22 (5.01)
Sex, No. (%)			
Men	3 (33)	3 (27)	6 (30)
Women	6 (67)	8 (73)	14 (70)
BMI, mean (SD)	24.42 (2.25)	24.88 (3.95)	24.68 (3.22)
Lifestyle characteristics			
Weekly alcohol consumption, No. (SD)	1.56 (1.74)	1.27 (1.79)	1.40 (1.73)
Daily caffeine consumption, No. (SD)	1.72 (1.60)	1.91 (1.36)	1.83 (1.44)
Sleep characteristics			
ISI, mean (SD)	1.89 (1.17)	11.36 (3.83)	7.10 (5.63)
PSQI, mean (SD)	3.22 (1.79)	7.73 (3.50)	5.70 (3.61)
Habitual Bedtime, mean time (SD), min	22:36 (61.68)	23:08 (67.43)	22:54 (65.50)
Habitual Wake Up Time, mean time (SD), min	07:17 (53.21)	07:56 (68.54)	07:38 (64.26)
Habitual TST, mean (SD), hrs	8.06 (0.88)	7.09 (1.50)	7.53 (1.32)

BMI = body mass index, ISI = Insomnia Severity Index, N = sample size, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

Results

Missing Data

Due to battery issues, three nights of Actiwatch data were missing. All nights were recorded successfully with the THIM and Fitbit devices.

Epoch by Epoch Analysis

Table 5-6 presents the descriptive statistics for the epoch-by-epoch analyses with each actigraphy device. Despite high sensitivity, a LMM revealed that there

were significant differences between the devices, $F(2, 54) = 14.52$, $p < .001$. Post hoc tests showed that THIM had a significantly lower sensitivity than the Actiwatch-2, $p < .001$, $d = 1.18$, and Fitbit Alta devices, $p < .001$, $d = 1.37$. There was no significant difference between the Actiwatch-2 and Fitbit Alta devices, $p = .99$.

Furthermore, a LMM indicated significant differences in the specificities between devices, $F(2, 54) = 7.72$, $p = .001$. Post hoc tests indicated that both THIM and the Actiwatch-2 had significantly higher specificities than the Fitbit Alta, $p < .005$, $d = 1.11$ and 1.05, respectively. There was no significant difference between THIM and the Actiwatch-2, $p = .99$.

A significant main effect of device was found on accuracy, $F(2, 54) = 5.14$, $p = .009$. Post hoc tests indicated THIM had significantly lower accuracy than the Actiwatch-2, $p = .01$, but was not significantly different compared to the Fitbit Alta, $p = .09$. There was no significant difference between the Actiwatch-2 and Fitbit Alta devices, $p = .99$.

Table 5-6. *Sensitivity, specificity and accuracy for each actigraphy device in Study 2.*

Device	Sensitivity (SD)	Specificity (SD)	Accuracy (SD)
THIM	0.89 (0.06)	0.59 (0.18)	0.85 (0.06)
Actiwatch-2	0.95 (0.04) *	0.59 (0.20)	0.91 (0.05) *
Fitbit Alta	0.96 (0.04) *	0.39 (0.18) *	0.89 (0.07)

* $p < .05$ between THIM and device. SD = standard deviation.

Sleep Parameter Estimations

Table 5-7 presents the descriptive statistics for the sleep parameter estimations. A LMM determined significant differences on estimations of SOL, $F(3, 73) = 4.07$, $p = .01$. Post hoc comparisons indicated a large significant difference between PSG and the Actiwatch-2, $p = .01$, $d = 1.41$, and no significant differences between PSG and THIM, $p = .99$, or the Fitbit Alta, $p = .81$. There were no significant

differences on estimations of TST, $F(3, 73) = 2.23$, $p = .75$. There were significant differences on WASO, $F(3, 73) = 7.44$, $p < .001$. However, post hoc comparisons indicated that these significant differences were not between PSG and any of the actigraphy devices, $p > .06$. Similarly, there were significant differences on sleep efficiency, $F(3, 73) = 6.95$, $p < .001$, but post hoc comparisons indicated that there were no significant differences between PSG and any actigraphy device, $p > .14$.

Table 5-7. *Sleep parameter descriptive statistics for PSG and the actigraphy devices from Study 2.*

Sleep Parameter	Actigraphy Device			
	PSG	THIM	Actiwatch-2	Fitbit Alta
SOL, mean (SD), min	21.68 (16.65)	19.15 (17.82)	4.53 (4.25) *	14.03 (19.22)
TST, mean (SD), min	424.60 (48.73)	403.25 (46.15)	434.35 (46.66)	438.05 (43.54)
WASO, mean (SD), min	40.68 (34.50)	64.55 (30.36)	45.03 (24.84)	22.28 (21.88)
Sleep efficiency, mean (SD), %	87.36 (8.62)	82.93 (7.82)	89.88 (4.86)	92.33 (4.56)

* p < .05 between PSG and actigraphy device. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Bland-Altman Plots

Figure 5-2 presents Bland-Altman plots for each actigraphy device. Overall, the plots are similar to those found in Study 1. The SOL plots show that THIM has a mean bias in close agreement with PSG. The limits of agreement are similar between THIM (range: 58.17) and the Actiwatch-2 device (range: 58.16), with the Fitbit Alta device having wider limits of agreement (range: 93.64). THIM also has a less steep line of best fit compared to the Actiwatch-2 and Fitbit Alta devices, indicating less of a bias to increasingly underestimate SOL as PSG-SOL increases.

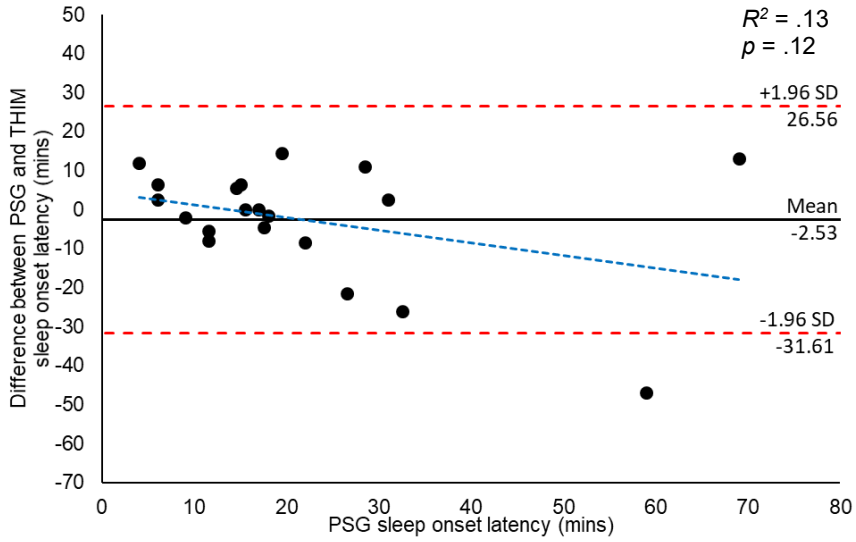
The TST plots indicate that both THIM and the Fitbit Alta devices have greater variability of agreement compared to the Actiwatch-2 device. THIM also appears to have a mean bias towards underestimating TST compared to the Actiwatch-2 device that shows a mean bias close to zero and the Fitbit Alta device that has a mean bias greater than zero. Yet, considering the findings above, these biases do not produce estimation of TST that are significantly different to PSG. The limits of agreement ranged between 112.61 minutes for the Actiwatch device to 193.39 minutes for the Fitbit Flex.

The WASO plots illustrate limits of agreement that ranged between 113.49 minutes for the Actiwatch device to 143.72 minutes for the THIM device. All devices have lines of best fit with steep negative slopes, suggesting a tendency to underestimate WASO to a greater extent as PSG-WASO increases in duration.

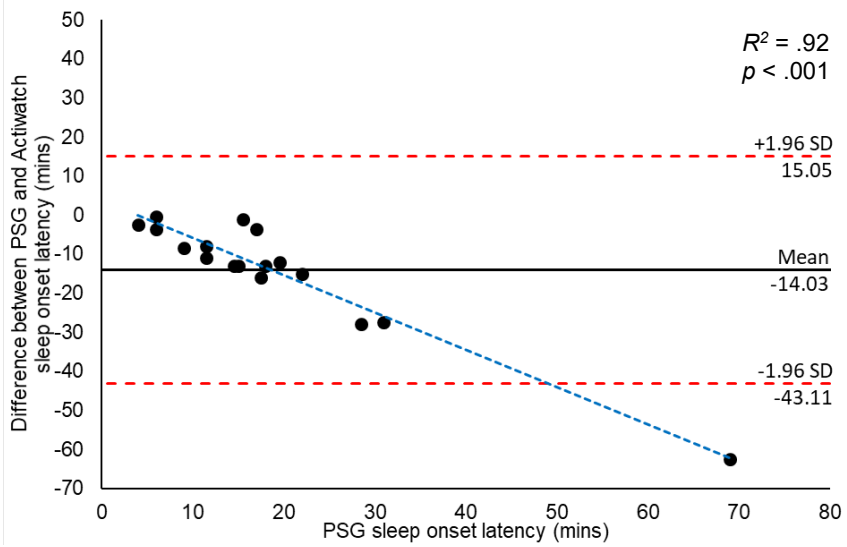
The sleep efficiency plots further illustrate that the Actiwatch-2 produces estimates that agree more closely with PSG than the THIM and Fitbit devices, with a mean bias closer to zero. THIM has a mean bias below zero (underestimate sleep efficiency), whilst the Fitbit Alta has a mean bias above zero (overestimate sleep efficiency). Nonetheless, the findings of the LMM analyses above indicate that these biases do not produce estimation of sleep efficiency significantly different to PSG. The limits of agreement ranged between 20.45% for the Actiwatch device to 35.42% for the Fitbit Flex.

Sleep Onset Latency

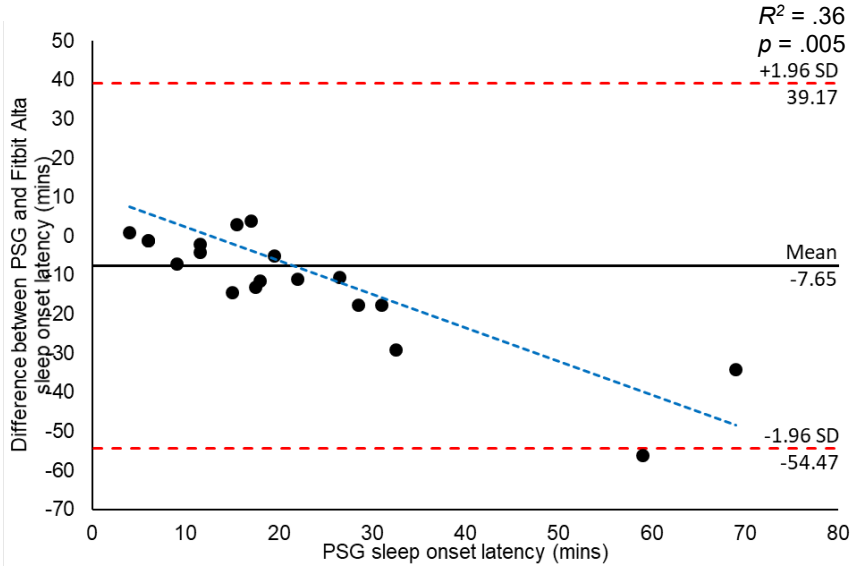
(a) THIM



(b) Actiwatch

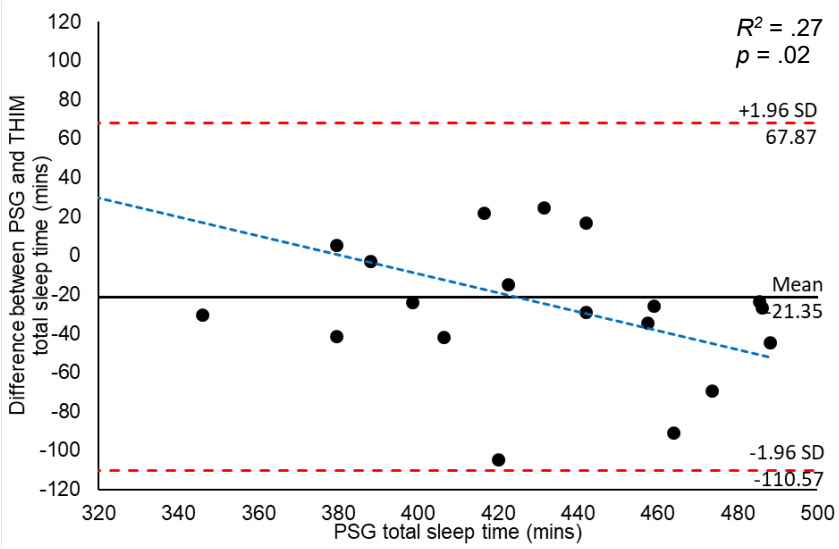


(c) Fitbit Alta

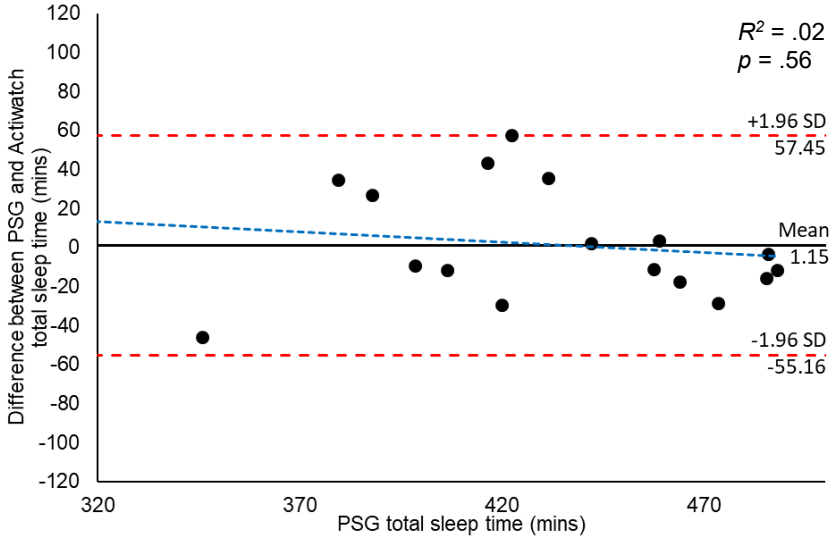


Total Sleep Time

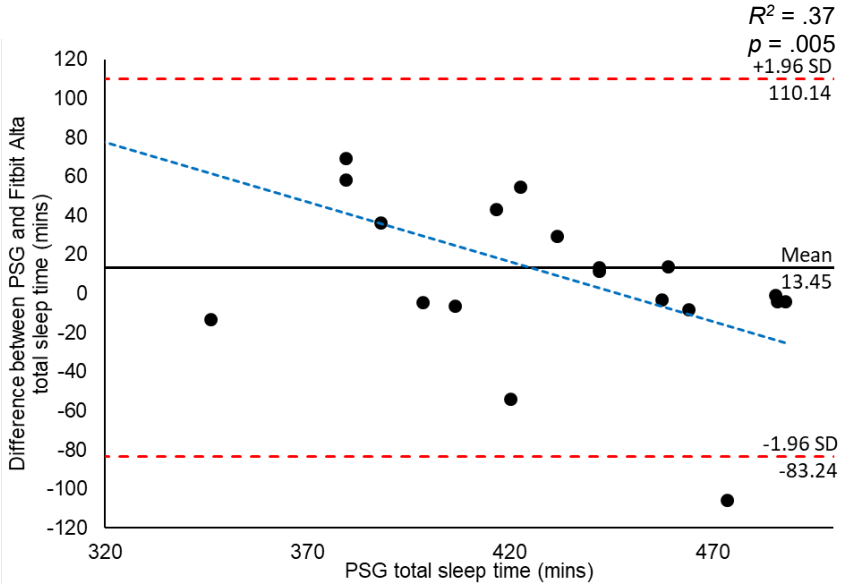
(a) THIM



(b) Actiwatch

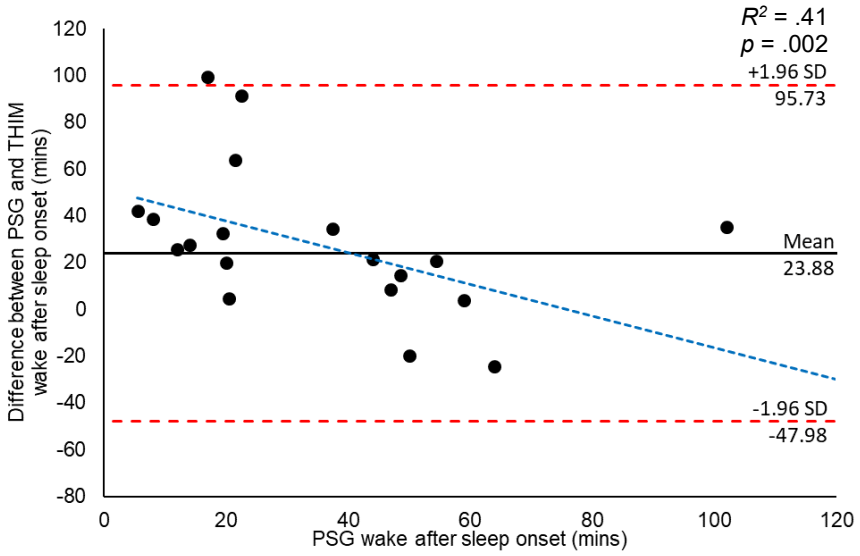


(c) Fitbit Alta

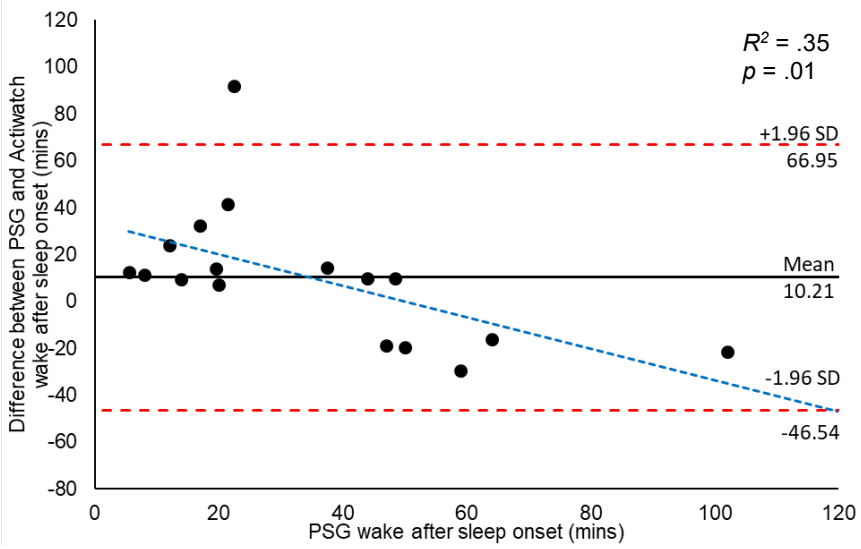


Wake After Sleep Onset

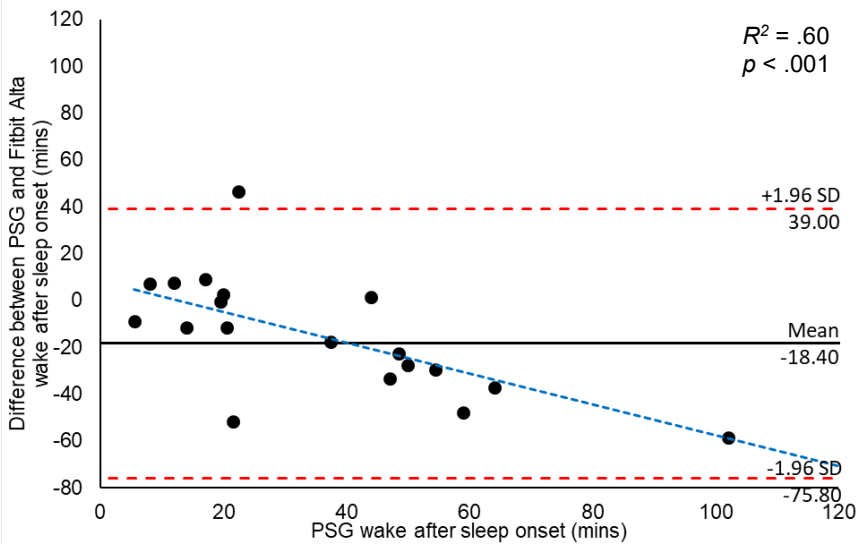
(a) THIM



(b) Actiwatch

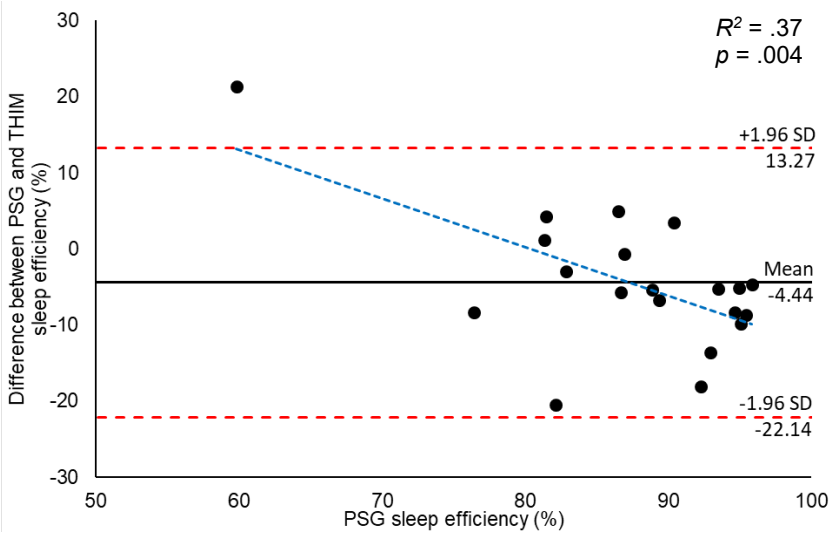


(c) Fitbit Alta

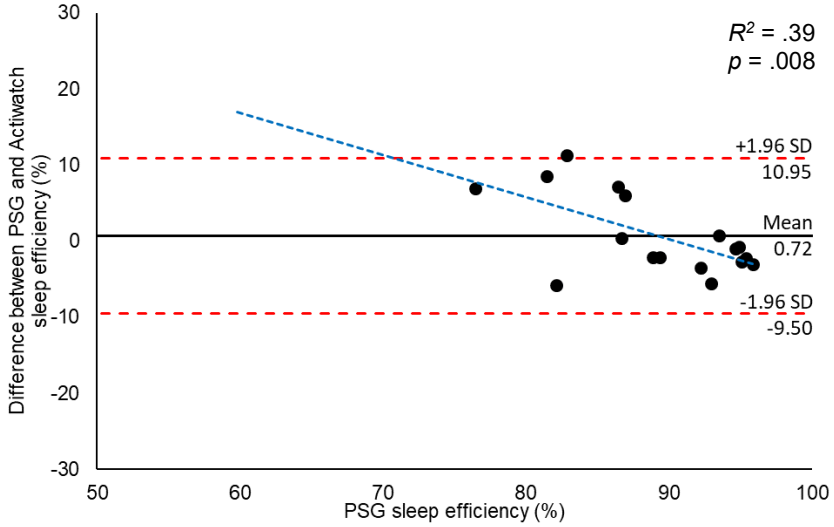


Sleep Efficiency

(a) THIM



(b) Actiwatch



(c) Fitbit Alta

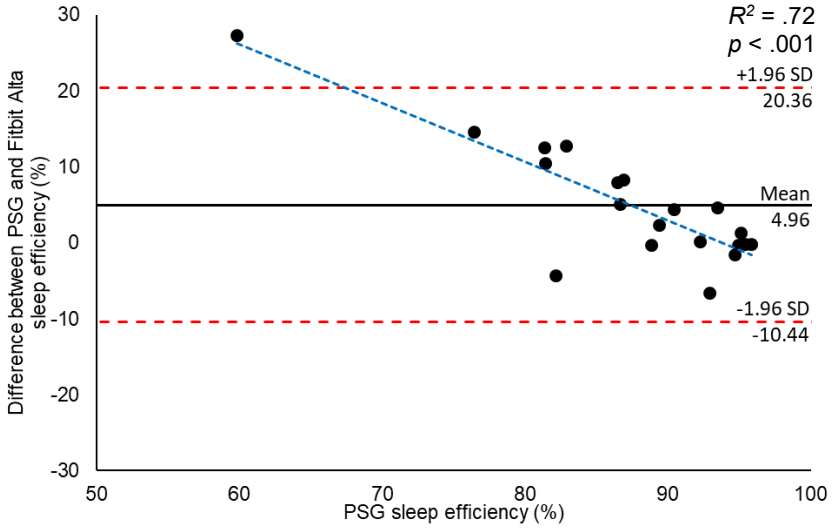


Figure 5-2. Bland-Altman plots showing the agreement between PSG and THIM, Actiwatch-2 and Fitbit Alta devices separately on sleep onset latency, total sleep time, wake after sleep onset and sleep efficiency for Study 2.

The solid black lines indicate the bias. The dashed red horizontal lines indicate the upper and lower limits of agreement, and the dotted blue lines are the lines of best fit.

Good and Poor Sleeper Comparison

Table 5-8 contains the descriptive statistics for LMM analyses conducted to determine whether there were any significant differences between good and poor sleepers on the sensitivity, specificity and accuracy of the actigraphy devices. The interactions between the actigraphy devices and sleeper type were not statistically significant for sensitivity, $p = .78$, specificity, $p = .43$, or accuracy, $p = .37$.

The descriptive statistics for LMM analyses comparing mean discrepancies between PSG and each actigraphy device on the sleep parameters are also presented in Table 5-8. There were no significant interactions between device and sleeper type on SOL, $p = .55$, TST, $p = .75$, WASO, $p = .47$, or sleep efficiency, $p = .95$.

Table 5-8. *Epoch-by-epoch and sleep parameter descriptive statistics comparing good and poor sleepers from Study 2.*

Variable	Actigraphy Device		
	THIM	**Actiwatch-2**	**Fitbit Alta**
Epoch-by-epoch analyses			
Sensitivity, mean (SD)			
Good sleepers	0.91 (0.05)	0.97 (0.01)	0.98 (0.01)
Poor sleepers	0.87 (0.07)	0.94 (0.04)	0.94 (0.05)
Specificity, mean (SD)			
Good sleepers	0.54 (0.16)	0.52 (0.15)	0.29 (0.10)
Poor sleepers	0.63 (0.19)	0.64 (0.22)	0.47 (0.20)
Accuracy, mean (SD)			
Good sleepers	0.87 (0.04)	0.91 (0.04)	0.90 (0.04)
Poor sleepers	0.83 (0.07)	0.89 (0.05)	0.88 (0.08)
Sleep parameters			
SOL discrepancy, mean min (SD)			
Good sleepers	-2.17 (13.56)	-12.07 (8.65)	-11.44 (8.95)
Poor sleepers	-2.82 (16.46)	-15.40 (18.34)	-4.55 (31.57)
TST discrepancy, mean min (SD)			
Good sleepers	-14.44 (24.75)	14.36 (28.02)	25.50 (28.84)
Poor sleepers	-27.00 (58.04)	-8.10 (26.70)	3.59 (61.00)
WASO discrepancy, mean min (SD)			
Good sleepers	16.61 (17.54)	-2.29 (20.77)	-19.67 (22.91)
Poor sleepers	29.82 (47.13)	18.95 (31.57)	-17.36 (34.74)
Sleep Efficiency discrepancy, mean % (SD)			
Good sleepers	-2.84 (4.95)	2.87 (5.54)	6.27 (5.97)
Poor sleepers	-5.74 (11.46)	-0.78 (4.68)	3.89 (9.27)

* $p < .05$ between good and poor sleepers with this device. SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Study 2: Discussion

The first aim of Study 2 was to test the accuracy of THIM with an independent sample. THIM had slightly lower sensitivity compared to the findings from Study 1,

reflecting a greater bias towards scoring an epoch as wake rather than sleep in this independent sample. However, this greater bias did not produce estimates of sleep parameters that significantly differed from PSG. The THIM Bland-Altman plots appeared comparable between Study 1 and Study 2, with relatively high variability in the discrepancy between PSG and each actigraphy device shown across all devices. Together, these findings suggest that THIM was similar in accuracy for estimating sleep and wake in the independent sample as the sample from which the algorithm was developed.

The second aim of Study 2 was to compare the accuracy of THIM to the Fitbit Alta and Actiwatch-2 devices. THIM had lower sensitivity yet higher specificity than the Fitbit device. Overall, THIM had lower accuracy than the Acitwatch-2 device. The Bland-Altman plots also indicated that THIM had a bias towards underestimating TST and sleep efficiency, and overestimating WASO compared to the other two actigraphy devices. This contrasts previous research with similar actigraphy devices that demonstrate a bias towards overestimating sleep and underestimating wake (Bianchi, 2017; Evenson et al., 2015). This may be due to differences between algorithms or device placement, or both, between THIM and other actigraphy devices.

Study 2 also aimed to determine whether there were any differences in the accuracy of these devices between good and poor sleepers. Similar to the findings of Study 1, there were no significant differences between good and poor sleepers for any of the dependent variables across actigraphy devices. This contrasts with previous research where actigraphy devices were less accurate for those with a range of sleep problems compared to good sleepers (Hedner et al., 2004).

Nonetheless, as found in Study 1, there is still considerably high variability of

agreement on the Bland-Altman plots for all actigraphy devices. The limits of agreement ranged from a minimum of 58 minutes for SOL, 112 minutes for TST, 113 minutes for WASO, and 20% for sleep efficiency across the actigraphy devices. Whilst there are no current recommendations about acceptable limits of agreement for actigraphy devices (de Zambotti et al., 2019) this degree of variability is presumably not acceptable to appropriately substitute for PSG, particularly when interpreting the data at the individual-level (Danzig, Wang, Shah, & Trotti, 2019). Additional individual characteristics that theoretically may explain high variability in the accuracy of actigraphy devices include age, gender, BMI, and the presence of sleep disorders (Danzig et al., 2019; de Zambotti et al., 2019). In additional LMM analyses, the main effects of gender and BMI were not significant across any of the sleep parameters, and therefore, these factors do not explain the variability in this sample. Due to none of the participants having a sleep disorder and a small age range, these factors are likely to have a small effect size in the current study.

It is important to note the limitations of the current study. The PSG data was scored by one qualified sleep technician. The interrater reliability of PSG sleep scoring amongst qualified individuals can be low (Rosenberg & Van Hout, 2013), increasing the error of measurement of our gold standard measure. Additionally, the sample size was relatively low compared to other validation studies. This likely impacted our ability to detect differences in the accuracy of the actigraphy devices between groups (good and poor sleepers). These factors should be considered when interpreting and generalising the findings of the current study to the general population.

Importantly, THIM has not been tested for people with insomnia and could be assessed in future research. Studies could also investigate night-to-night variability

in the accuracy of THIM, which is particularly important to assess since people with insomnia have high variability in sleep across nights (Buysse et al., 2010). Additionally, considering that people with insomnia experience different sleep quality in the sleep laboratory compared to their home environments (Edinger et al., 1997; Edinger et al., 2001), it is particularly important to test THIM in individuals' homes. Future studies could also collect data from larger, more heterogenous samples to provide stronger conclusions about the accuracy of THIM than the present studies.

Conclusion

The two current studies aimed to develop (Study 1) and provide preliminary evidence for the accuracy (Study 2) of the THIM wearable device for estimating sleep and wakefulness. With an independent sample in Study 2, THIM had slightly lower sensitivity compared to the findings with the algorithm training sample in Study 1. However, specificity remained relatively high compared to other actigraphy devices. The studies also examined whether THIM performed comparably to two popular actigraphy devices: the Actiwatch and Fitbit devices. In these preliminary studies, it appears that THIM is relatively similar in accuracy for estimating sleep and wake compared to the Actiwatch and Fitbit devices. However, THIM showed a tendency towards underestimating sleep and overestimating wakefulness. Whilst these studies did not find differences in the accuracy of actigraphy devices between good and poor sleepers, there was high variability in the devices' accuracies between individuals that could be explored in future research. The accuracy of THIM for estimating sleep and wake in individuals with insomnia could also be explored to further the long-term goal of improving the treatment of insomnia.

Chapter 6: The Accuracy of the THIM Wearable Device for Monitoring Sleep and Wake over Multiple Nights.

Hannah Scott[1], Nicole Lovato[2] and Leon Lack[1,2]

[1] College of Education, Psychology and Social Work, Flinders University, Adelaide 5001, Australia.

[2] College of Medicine and Public Health, Adelaide Institute for Sleep Health: A Flinders Centre of Research Excellence, Flinders University, Adelaide 5001, Australia.

Author contributions: HS contributed to study design, data collection, analysis and interpretation of the data, and manuscript preparation. LL and NL contributed to conception and study design, interpretation of the data and drafting of the manuscript.

Conflict of interests: The research costs were partially funded by Re-Time Pty Ltd, the company that sell the THIM device. None of the study authors were financially supported by Re-Time for this project. LL is a shareholder of Re-Time Pty Ltd. LL and HS have a patent pending regarding the THIM device. NL has no conflicts of interest to declare.

Abstract

In-laboratory validation studies of actigraphy devices have typically been performed on one night of data. The accuracy of actigraphy devices over multiple nights is largely unknown, particularly for consumer sleep trackers. The aim of this study was to evaluate the accuracy of the THIM wearable device for estimating sleep and wakefulness over three nights compared to polysomnography.

Twenty individuals slept overnight in the sleep laboratory on three nights. The accuracy of THIM was assessed by conducting epoch by epoch analyses and examining agreement with polysomnography on sleep parameters. The degree of accuracy of THIM against polysomnography (PSG) was compared between nights.

THIM demonstrated consistently high sensitivity (0.89) and specificity (range: 0.42-0.59) across the three laboratory nights. There were no significant differences between PSG and THIM for estimations of total sleep time on any night, $p > .23$. However, THIM overestimated WASO compared to PSG on Night 1 (*M difference* = +23.88, *SD* = 36.66), p = .002, $d = 0.74$, and Night 2 (*M difference* = +23.03, *SD* = 20.78), $p = .004$, $d = 1.41$. THIM also underestimated sleep efficiency on all three nights, $p < .03$, $d > 0.54$. There were no significant differences between good and poor sleepers on any accuracy variable, $p > .14$.

The accuracy of THIM was consistent across nights, with the device showing a bias towards the overestimation of wakefulness. Future research could investigate the accuracy of wearable devices across multiple nights in a naturalistic environment to ascertain the usefulness of these devices for long-term monitoring in the home environment.

Keywords: wearable technology; consumer sleep technology; polysomnography; actigraphy; activity-monitor; night-to-night variability.

The accurate estimation of objective sleep in the home environment is essential towards advancing sleep science and healthcare. Yet, the objective long-term measurement of sleep is difficult to undertake in the home environment. It is desirable to monitor objective sleep over multiple nights to obtain an accurate assessment of individuals' sleep as one night of data may not be representative of their typical sleep. This can be due to first night effects (Agnew et al., 1966) or large night-to-night variability in sleep quality, as is particularly the case for individuals with some sleep disorders, such as insomnia (Buysse et al., 2010). The gold-standard method of objective sleep measurement, PSG, is impractical for use over multiple nights outside the laboratory setting. PSG requires expensive equipment and specialised individuals to setup and score the data. Whilst simplified ambulatory PSG devices with automatic scoring algorithms overcome many limitations, these devices are still expensive and cumbersome to monitor sleep over multiple nights. Individuals are also unlikely to adhere to wearing PSG devices over a long period of time due to discomfort and the time required to setup and remove these devices.

Actigraphy overcomes most of the disadvantages of PSG recording for long-term sleep measurement in the home environment. The actigraphy method involves estimating sleep and wakefulness by measuring limb movement (Ancoli-Israel et al., 2015; Sadeh & Acebo, 2002). Epochs with little/no movement are scored as sleep and larger movements are scored as wake. Thus, actigraphy devices are simple to score, require little action from the individual, are typically inexpensive and are practical to administer over long periods of time. Due to their ease of use and the substantial amount of valuable data that can be obtained from them, actigraphy devices have become a popular method of estimating sleep for researchers, clinicians and consumers (Bianchi, 2017). Current AASM clinical guidelines indicate

that research-grade actigraphy devices may be used to provide objective sleep metrics to assess and monitor treatment response under certain conditions (Smith et al., 2018). However, guidelines recommend against the use of consumer sleep trackers for similar purposes due to the lack of validation studies and regulation of the industry (Khosla et al., 2018). Nonetheless, sleep research has extensively utilised actigraphy devices. There is also a burgeoning market for consumer sleep trackers that estimate sleep, with consumers wanting to continuously monitor their sleep in an effort towards improving their health and wellbeing (Bianchi, 2017), although data from these devices may not always be helpful (Baker, 2020; Baron, Abbott, Jao, Manalo, & Mullen, 2017).

Despite their widespread use for long-term sleep monitoring, there is little empirical evidence available about the accuracy of actigraphy devices for estimating sleep over multiple nights. Validation studies with research and consumer actigraphy devices are largely conducted on single nights in the sleep laboratory (Evenson et al., 2015; Van den Water et al., 2011). Few studies have investigated the accuracy of actigraphy devices over multiple nights (Hamill et al., 2020; Marino et al., 2013; Paquet et al., 2007; Sanchez-Ortuno et al., 2010). These studies largely find little variability in the accuracy of actigraphy devices for estimating sleep and wakefulness across multiple nights of typical sleep. However, the accuracy of actigraphy devices can vary with varying sleep quality between nights (Paquet et al., 2007). These studies are also largely conducted with research-grade actigraphy devices, and therefore the accuracy of consumer sleep trackers for estimating sleep over multiple nights compared to PSG is largely unknown.

Additionally, night-to-night sleep can vary considerably within individuals, particularly for certain clinical populations such as those with insomnia (Buysse et

al., 2010). It is likely that the accuracy of consumer sleep trackers may differ across nights more so for some individuals (e.g. with poor sleep) than others. Therefore, it is important to understand the accuracy of consumer sleep trackers over multiple nights and between individuals varying in sleep quality, since these devices are currently being utilised to assess sleep patterns under these conditions. The aim of this study was to evaluate the night-to-night variability in the accuracy of a new consumer sleep tracker, the THIM device, for estimating sleep and wakefulness compared to PSG with both good and poor sleepers.

Method

Participants and Study Protocol

The study protocol has been described previously in Study 2, Chapter 3. Briefly, twenty-one healthy individuals with good or poor sleep as defined by scores on the Insomnia Severity Index ([ISI], good sleepers ISI < 7, poor sleepers ISI ≥ 7) were recruited for this study. Participants slept overnight in the sleep laboratory on three occasions, however one poor sleeper withdrew after the first night. See Table 6-1 for participant characteristics information. Night 1 was an adaptation night where participants slept from their typical bedtime until their typical wake up time. On the following night, Night 2, participants completed sleep onset trials as part of a larger research project with THIM unrelated to the current study (see Chapters 3 and 4). During these trials, participants repeatedly fell asleep but were woken up shortly thereafter for four hours, meaning that they achieved little sleep during this four-hour period. Participants then slept uninterrupted until they woke spontaneously in the morning. Night 3 occurred one week later and followed the same protocol as Night 2. Thus, participants experienced their typical sleep opportunity on Night 1 and experimentally restricted sleep opportunities on both Night 2 and Night 3.

Table 6-1. *Participant characteristics for good sleepers, poor sleepers and the whole sample.*

Characteristic	Good Sleepers (N = 10)	Poor Sleepers (N = 11)	Total Sample (N = 21)
Age, mean (SD), y	24.90 (5.28)	21.88 (4.29)	23.32 (4.92)
Sex, No. (%)			
Men	4 (40)	3 (27)	7 (33.33)
Women	6 (60)	8 (73)	14 (66.67)
BMI, mean (SD)	25.48 (3.97)	24.88 (3.95)	25.17 (3.87)
Lifestyle characteristics			
Weekly alcohol consumption, No. (SD)	1.80 (1.81)	1.27 (1.79)	1.52 (1.78)
Daily caffeine consumption, No. (SD)	1.85 (1.56)	1.91 (1.38)	1.88 (1.43)
Sleep characteristics			
ISI, mean (SD)	2.00 (1.15)	11.36 (3.83)	6.90 (5.56)
PSQI, mean (SD)	3.10 (1.73)	7.73 (3.50)	5.52 (3.61)
Habitual Bedtime, mean time (SD), min	22:45 (64.59)	23:08 (67.43)	22:57 (65.66)
Habitual Wake Time, mean time (SD), min	07:27 (61.27)	07:56 (68.54)	07:42 (65.51)
Habitual TST, mean (SD), hrs	8.05 (0.83)	7.09 (1.50)	7.55 (1.29)

BMI = body mass index, ISI = Insomnia Severity Index, N = sample size, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

On all three nights, PSG and THIM recorded sleep simultaneously. The accuracy of THIM compared to PSG on Night 1 has been discussed elsewhere (see Chapter 5), but this chapter will incorporate this data in order to examine the differences between the device's accuracy across all three nights.

THIM

As described elsewhere, THIM is a ring-like device worn on the index finger of

the dominant hand. The device's tri-axial accelerometer measures acceleration and stores an average value for each 30-second epoch. Participants started the sleep tracking function when they lay down in bed at lights out on Night 1 by tapping their index finger on which THIM was placed onto their thumb, twice in quick succession. On Night 2 and Night 3, the THIM sleep tracking function automatically began within several minutes after the last sleep onset trial. However, participants were already attempting to fall asleep (some had already fallen asleep) when the THIM sleep tracking function began. As such, the estimations of sleep onset latency (SOL) are misleading as the start of this period is not the start of their attempt to fall asleep, as is customary. Therefore, the THIM estimations of SOL will be presented but not discussed in great detail. Rather, the device's estimates of other sleep parameters will be the main consideration in evaluating the accuracy of the device.

To retrieve the THIM sleep tracking data, the data was transmitted via Bluetooth from the device to the THIM smartphone application (v 1.0.1, operated on an Apple iPhone 5s with an iOS 8.0 operating system). The data is then sent to cloud-based servers for scoring. The manufacturers of THIM, Re-Time Pty. Ltd., sent us the 30-second epoch data for analysis as this data is not currently accessible for download via the smartphone app.

Statistical Analysis

The accuracy of THIM compared to PSG across all three nights was assessed in accordance with proposed guidelines for device validation studies (de Zambotti et al., 2019; Depner et al., 2019). Epoch-by-epoch analyses were conducted to calculate the sensitivity (proportion of epochs that THIM scored as *sleep* when the individual was asleep according to PSG), specificity (proportion of epochs that THIM scored as *wake* when the individual was awake according to PSG)

and accuracy (proportion of correctly-scored epochs) separately for each participant and then averaged together. We then performed Linear Mixed Model (LMM) analyses to examine whether there were any significant differences on the sensitivity, specificity, and accuracy of THIM across the laboratory nights (IBM SPSS, v 23). All LMM analyses used a first-order autoregressive covariance structure with laboratory night as a fixed effect. Where appropriate, post hoc comparisons were conducted with the Bonferroni correction.

Estimations of sleep parameters were compared between THIM and PSG, including total sleep time (TST), SOL, wake after sleep onset (WASO), and sleep efficiency. LMM analyses were conducted to examine whether there were any significant differences between PSG and THIM (the fixed effect) for estimating these sleep parameters. A statistically significant main effect was further examined using Bonferroni adjusted pairwise comparisons. Bland-Altman plots illustrated the degree of discrepancy between PSG and THIM estimations (y axis) against PSG (x axis) on separate plots for each sleep parameter (Bland & Altman, 1986). These plots display the mean difference (bias), the line of best fit, and the limits of agreement (± 1.96 SD of the mean difference). The type of sleeper was subsequently included in the model to determine whether the accuracy of THIM across all three nights differed between good and poor sleepers.

Results

Missing Data

On laboratory Nights 1 and 2, one PSG recording failed due to technical error. On Night 2 and Night 3, one THIM recording is missing due to a technical error relating to internet access when retrieving the data.

Epoch by Epoch Analysis

Table 6-2 presents the descriptive statistics for the epoch-by-epoch analyses on each night. LMM analyses revealed no significant differences between nights on sensitivity, $F(2, 57) = 0.33$, $p = .97$, specificity, $F(2, 57) = 2.84$, $p = .07$, or accuracy, $F(2, 57) = 0.60$, $p = .55$.

Table 6-2. *Sensitivity, specificity and accuracy for THIM across all three nights.*

Variable	Night 1	Night 2	Night 3
Sensitivity, mean (SD)	0.89 (0.06)	0.89 (0.05)	0.89 (0.07)
Specificity, mean (SD)	0.59 (0.18)	0.42 (0.27)	0.47 (0.25)
Accuracy, mean (SD)	0.85 (0.06)	0.87 (0.05)	0.87 (0.07)

SD = standard deviation.

Sleep Parameter Estimations

Table 6-3 presents the descriptive statistics for the sleep parameter estimations from PSG and THIM across all three nights. A LMM determined no significant interaction between the device (PSG and THIM) and laboratory nights on SOL estimations, $F(2, 114) = 1.45$, $p = .24$, TST estimations, $F(2, 114) = 0.37$, $p = .97$, sleep efficiency estimations, $F(2, 114) = 0.45$, $p = .66$, or WASO estimations, $F(2, 114) = 0.73$, $p = .49$. Despite this, pairwise comparisons indicated significant differences between PSG and THIM for estimations of sleep efficiency across all three laboratory nights, $p < .03$, and for estimations of WASO across Night 1, $p = .002$, and Night 2, $p = .004$, but not Night 3, $p = .11$. The effect sizes between PSG and THIM for sleep efficiency estimations were large across Night 1, $d = 0.54$, Night 2, $d = 0.94$, and Night 3, $d = 1.00$. Similarly, the effect sizes between PSG and THIM for WASO estimations were large on Night 1, $d = 0.74$, and Night 2, $d = 1.41$.

Bland-Altman Plots

Figure 6-1 presents Bland-Altman plots for key sleep parameters on Nights 1,

2 and 3. Both TST plots show a bias for THIM towards underestimating TST. Yet, the LMM analyses above indicate no significant difference between THIM and PSG on TST for these nights. The WASO plots illustrate a similar pattern. The mean bias for THIM is positive, meaning that THIM slightly overestimates WASO compared to PSG, albeit not significantly on Night 3 according to the LMM analyses. The sleep efficiency plots further illustrated that THIM slightly underestimates sleep, evident by a mean bias below zero on the y axis.

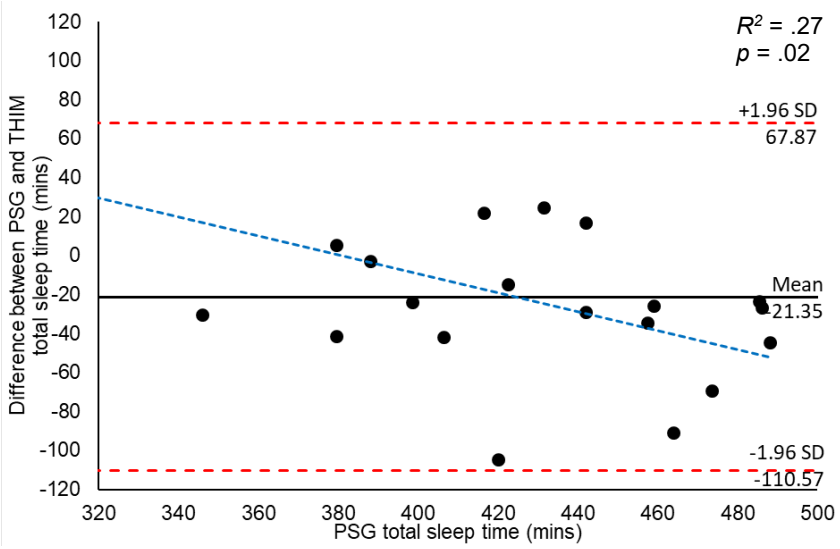
Table 6-3. *Sleep parameter descriptive statistics for PSG and THIM and the mean discrepancy between these two measures.*

Sleep Parameters	Night 1			Night 2			Night 3		
	PSG	**THIM**	**Difference**	**PSG**	**THIM**	**Difference**	**PSG**	**THIM**	**Difference**
SOL, mean (SD), min	21.68 (16.65)	19.15 (17.82)	-2.53 (14.84)	1.42 (3.25)	4.61 (5.39)	+3.19 (3.58)	1.21 (2.22)	6.82 (8.49)	+5.61 (8.06)
TST, mean (SD), min	424.60 (48.73)	403.25 (46.15)	-21.35 (45.52)	339.83 (56.85)	314.00 (53.85)	-25.83 (21.75)	287.55 (88.32)	269.95 (90.83)	-17.61 (13.63)
WASO, mean (SD), min	40.68 (34.50)	64.55 (30.36) *	+23.88 (36.66)	17.06 (9.75)	40.08 (20.96) *	+23.03 (20.78)	21.34 (14.14)	33.68 (22.20)	+12.34 (12.78)
sleep efficiency, mean (SD), %	87.36 (8.62)	82.93 (7.82) *	-4.44 (9.04)	94.74 (2.94)	87.59 (5.93) *	-7.14 (6.03)	92.83 (4.12)	86.36 (8.14) *	-6.48 (6.24)

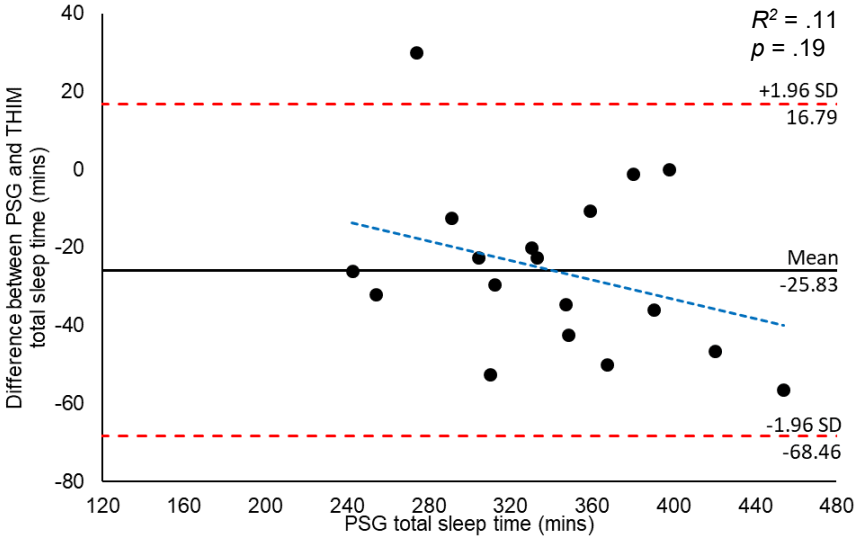
* $p < .05$ between PSG and THIM. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Total Sleep Time

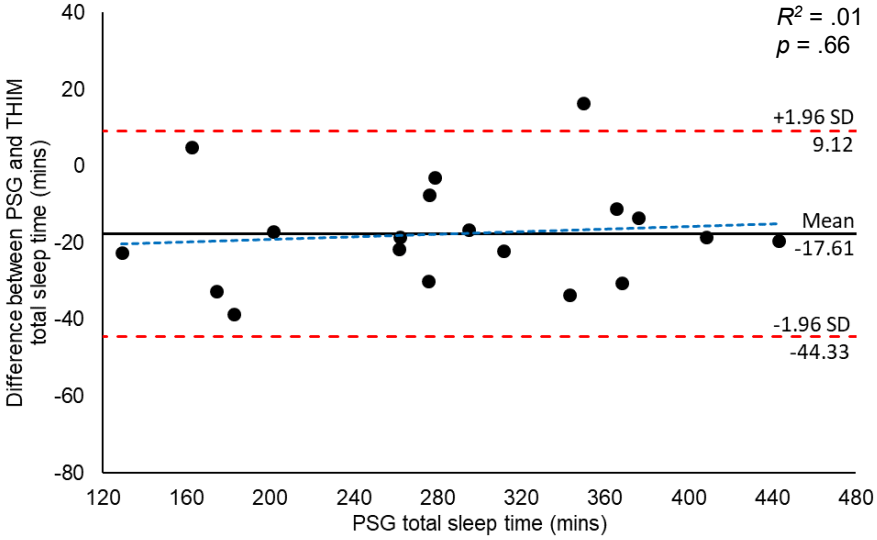
(a) Night 1



(b) Night 2

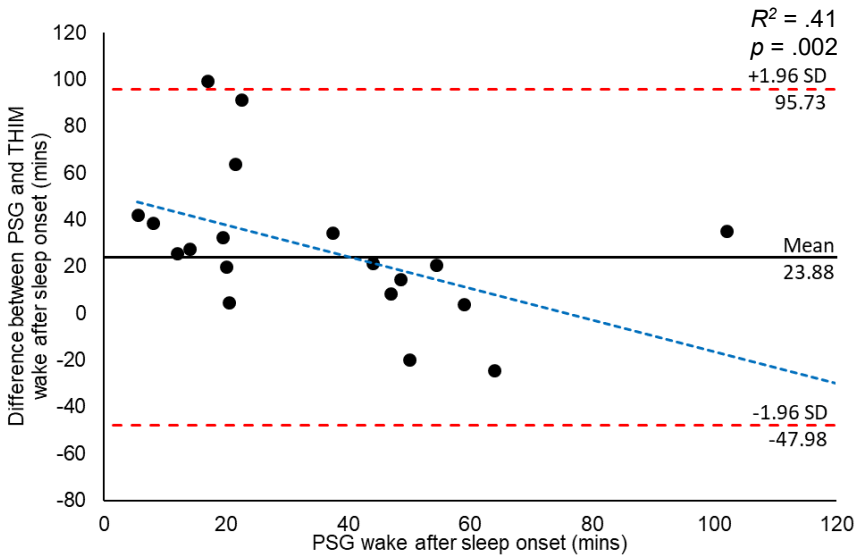


(c) Night 3

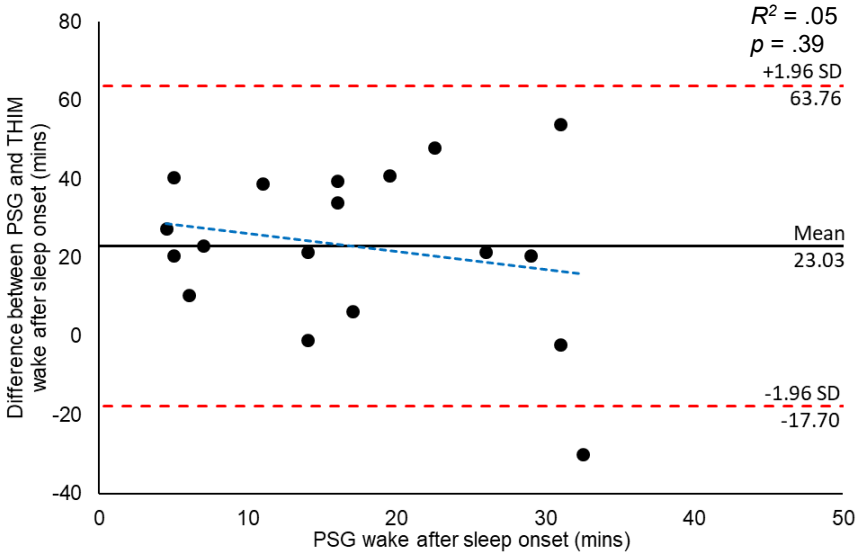


Wake After Sleep Onset

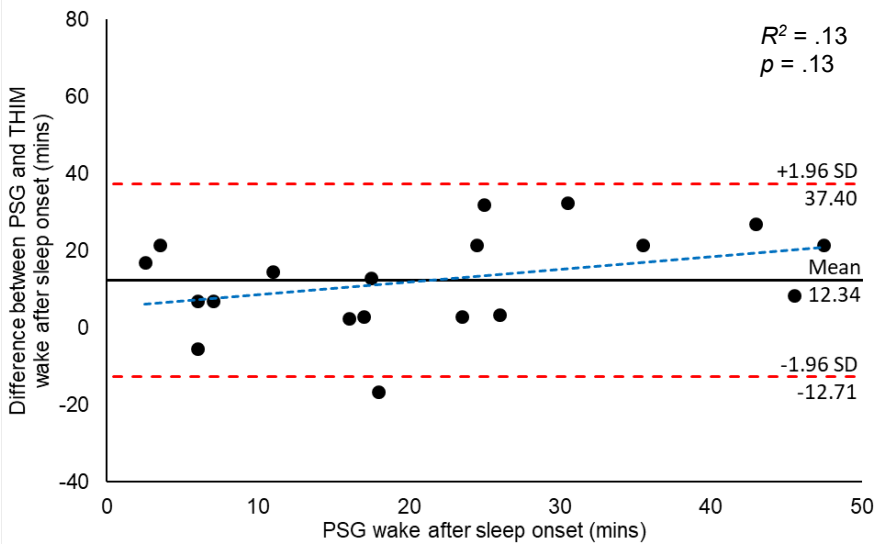
(a) Night 1



(b) Night 2

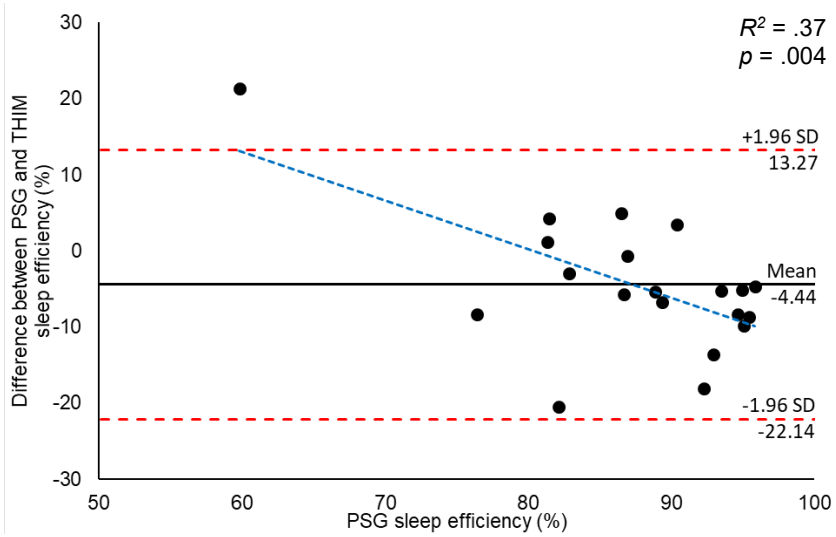


(c) Night 3

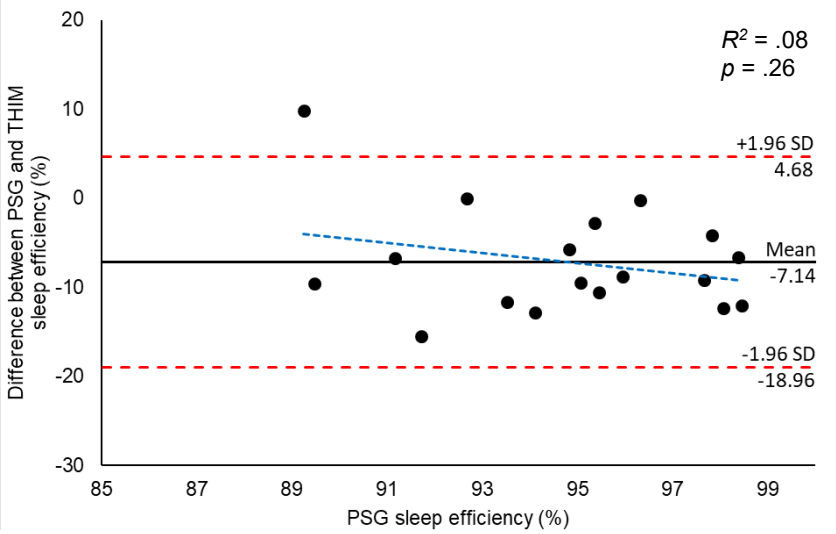


Sleep Efficiency

(a) Night 1



(b) Night 2



(c) Night 3

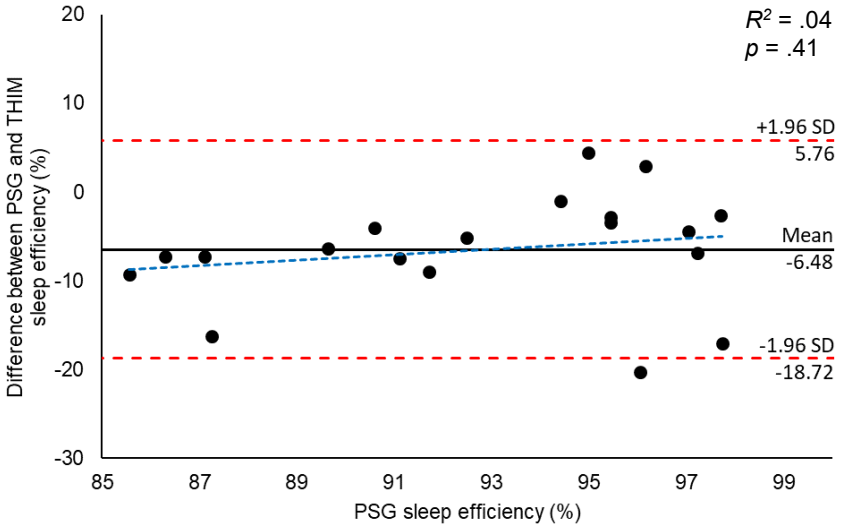


Figure 6-1. Bland-Altman plots showing the agreement between PSG and THIM on total sleep time, wake after sleep onset and sleep efficiency, separately for each night.
The solid black horizontal line indicates perfect agreement with PSG. The solid coloured horizontal line indicates the bias, the dashed coloured horizontal lines indicates the upper and lower limits of agreement, and the dotted coloured lines are the lines of best fit.

Good and Poor Sleeper Comparison.

Table 6-4 contains the descriptive statistics for sensitivity, specificity and accuracy on each night separately for good and poor sleepers. LMM analyses showed no significant differences between good and poor sleepers on any night. The interactions between the laboratory nights and the type of sleeper were also not statistically significant on sensitivity, $p = .56$, specificity, $p = .23$, or accuracy, $p = .77$.

The descriptive statistics for discrepancies between PSG and THIM determined sleep parameters are also present in Table 4. LMM analyses comparing mean discrepancies between PSG and THIM found no significant differences between nights or between good and poor sleepers. There were also no significant interactions between nights and sleeper type on SOL, $p = .91$, TST, $p = .14$, WASO, $p = .42$, or sleep efficiency, $p = .47$.

Table 6-4. *Epoch-by-epoch and sleep parameter descriptive statistics comparing good and poor sleepers.*

Variable	Laboratory Night		
	Night 1	**Night 2**	**Night 3**
Epoch-by-epoch analyses			
Sensitivity, mean (SD)			
Good sleepers	0.91 (0.05)	0.91 (0.06)	0.90 (0.08)
Poor sleepers	0.87 (0.07)	0.88 (0.04)	0.88 (0.06)
Specificity, mean (SD)			
Good sleepers	0.54 (0.16)	0.44 (0.27)	0.49 (0.23)
Poor sleepers	0.63 (0.19)	0.41 (0.28)	0.44 (0.29)
Accuracy, mean (SD)			
Good sleepers	0.87 (0.04)	0.88 (0.04)	0.87 (0.08)
Poor sleepers	0.83 (0.07)	0.86 (0.05)	0.86 (0.06)
Sleep parameters			
SOL discrepancy, mean (SD), min			
Good sleepers	-2.17 (13.56)	3.06 (4.65)	5.70 (5.63)
Poor sleepers	-2.82 (16.46)	3.33 (2.36)	5.50 (10.51)
TST discrepancy, mean (SD), min			
Good sleepers	-14.44 (24.75)	-21.33 (26.77)	-14.50 (14.97)
Poor sleepers	-27.00 (58.04)	-30.33 (15.59)	-21.06 (11.86)
WASO discrepancy, mean (SD), min			
Good sleepers	16.61 (17.54)	18.67 (25.03)	9.10 (14.92)
Poor sleepers	29.82 (47.13)	27.39 (15.76)	15.94 (9.47)
Sleep efficiency discrepancy, mean (SD), %			
Good sleepers	-2.84 (4.95)	-5.71 (7.73)	-4.93 (5.92)
Poor sleepers	-5.74 (11.46)	-8.58 (3.60)	-8.20 (6.48)

* $p < .05$ between good and poor sleepers with this device. SD = standard deviation, SOL = sleep onset latency, TST = total sleep time, WASO = wake after sleep onset.

Discussion

The aim of the current study was to evaluate the consistency in the accuracy of the THIM device for estimating sleep and wakefulness compared to PSG across

three nights. The epoch-by-epoch analyses indicated that THIM was consistent in its accuracy overall and for estimating sleep (sensitivity) and wake (specificity), with no significant differences found across the three laboratory nights. The interactions between the device (THIM and PSG) and laboratory nights were not significant for any sleep parameter, yet the pairwise comparisons indicated significant differences. These pairwise comparisons have greater statistical power to detect significant differences than the interactions, but solely interpreting them increases the risk of making a type I error due to multiple comparisons. For the purpose of the current study, this potential consequence was considered acceptable.

From the multiple comparisons, THIM estimations of SOL and TST did not significantly differ from PSG estimations on all three nights. However, THIM estimations of WASO were significantly higher than PSG on Nights 1 and 2, and a trend towards higher estimations on Night 3. This contributed towards significantly lower THIM estimations of sleep efficiency compared to PSG on all three nights. Additionally, the Bland-Altman plots illustrate little, mostly not significant, proportional bias across the three laboratory nights. These findings suggest that THIM remains accurate across typical and experimentally-restricted sleep. This study also found no significant differences between good and poor sleepers in the accuracy of THIM across all three laboratory nights. Overall, THIM performs consistently in its accuracy (sensitivity, specificity and accuracy) and estimations of sleep parameters for both good and poor sleepers across multiple nights of varying sleep opportunity (typical and experimentally-restricted).

To determine whether THIM may be useful for the long-term monitoring of sleep, an answer is required to the question, “Is the device accurate enough?” Yet, it is difficult to provide an appropriate answer to this question because the degree of

accuracy that is required to obtain useful data about sleep over multiple nights will depend upon its intended use. For diagnostic purposes that are typically conducted with PSG, THIM may not be 'accurate enough' because it fairly consistently produced overestimations of WASO and underestimations of sleep efficiency compared to PSG. For understanding long-term sleep duration of a large sample of individuals for research purposes, THIM may be considered 'accurate enough' as the accuracy of the device for estimating TST was comparable to PSG (underestimation of approximately 6%) and was consistent across nights. For consumers wanting to monitor their sleep patterns, the tendency of THIM to overestimate wake may cause unnecessary alarm and should therefore be used cautiously. Instead of concluding with an answer to the proposed question, the findings of the current study should be used by others as a guide to determine whether the device is 'accurate enough' for their particular purpose.

Having said this, it would be advantageous for all intended purposes to improve the accuracy of actigraphy devices for estimating objective wake time (Goldstone, Baker, & de Zambotti, 2018). Currently, THIM overestimates wakefulness across the night. This could be problematic if it leads to individuals seeking unneeded treatment on the basis of their sleep tracking data (Baron et al., 2017; Gavriloff et al., 2018). Modifications to the THIM sleep tracking algorithm should be made to prevent the device from overestimating wakefulness. Making any changes to the algorithm would require the accuracy of the device to be tested again with an independent sample to ascertain whether the changes reduce the wake bias.

It is important to consider the limitations of the current study when interpreting these findings. The sample size was relatively small compared to validation studies with other actigraphy devices, which limits the reliability of the findings. This is

particularly important for the comparison between good and poor sleepers as the study was not adequately powered to detect small differences in the accuracy of THIM between good and poor sleepers. Another limitation is the age range of participants, which was restricted to young, healthy adults. Considering that comorbidities and age may impact the accuracy of other actigraphy devices (de Zambotti, Baker, & Colrain, 2015b; de Zambotti et al., 2019) the findings of this study may not generalise to older individuals and/or to those with health conditions. Additionally, Night 2 and Night 3 were experimentally restricted sleep opportunities. On the one hand, these findings could generalise to restricted sleep opportunities, such as those experienced during behavioural insomnia treatments like sleep restriction therapy. On the other hand, the findings of the current study should not be extended to other types of sleep, including the fragmented sleep experienced by people with insomnia. The accuracy of THIM for estimating sleep and wakefulness across nights of disrupted/fragmented sleep is unknown and warrants further investigation.

With the long-term goal of using THIM to accurately monitor sleep with individuals over a long period of time, there are many aims to potentially address in future research. Firstly, the accuracy of THIM should be investigated in the home environment. Individuals' sleep in the artificial laboratory environment can substantially vary in quality compared to sleep in their bedroom environment (Edinger et al., 1997; Edinger et al., 2001). Relatedly, individuals' sleep with a full PSG montage can vary substantially compared to their sleep without PSG recording. Whilst the accuracy of THIM for estimating objective sleep should be compared to PSG (Depner et al., 2019), the use of simplified PSG devices may largely overcome discomfort (Svensson, Chung, Tokuno, Nakamura, & Svensson, 2019). Therefore,

the accuracy of THIM should be tested in participants' bedroom environments and compared against simplified PSG devices to ascertain whether THIM is suitable for monitoring sleep outside of the laboratory setting.

Secondly, the accuracy of THIM should be investigated in various populations, such as older individuals and those with health conditions and/or disordered sleep. THIM is intended to be used by individuals with insomnia, who are typically older than this study's sample. Nonetheless, no significant differences were found in the accuracy of THIM between good and poor sleepers in this study, so the device may be similarly accurate in an insomnia sample. To understand the utility of the THIM sleep tracking function for individuals with insomnia, the device should be tested with a sample representative of this population.

Thirdly, the accuracy of THIM may need to be tested over a longer period than tested in the current study for certain populations, i.e. 7-14 nights (Acebo et al., 1999; Van Someren, 2007). This is particularly important for monitoring individuals with high night-to-night variability in sleep, such as those with insomnia (Buysse et al., 2010). These three directions for future research would determine whether THIM could accurately monitor sleep and wakefulness in the home environment for individuals of varying demographics and sleep quality, rendering the device useful for a variety of clinical and research purposes including the management of insomnia.

Conclusion

This study investigated the consistency in the accuracy of THIM for estimating sleep and wakefulness compared to PSG across three nights in the laboratory. Whilst sensitivity, specificity, and accuracy remained high and did not vary significantly across the three laboratory nights, THIM estimations of WASO and

sleep efficiency significantly differed from PSG on some laboratory nights. Nonetheless, THIM performed comparably with both good and poor sleepers across all three laboratory nights. Whether THIM is accurate enough for monitoring sleep and wakefulness over multiple nights will depend on the intended purpose. Regardless, modifications should be made to THIM to improve its accuracy for estimating wake time. Future research could test THIM in the home environment with a larger sample of individuals with varying demographics, health conditions, and sleep quality. This is necessary to ascertain whether THIM will be useful for long-term sleep monitoring to assess, manage and treat individuals with insomnia.

Chapter 7: General Discussion

Overview

The broad aims of this dissertation were to develop and test the accuracy of the THIM device for, firstly, estimating sleep onset using behavioural responses to external stimuli and, secondly, for monitoring sleep and wakefulness overnight using actigraphy. The purpose was to create a wearable device that could appropriately and accurately administer Intensive Sleep Retraining (ISR), and passively monitor sleep for its intended use. This chapter summarises the findings of this dissertation and discusses the significance and implications of the findings for the conceptualisation of the sleep onset period and the clinical treatment of insomnia. Methodological considerations and limitations of this dissertation will be presented as well as directions for future research with THIM and its use for administering ISR for the treatment of insomnia.

Summary of Dissertation Findings and Original Contribution to Knowledge

Chapter 2: Systematic review of wearable devices

Chapter 2 systematically identified studies that examined the accuracy of wearable devices for estimating sleep onset latency (SOL) compared to polysomnography (PSG). The aim of this review was to determine whether any currently available wearable devices are suitable for administering ISR in the home environment. Whilst previous reviews summarised the accuracy of wearable devices for estimating sleep (Evenson et al., 2015; Van den Water et al., 2011), this review was the first to specifically focus on the accuracy of SOL estimations. A focused review was warranted, considering that the accurate measurement of SOL is crucial for various research and clinical purposes, including the administration of ISR, power naps, and objective daytime sleepiness tests outside of the laboratory setting.

Although the reviewed actigraphy devices produced estimations of SOL that were often not significantly different from PSG, there was large interindividual variability depending on, but not entirely explained by, participant characteristics. Actigraphy devices are therefore not suitable for purposes dependent on accurate estimation of sleep onset because they do not achieve the required degree of accuracy across all individuals. This is particularly the case for individuals with sleep disorders, who would benefit the most from the clinical applications. As predicted, electroencephalography-based (EEG) devices produced more accurate and less variable estimates of SOL. However, these devices are expensive, require trained personnel to operate, and rarely produce sleep data in real-time as scoring largely occurs retrospectively. This limits their usefulness for the purposes of ISR, power naps, and daytime diagnostic tests. The review concluded that devices measuring behavioural sleep onset were most suitable for the administration of ISR because they consistently overestimated PSG-determined SOL, which is more suitable for administering ISR than a device that underestimates SOL. These devices also showed less variability in their accuracy across individuals than other wearable devices. This finding justified the stimulus-response method that THIM relies upon to estimate sleep onset for the purposes of ISR.

Chapter 3: Development of THIM for sleep onset latency detection

Chapter 3 discussed the development and refinement of the THIM device for estimating SOL in comparison to PSG. The findings about the accuracy of the initial algorithm in Study 1 informed the refinement of the algorithm, which was subsequently tested in Study 2. These two studies made an original contribution to knowledge as they were the first to test the accuracy of the novel THIM device. They were also the first to assess the use of behavioural responses to external stimuli for

sleep onset detection using minimal intensity vibratory stimuli. This was significant because THIM showed much closer agreement to PSG than other similar devices that use auditory stimuli and larger hand/wrist movements as behavioural responses. With the final version of the algorithm, THIM overestimated PSG-SOL by < 1 minute on average on both testing nights compared to similar devices that had discrepancies of 2-3 minutes compared to PSG (Mair, 1994; Scott et al., 2018). Importantly, THIM remained accurate after repeated use and was similar in accuracy across good and poor sleepers. It was concluded that THIM was accurate enough for the purpose of administering ISR as its slight overestimation of sleep onset means that individuals would achieve a similar sleep duration during each sleep onset trial as those whose insomnia was effectively treated during laboratory-based ISR studies (Harris et al., 2012; Harris et al., 2007). Thus, the THIM administration of ISR aligns with the laboratory-based protocol in this aspect. Nonetheless, this needs to be confirmed with a sample of individuals with insomnia to ensure that THIM can estimate sleep onset to a similar degree of accuracy with the target population for ISR.

Chapter 4: Quantitative EEG analysis of sleep microstructure whilst responding to vibratory stimuli

Chapter 4 further investigated the data from Chapter 3 (Study 2). A quantitative electroencephalography (qEEG) analysis was performed during the THIM-administered sleep onset trials to characterise sleep microstructure using a finer-grained analysis than traditional EEG sleep staging. The aim was to examine the correspondence between sleep microstructure and responses to the vibratory stimuli emitted by THIM during the sleep onset period. This study was the first to perform a qEEG analysis on responses to vibratory stimuli during the sleep onset period, with previous research typically using auditory stimuli (Colrain, Di Parsia, &

Gora, 2000; Cote, De Lugt, & Campbell, 2002; Cote, Etienne, & Campbell, 2001; Harsh et al., 1994). The findings indicated increases in higher frequency brainwaves when participants responded to the vibratory stimulus compared to when they did not respond across all sleep stages. This suggests that a shift to wakefulness or an arousal occurs prior to or coincident with the vibratory stimulus. A lack of response to the stimulus was associated with increases in delta activity, signalling greater sleep depth. Together, these findings illustrate that during N1 sleep, participants consistently exhibited both wake and sleep-like physiological (shifts to higher EEG frequencies) and behavioural responses (probability of responding or not responding to the stimulus) to vibratory stimuli. This implies that the 30-second epochs scored as N1-sleep, in which approximately 54% of vibratory stimuli evoked a behavioural response, is a transitional state between wake and sleep that contains both sleep-like behaviour and brief arousals indicative of wakefulness.

Chapter 5: Development of THIM sleep tracking function

Chapter 5 tested the THIM sleep tracking function and its accuracy for estimating sleep and wakefulness compared to PSG. The device uses actigraphy to passively estimate sleep and wakefulness during the sleep period. Whilst the accuracy of this method is well-known, this study made an original contribution to knowledge as it was the first to test the accuracy of the THIM device and its novel algorithm. Aside from a different actigraphy algorithm, THIM also differs from most common sleep trackers because it is worn on the index finger as opposed to the wrist. The placement and algorithm of actigraphy devices greatly impact their accuracy (Kim et al., 2013; Quante et al., 2018; Slater et al., 2015; Zinkhan et al., 2014), and finger-worn actigraphy devices may be more accurate than wrist-based devices (de Zambotti et al., 2017). Therefore, it was warranted to investigate the

accuracy of THIM specifically and we predicted that the device would be more accurate than wrist actigraphy devices.

Contrary to our predictions, THIM had similar accuracy for estimating sleep and wake compared to the Actiwatch and Fitbit devices. However, THIM showed a greater tendency to underestimate sleep and overestimate wakefulness (which may have been significant with a larger sample size), which may be due to the placement of the device or the algorithm, or both. No differences were found in the accuracy of THIM between good and poor sleepers, which may extrapolate to individuals with insomnia but this would require confirmation in future research. There was high unexplained variability between individuals, as is common with actigraphy devices. This is an area that warrants further investigation to improve the usefulness of actigraphy devices for sleep monitoring.

Chapter 6: Accuracy of THIM sleep tracking over multiple nights

Chapter 6 assessed the variability in the accuracy of the THIM sleep tracking function for estimating sleep and wakefulness during three laboratory nights. This was an important contribution to knowledge because sleep trackers intended use is over multiple nights, not just one night. Yet, few studies have investigated the accuracy of actigraphy devices compared to PSG across multiple nights (Hamill et al., 2020; Marino et al., 2013; Paquet et al., 2007; Sanchez-Ortuno et al., 2010). Since sleep quality varies between nights for some individuals more than others, such as those with insomnia (Buysse et al., 2010), understanding the variability in the accuracy of sleep trackers across multiple nights is necessary.

The THIM device showed consistently high sensitivity, specificity and accuracy compared to PSG across the three laboratory nights comprising of a typical sleep opportunity (Night 1) and experimentally-restricted sleep opportunities (Nights

2 and 3). Additionally, there were no significant differences between good and poor sleepers in the agreement between THIM and PSG for all outcome variables. This suggests that THIM notably maintains a degree of accuracy across types of sleepers and across typical and reduced sleep opportunities. However, THIM produced consistently and significantly lower estimations of sleep efficiency due to higher estimations of wake after sleep onset across all three nights. This contrasted the findings of Chapter 5 due to the greater statistical power of these analyses, although the risk of making a type I error was greater in this study due to the interpretation of multiple comparisons. Therefore, whilst THIM was reliably accurate for good and poor sleepers across a typical and experimentally-restricted nights of sleep, the device consistently performed at a sub-optimal level for estimating wakefulness. The improvement of the accuracy of THIM for estimating wake is thus required to render the device useful for many purposes and may be achieved with modifications to the algorithm, the addition of physiological sensors or the extension of the stimulus/response method of sleep/wake determination, as discussed further below.

Theoretical Implications of Dissertation Findings: When does wake end, and sleep begin?

The goal of Chapters 3 and 4 was to quantify the discrepancy between PSG-sleep onset and THIM-derived sleep onset, yet the findings had unexpected theoretical implications. Contrary to findings with auditory stimuli, responsiveness to minimal intensity vibratory stimuli aligned closely with PSG-sleep onset, as shown in Chapter 3. Analysis of the sleep microstructure using qEEG elucidated this finding from Chapter 4. Responses to vibratory stimuli were frequent during wake, rare in N2 sleep, and somewhere in-between for N1 sleep. Overall, the theoretical findings of responsiveness to vibratory stimuli are similar to those found with auditory stimuli.

This research provides further support for the conceptualisation of N1 sleep as a transitional sleep/wake period. Sleep onset is not a definitive point in time. Rather, it is a transitional period that begins with quiet wakefulness, contains rapid fluctuations between sleep and wake, and if left undisturbed, ultimately ends with more continuous sleep. It involves many physiological and psychological changes including in respiration, heart rate, muscle activity, vigilance, memory consolidation and responsiveness to auditory stimuli (Ogilvie, 2001). The change in responsiveness to tactile stimuli can now be added to this list.

The next promising step in this research is to explore localised brain activity with high density EEG during the sleep onset period whilst participants respond to vibratory stimuli. During the sleep onset period, there is an overall reduction in alpha and increase in theta brain waves. Yet, this change is not uniform across all brain regions: some cortical brain regions display this change earlier during the sleep onset period than other regions (Ferrara & de Gennaro, 2011). Fernandez Guerrero and Achermann (2019) examined localised changes in brain activity during the sleep onset period. Of particular importance to the current study is the finding that the postcentral gyrus (Brodmann area 3, the primary somatosensory cortex) exhibited higher sigma and delta power than other areas of the brain at the beginning of the sleep onset period (early N1 sleep). Therefore, the brain region that processes vibratory stimuli appears to fall asleep sooner after N1-sleep onset than other measured brain regions, including the primary auditory cortex that processes auditory stimuli (Brodmann areas 41 and 42). It would therefore be beneficial to examine the correspondence between responsiveness to vibratory stimuli and localised brain activity. Furthermore, responsiveness to different stimulus types (vibratory and auditory) could be investigated to compare when different processing

pathways are inhibited during the sleep onset period.

Of particular interest for the use of THIM is whether responsiveness to vibratory stimuli occurs similarly for individuals with insomnia. The neurocognitive model for insomnia includes greater cortical arousal as a conditioned factor of the disorder (Perlis et al., 1997). Yet, differences in EEG spectral power have not been consistently observed between good sleepers and those with insomnia (Buysse et al., 2008; Perlis, Merica, Smith, & Giles, 2001; Spiegelhalder et al., 2012; St-Jean, Turcotte, Pérusse, & Bastien, 2013). Whether this theorised, yet inconsistently observed, greater cortical arousal extends to heightened responsivity to stimuli in the external environment is yet to be investigated. There is evidence of individual variability in the correspondence between sleep microstructure and behavioural indications of sleep onset that are currently unexplained in the literature. Prerau et al. (2014) combined qEEG with behavioural measures of sleep to characterise the probability of individuals being awake during the sleep onset period. The authors found close temporal alignment between the cessation of correct responses on a behavioural task and drop-out in alpha power. However, some individuals continued to behaviourally respond well after alpha drop-out, and until increases in delta and theta power occurred.

In Chapter 4, some individuals responded to a considerable number of stimuli during N1 and N2 sleep. We theorised that these differences may depend on whether participants were good or poor sleepers. Yet, we found no significant differences on sleep microstructure when participants responded or did not respond to vibratory stimuli during any sleep stage between good sleepers and poor sleepers (as defined by scores below or above 7 on the Insomnia Severity Index [ISI], see Table A6 in the Appendix). Future research could examine the sleep microstructure

of good sleepers compared to individuals with insomnia to determine whether responsiveness to vibratory stimuli during sleep depends on sleep characteristics. This may have important theoretical and clinical implications for the treatment of insomnia, since abnormally sensitive/acute responsivity and information processing when attempting sleep is a theorised neurocognitive factor of insomnia (Perlis et al., 1997). Furthermore, it will be important to assess whether patients with insomnia respond similarly to good sleepers to determine whether THIM will remain as accurate at estimating sleep onset in this population.

Clinical Implications of Dissertation Findings

THIM-administered Intensive Sleep Retraining in the Home

The findings of this dissertation have potential clinical implications for the treatment of insomnia. Chapters 3 and 4 showed that the THIM device can administer ISR appropriately by precisely estimating when sleep onset occurs to wake individuals up after a very short period of sleep. These findings demonstrate THIM-administered ISR aligns closely with the laboratory-based protocol. The next step is to test the efficacy of THIM-administered ISR in the home environment.

The main question remaining is whether THIM-administered ISR will be successful in the home environment. This raises two potential issues. The first uncertainty lies in whether patients will correctly use THIM and comply with the device's ISR instructions in the uncontrolled home environment without the aid or supervision of trained staff. As sleep pressure builds across the ISR treatment period, the temptation to deviate from instructions and prematurely end treatment will increase. This is an issue for all home-based treatments of insomnia that may be minimised with cognitive techniques, including motivational therapy, psychoeducation about homeostatic sleep drive and reframing sleepiness as a

positive experience indicative of treatment efficacy.

The second uncertainty is whether THIM-administered ISR will produce a therapeutic effect for sleep onset insomnia when used in the home environment. Mair et al. (2020) tested whether a home-based application of ISR can be effective using a phone-based application called Sleep On Cue (SOC) to administer ISR at-home with twelve patients diagnosed with insomnia. Compared with their ISI scores at baseline (M = 21.08, SD = 3.37), patients with insomnia had reduced insomnia severity at post-treatment (M = 14.18, SD = 6.10) and at four-week follow-up (M = 14.08, SD = 5.04). Patients also adequately adhered to the SOC-administered ISR protocol, suggested by actigraphy data that showed peaks in activity at the end of each sleep onset trial. They were inferred to correspond with patients' compliance to the ISR instruction to get out of bed in-between trials during the overnight treatment session. However, this study did not have a control group so the observed improvements in sleep and daytime functioning outcomes cannot necessarily be attributed only to the treatment. Nonetheless, treatment outcomes were comparable to those found with the waitlist controlled laboratory-based ISR procedure (Harris et al., 2012; Harris et al., 2007). Whether home-based ISR will produce similarly sustained therapeutic effects over time as the laboratory-based procedure could be a topic investigated in future research.

Since the THIM-administered ISR procedure operates similarly to the phone-based procedure, it is predicted that the device will successfully administer ISR in the home environment and improve sleep quality and daytime functioning for individuals with insomnia. A study to test this prediction is ongoing, however financial and time restrictions meant that this further study was not feasible within the PhD timeframe. If THIM-administered ISR is effective for treating insomnia, then the

device will enable the practical administration of this behavioural treatment in the home environment, improving the public's accessibility to an effective behavioural treatment of insomnia.

Combining THIM-administered ISR with other insomnia treatments

If THIM can effectively treat insomnia, it could become an adjunct tool for clinicians treating insomnia. Harris et al. (2012) found that whilst the laboratory-based ISR protocol alone was as effective for treating insomnia symptoms as stimulus control therapy (SCT), the combination of ISR followed by SCT was particularly effective for reducing insomnia severity. ISR was theorised to have provided a 'kick-start' to treatment effectiveness such that when the patient underwent SCT, their insomnia had showed signs of improvement (Harris et al., 2012). This meant that the daytime sleepiness and the associated challenges experienced during the first 3-4 weeks of SCT were avoided, or at least mitigated. Therefore, the combination of ISR and SCT may have been particularly efficacious due to the additive effects of these two treatment components and/or due to greater treatment adherence by starting SCT with less severe insomnia after ISR (Lack et al., 2019). Either way, combining THIM-administered ISR with SCT or another behavioural treatment such as sleep restriction therapy (SRT) as part of cognitive behavioural therapy for insomnia (CBT-I) may lead to better treatment outcomes than CBT-I alone, which can be tested in future research. Additionally, if THIM-administered ISR 'kick-starts' treatment efficacy, then remission may be achieved with a reduced number of CBT-I treatment sessions. This would reduce public healthcare costs and burden on clinicians, since there is a shortage of CBT-I specialists in Australia and, for that matter, worldwide (Thomas et al., 2016). Future research could conduct a dose-response study of CBT-I treatment sessions following

THIM-administered ISR to test this idea.

There is also the potential to combine THIM-administered ISR with digital CBT-I programs. When CBT-I administered by a qualified specialist is unavailable, digital CBT-I interventions are a viable solution that are inexpensive, and most importantly, lead to meaningful improvements in insomnia, health and wellbeing. In a large randomised controlled trial (N = 164), greater improvements in sleep and daytime functioning were reported for a digital CBT-I intervention, Sleepio, compared to a waitlist control group of mild-moderate insomnia cases (Espie et al., 2012). Effects sizes at two-month follow-up were substantial for key sleep outcomes, including sleep efficiency (*d* = 1.37), total wake time (*d* = 1.21), and SOL (*d* = 0.80, Espie et al., 2012). Like clinician-administered CBT-I, individuals could receive THIM-administered ISR before commencing digital CBT-I to kick-start therapy. Individuals would presumably experience better sleep than normal during the first weeks of digital CBT-I. This should make adherence to the behavioural therapy instructions easier, increasing treatment adherence and thus efficacy. Therefore, the combination of THIM with digital CBT-I would be expected to produce greater improvements in sleep and daytime functioning in a greater proportion of patients than with either treatment alone. Alternatively, THIM-administered ISR may be effective enough at alleviating the insomnia for individuals with mild cases, causing them to withdraw from digital CBT-I. Either way, future research could investigate whether combining THIM-administered ISR with digital CBT-I would lead to better treatment outcomes and adherence than to the stand-alone treatments.

Another promising avenue for insomnia treatment is combining ISR and CBT-I with circadian rhythm science to treat insomnia. For patients with difficulties initiating sleep (sometimes referred to as sleep onset insomnia), a circadian phase delay may

be exacerbating their insomnia symptoms because they are attempting sleep at a non-optimal circadian time, making it more difficult to fall sleep (Lack, Wright, & Paynter, 2007; Lack & Wright, 2007; Morris, Lack, & Dawson, 1990). Under these conditions, THIM-administered ISR and CBT-I could be combined with early morning bright light therapy, which would be administered after circadian core temperature nadir. The THIM-administered ISR may fully or partially extinguish the conditioned insomnia arousal response, tailored CBT-I would presumably address the perpetuating factors of the insomnia, and morning bright light therapy would correct the circadian misalignment component by phase advancing the rhythm to facilitate easier/earlier sleep onset on subsequent evenings. Our previous findings suggest that administering one week of morning bright light therapy is effective at phase advancing and improving sleep quality for individuals with sleep onset insomnia (Dubiel, 2019; Lack et al., 2007). With simple administration methods for ISR (THIM) and bright light therapy (portable light devices), clinicians are now able to easily combine these therapeutic techniques, to their patients' benefit.

THIM sleep tracking as an assistive tool for CBT-I

The THIM sleep tracking function may also be useful for insomnia treatment. Individuals may struggle to consistently maintain a sleep diary in the long-term. THIM may be useful in lieu of sleep diaries, such as to record time in/out of bed to monitor adherence to behavioural therapies. Additionally, THIM may provide more accurate data about non-adherence to therapy instructions than sleep diaries, as found with actigraphy devices by Carney, Lajos, and Waters (2004). THIM may also be useful for titrating sleep restriction by monitoring sleep efficiency. THIM could input the sleep tracking data either through a validated algorithm or relay these data to the clinician to produce tailored sleep restriction instructions for the individual. This

would enable closer monitoring of insomnia patients than is currently feasible with paper-based sleep diaries collected form patients at each CBT-I session. It is important to note that this technique would not be appropriate with patients with significant paradoxical insomnia contributing to the condition, as the clinical goal would be to treat subjective sleep perceptions measured via sleep diaries, not objective sleep estimated with THIM.

For individuals with substantial paradoxical insomnia, an accurate THIM sleep tracking function may have therapeutic benefits. Individuals could monitor their sleep at home with THIM for many nights and review their data, with guidance from a clinician to aid interpretation, to understand discrepancies between their sleep as measured by THIM and their perceptions of sleep. A daily sleep diary could be incorporated into the THIM smartphone application to assist with the recording of patients' perceptions of sleep. Data could then be downloaded by patients and physicians for use in treatment. This treatment technique has shown small-moderate therapeutic benefits when patients were presented with PSG data (Downey & Bonnet, 1992; Tang & Harvey, 2006). Whether this technique would be successful using the THIM sleep tracking function is a potential direction for future research.

These potentially therapeutic techniques would require THIM to be accurate for monitoring sleep and wakefulness in individuals with insomnia. However, the findings of this dissertation are unable to shed light on whether THIM will be accurate for people with insomnia. THIM remained accurate at tracking sleep and wakefulness for both good and poor sleepers, and across typical and experimentally-restricted nights of sleep, despite consistently overestimating wakefulness. Yet, we did not explore the accuracy of THIM for tracking the fragmented sleep (sleep characterised by a high number and duration of awakenings) that people with

insomnia typically experience, particularly those experiencing difficulty maintaining sleep. Instead, participants largely experienced consolidated sleep during this study, although restricted on Nights 2 and 3. THIM may be less accurate at estimating sleep and wakefulness across a fragmented sleep period as it may 'miss' brief wake periods, reducing specificity and overall accuracy. Therefore, the accuracy of THIM for estimating sleep and wakefulness with insomnia patients should be investigated before testing the utility of the sleep tracking function for treating chronic insomnia. Modifications to THIM to improve its accuracy for estimating wake is all the more imperative to develop a device that could accurately track the fragmented sleep of insomnia patients.

Methodological Considerations

Methodological considerations are discussed throughout this dissertation where appropriate. Important points to consider when interpreting the findings and implications of this dissertation are discussed here. In terms of study limitations, the samples sizes of each study were too small to be adequately powered to detect small differences between groups (i.e. between good and poor sleepers). This is particularly a concern when considering how the accuracy of THIM for monitoring sleep and wakefulness overnight differs between individuals. Research with larger sample sizes is required to determine whether individual characteristics can explain the variability in the accuracy of THIM across individuals. Furthermore, the accuracy of THIM for estimating sleep onset and monitoring sleep and wakefulness overnight was not tested in patients with insomnia. Therefore, caution must be taken when considering the implications of this research for the management and treatment of insomnia with THIM. Additional research with a sample of patients with insomnia is required to confirm whether THIM can successfully perform its intended functions

with this population, and whether this leads to the anticipated therapeutic benefits.

A methodological strength of this dissertation was the comparison of THIM against the gold standard measure for measuring objective sleep, PSG, as is recommended in guidelines for validation studies of wearable devices (Depner et al., 2019). These studies were also conducted in the controlled laboratory environment, which was particularly important for confirming that participants adhered correctly to THIM's instructions during ISR. However, laboratory-based testing is not representative of home-based settings and so the accuracy of THIM must be confirmed in the home environment. The choice of testing the accuracy of THIM with a sample of varied sleep quality could also be considered a strength because it allowed us to evaluate THIM's performance with a sample that represented the general population, as opposed to validation studies conducted with healthy, good sleepers that often fail to replicate with more representative samples (Van den Water et al., 2011). Additionally, whilst the sample sizes of the studies discussed in this dissertation are too small to achieve adequate statistical power for the detection of between-group differences, many datapoints were obtained per individual which provided greater power to detect within subject differences. This was particularly the case for Chapter 3 which tested the accuracy of THIM for estimating SOL and allowed us to evaluate variability in the device's performance within individuals: a consideration that is typically overlooked in validation studies of wearable devices.

Directions for Future Research

Possible directions for future research are discussed extensively throughout the dissertation, with the broad aim of incorporating THIM into the treatment and management of insomnia. Additional directions for future research are discussed below.

Future research into Intensive Sleep Retraining

Therapeutic components to ISR

Now that THIM has enabled the practical administration of ISR, experimental studies to further investigate and refine this treatment are easier to conduct. A theoretical research question with potentially important implications for the treatment of insomnia is what makes ISR effective? Our group has theorised that the mechanism of action of ISR is the extinguishing of the conditioned cortical arousal response that perpetuates sleep onset difficulties by the re-establishment of the bedroom environment and intention to sleep with rapid sleep onset instead of wakefulness. This mechanism is common with other behavioural treatments, such as SCT that reduces time spent awake in bed over many weeks to reinforce sleep. However, there are other potentially therapeutic components to ISR that may substantially contribute to its efficacy.

One potentially therapeutic component of ISR is near-total sleep deprivation. Patients experienced one night of sleep deprivation (except for the very brief sleep episodes that they experienced after sleep onset during the trials) during the laboratory-based ISR studies (Harris et al., 2012; Harris et al., 2007). This increased homeostatic sleep drive on the following day and would have contributed to the robust recovery sleep that patients experienced the following night. Sleep pressure may not have been entirely alleviated during the recovery sleep and may have remained higher than normal over multiple nights following treatment, contributing to the improved sleep quality observed in the week after treatment. Increased homeostatic sleep drive is a necessary component to other behavioural treatments for insomnia, such as SRT. The difference between ISR and SRT is that sleep pressure was rapidly increased over one night compared to relatively slowly increased over many weeks. Therefore, this component of ISR is expected to have

some therapeutic benefit – the degree to which this component is responsible for the efficacy of ISR could be investigated in future research.

Similarly, another potentially cognitively therapeutic component of ISR is patient reassurance on multiple occasions that their sleep mechanism is not 'broken', and they retain the ability to sleep. A randomised controlled trial could be conducted with each condition largely representing one of these theorised therapeutic components of ISR. This trial could test ISR as it is currently conducted (sleep retraining condition) compared to a night of sleep deprivation (homeostatic sleep drive condition) and to a protocol where participants receive feedback after each sleep onset trial that they are falling asleep quickly to enhance the cognitive benefits of ISR (the cognitive reassurance condition). Such a study will be practical to conduct with THIM, once its performance has been tested with individuals with insomnia. Understanding the mechanism of action of ISR may result in alterations to the treatment instructions to enhance its effectiveness for treating insomnia.

One night versus many nights of ISR

One practical question about the administration of ISR is whether sleep onset trials need to be conducted over one night or whether they can be spread over multiple nights. Individuals sometimes express concern over the requirement of the ISR procedure to experience very little sleep on the treatment night. For some individuals for whom this necessary sleep restriction is not advised (such as professional drivers, emergency services workers, etc.), a modified treatment protocol whereby patients experience a similar number of sleep onset trials spread over 2-7 nights and experiencing only 1-2 hours of training and only moderately restricted sleep per night may be more agreeable. If ISR's mechanism of action is the retraining of the bedroom environment for sleep, then spreading these retraining

experiences over multiple nights should achieve a similar level of therapeutic benefit. However, the strongest re-association of the bedroom environment for sleepiness may occur through repeated experiences of *rapid* sleep onsets. ISR spread over multiple nights may not result in such rapid sleep onsets. This would be due to 1) reduced homeostatic sleep drive compared to later trials in the overnight procedure and 2) the strongest circadian rhythm sleep pressure in the early morning not being utilised to reduce SOL. Therefore, spreading the ISR treatment over multiple nights may not be as effective compared to one overnight treatment session. Whether this modified ISR protocol would result in similar improvements to sleep and daytime functioning as the original ISR protocol is a theoretically and clinically important topic for further investigation that could be explored using THIM.

Future research into THIM sleep tracking

In addition to assessing the accuracy of THIM for individuals with insomnia, several modifications could be made that would likely improve the accuracy of the THIM sleep tracking function. These modifications include the incorporation of additional physiological sensors into the THIM device hardware and refining the sleep tracking algorithm.

Addition of sensors to the THIM hardware

One additional physiological signal added to the next generation of consumer sleep trackers is sensors for photoplethysmography (PPG). These sensors emit light from the underside of the device onto blood vessels to measure light absorption and provide estimates of heart rate and heart rate variability. Preliminary research suggests that PPG improves the accuracy of wrist actigraphy-based devices (Fonseca et al., 2017; Walch, Huang, Forger, & Goldstein, 2019), and may therefore improve the accuracy of THIM. With the miniaturisation of physiological sensors and

the central processing units (CPUs) to process this additional information in real-time, there are many other potential signals that could be used to improve the device's accuracy. These include those that could be incorporated into the THIM hardware such as skin temperature and pulse oximetry (Hedner et al., 2004). Such additions to THIM would need to undergo empirical validation following recommended guidelines (Depner et al., 2019) to ensure that the device produces sufficiently accurate estimates of sleep and wakefulness for its intended purposes.

Refinement to the THIM sleep tracking algorithm

Another promising approach is to refine the THIM sleep tracking algorithm. In Chapter 5, the tendency for greater overestimation in wakefulness by THIM was theorised to be due to the device placement, the algorithm, or both. With respect to device placement, recent research has suggested that finger twitches occur relatively frequently during N1-sleep and REM, and less frequently in deeper stages of sleep (Reiter et al., 2020). Consequently, THIM may be sensitive enough to detect finger twitches during sleep. If this is the case, then the acceleration from these finger twitches would have increased the average amount of movement in the 30-second epochs. Thus, more epochs may have potentially exceeded the threshold to score an epoch as wake, contributing to THIM's overestimation of wakefulness. It is unlikely that actigraphy devices placed on the wrist would be able to detect these twitches that are largely localised to the finger (Reiter et al., 2020). This may partly explain why THIM appeared to overestimate wakefulness more so than wrist actigraphy devices.

The current THIM algorithm is unable to differentiate between high acceleration values in 30-second epochs from finger twitches and larger body movements. An adjunctive algorithm could be developed that identifies finger

twitches (small movements of a duration less than 1-2 seconds) through the sleep period, separately from larger body movements. This information could be used to more accurately categorise 30-second epochs as wake or sleep. The adjunctive algorithm could also potentially be used to differentiate between sleep stages, as finger twitches tend to occur more frequently in lighter stages of sleep (Reiter et al., 2020). Therefore, this proposed algorithm development could lead to improve accuracy for estimating both wake and sleep stages.

Refinements to the algorithm may incorporate machine learning techniques. Rather than pre-emptively identifying factors that may impact the accuracy of actigraphy devices, machine learning approaches identify patterns and make inferences to inform decisions based on the data instead (with one extracted feature potentially being finger twitches). These approaches produce more accurate estimations of sleep and wakefulness in independent sample actigraphy datasets after training compared to currently validated algorithms (Palotti et al., 2019). Furthermore, personalised machine learning models (one separate model for each individual, trained for each individual) produce more accurate estimates of sleep and wakefulness compared to general machine learning models (one model for all individuals, trained on all individuals), as there is considerable variability between individuals that cannot be accounted for in generalised models (Khademi, El-Manzalawy, Master, Buxton, & Honavar, 2019). If this approach could be implemented into actigraphy scoring software, it may largely overcome the limitation with actigraphy devices of variability in accuracy across individuals.

Responses to vibratory stimuli and sleep tracking

The two approaches to improving the THIM sleep tracking function described above would require extensive resources, time and expertise to develop, research,

and implement. There is another approach that utilises the current THIM hardware and algorithms that would require only slight changes to the THIM smartphone software. The detection of wakefulness using finger tap responses to vibratory stimuli as used in the THIM-ISR function could be incorporated into the THIM sleep tracking function. With this method, individuals would attempt to fall asleep whilst responding to vibratory stimuli. Once THIM detects that sleep onset has occurred, the device could begin monitoring sleep using actigraphy. Once enough movement is detected to suggest that the individual is awake, THIM could re-commence administering vibratory stimuli to which the individual would respond until they fell asleep again. Alternatively, THIM could continue to administer vibratory stimuli after the initial sleep onset has occurred, as the stimuli are presumably at a supra-threshold intensity that is only just perceptible when awake and may be too weak to interrupt sleep once established. With this alternative, the individual would respond to vibratory stimuli as soon as they woke up. The latter method may be more advantageous as using the former method would result in some awakenings being undetected throughout the night, since movement does not always precisely co-occur with awakenings during the sleep period (Pollak et al., 2001). However, the latter method would require greater CPU usage and battery consumption, which may be unfeasible to implement with the current THIM technology.

Preliminary research with a similar device that utilised responsiveness to auditory stimuli emitted after detecting movement with actigraphy (i.e. the former method described above) was promising (Scott, 2016). This device showed substantially higher specificity (M = 0.79, SD = 0.25) and comparable sensitivity (M = 0.89, SD = 0.06), as compared to THIM and other actigraphy devices discussed in this dissertation. Future research could compare the accuracy of the current THIM

sleep tracking function to a modified THIM that utilises finger tap responses to vibratory stimuli in comparison to PSG to determine whether this method is more accurate for estimating sleep and wake.

Conclusion

Based on the findings of this dissertation, the THIM device can accurately estimate sleep onset compared to PSG with good and poor sleepers. THIM and PSG-SOL correspond highly. QEEG analysis further indicated that responsiveness to THIM's vibrations can occur during brief arousal/wakeful periods that are often overlooked by traditional PSG scoring criteria for sleep. The THIM sleep tracking algorithm has comparable accuracy for estimating sleep and wakefulness compared to other actigraphy devices, and its accuracy remains consistent across multiple nights. These promising findings suggest that THIM may be useful for the administration of ISR and for the long-term monitoring of objective sleep in the home environment. The next step is to test THIM-administered ISR and sleep tracking with insomnia patients to ascertain the utility of the device for the treatment and management of this sleep disorder. This practical device may allow for the wider dissemination of the effective ISR behavioural technique to treat insomnia, assisting physicians and ultimately benefiting patients.

References

Abraira, V. E., & Ginty, D. D. (2013). The sensory neurons of touch. *Neuron, 79*(4), 618-639. doi:10.1016/j.neuron.2013.07.051

Acebo, C., Sadeh, A., Seifer, R., Tzischinsky, O., Wolfson, A. R., Hafer, A., & Carskadon, M. A. (1999). Estimating sleep patterns with activity monitoring in children and adolescents: How many nights are necessary for reliable measures? *Sleep, 22*(1), 95-103.

Adams, R. J., Appleton, S. L., Taylor, A. W., Gill, T. K., Lang, C., McEvoy, R. D., & Antic, N. A. (2017). Sleep health of Australian adults in 2016: Results of the 2016 Sleep Health Foundation national survey. *Sleep Health, 3*(1), 35-42. doi:10.1016/j.sleh.2016.11.005

Agnew, H. W., Webb, W. B., & Williams, R. L. (1966). The first night effect: An EEG study of sleep. *Psychophysiology, 2*(3), 263-266.

Ajilore, O., Stickgold, R., Rittenhouse, C. D., & Hobson, J. A. (1995). Nightcap: Laboratory and home‐based evaluation of a portable sleep monitor. *Psychophysiology, 32*(1), 92-98. doi:10.1111/j.1469-8986.1995.tb03410.x

Alsaadi, S. M., McAuley, J. H., Hush, J. M., Bartlett, D. J., McKeough, Z. M., Grunstein, R. R., . . . Maher, C. G. (2014). Assessing sleep disturbance in low back pain: The validity of portable instruments. *PLoS One, 9*(4). doi:10.1371/journal.pone.0095824

American Academy of Sleep Medicine. (2014). *International Classification of Sleep Disorders* (3rd ed.). Darien, IL: American Academy of Sleep Medicine.

American Academy of Sleep Medicine. (2018). *The AASM manual for the scoring of sleep and associated events (version 2.5)*: American Academy of Sleep Medicine.

Ancoli-Israel, S., Martin, J. L., Blackwell, T., Buenaver, L., Liu, L., Meltzer, L. J., . . . Taylor, D. J. (2015). The SBSM guide to actigraphy monitoring: Clinical and research applications. *Behavioral Sleep Medicine, 13 Suppl 1*, S4-S38. doi:10.1080/15402002.2015.1046356

Arroll, B., Fernando, A., 3rd, Falloon, K., Goodyear-Smith, F., Samaranayake, C., & Warman, G. (2012). Prevalence of causes of insomnia in primary care: A cross-sectional study. *British Journal of General Practice, 62*(595), e99-103. doi:10.3399/bjgp12X625157

Baandrup, L., & Jennum, P. J. (2015). A validation of wrist actigraphy against polysomnography in patients with schizophrenia or bipolar disorder. *Neuropsychiatric Disease and Treatment, 11*, 2271-2277. doi:10.2147/NDT.S88236

Baglioni, C., Battagliese, G., Feige, B., Spiegelhalder, K., Nissen, C., Voderholzer, U., . . .

Riemann, D. (2011). Insomnia as a predictor of depression: A meta-analytic evaluation of longitudinal epidemiological studies. *Journal of Affective Disorders, 135*(1-3), 10-19. doi:10.1016/j.jad.2011.01.011

Baker, D. A. (2020). Four ironies of self-quantification: Wearable technologies and the Quantified Self. *Science and Engineering Ethics*. doi:10.1007/s11948-020-00181-w

Baron, K. G., Abbott, S., Jao, N., Manalo, N., & Mullen, R. (2017). Orthosomnia: Are some patients taking the Quantified Self too far? *Journal of Clinical Sleep Medicine, 13*(2), 351-354. doi:10.5664/jcsm.6472

BeHealth Solutions. (2018). SHUTi. Retrieved from http://www.myshuti.com/

Benbadis, S. R., Mascha, E., Perry, M. C., Wolgamuth, B. R., Smolley, L. A., & Dinner, D. S. (1999). Association between the Epworth Sleepiness Scale and the Multiple Sleep Latency Test in a clinical population. *Annals of Internal Medicine, 130*(4), 289-292. doi:10.7326/0003-4819-130-4-199902160-00014

Benca, R. M. (2005). Diagnosis and treatment of chronic insomnia: A review. *Psychiatric services, 56*(3), 332-343.

Bianchi, M. T. (2017). Sleep Devices: Wearables and nearables, informational and interventional, consumer and clinical. *Metabolism, 84*, 99-108. doi:10.1016/j.metabol.2017.10.008

Blackwell, T., Ancoli-Israel, S., Redline, S., & Stone, K. L. (2011). Factors that may influence the classification of sleep-wake by wrist actigraphy: The MrOS sleep study. *Journal of Clinical Sleep Medicine, 7*(4), 357-367. doi:10.5664/jcsm.1190

Bland, J. M., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet, 327*(8476), 307-310. doi:10.1016/S0140-6736(86)90837-8

Blume, C., Del Giudice, R., Lechinger, J., Wislowska, M., Heib, D. P., Hoedlmoser, K., & Schabus, M. (2017). Preferential processing of emotionally and self-relevant stimuli persists in unconscious N2 sleep. *Brain Language, 167*, 72-82. doi:10.1016/j.bandl.2016.02.004

Bonnet, M. H. (1986). Auditory thresholds during continuing sleep. *Biological Psychology, 22*(1), 3-10.

Bonnet, M. H., & Moore, S. E. (1982). The threshold of sleep: Perception of sleep as a function of time asleep and auditory threshold. *Sleep, 5*(3), 267-276.

Bootzin, R., & Epstein, D. R. (2011). Understanding and treating insomnia. *Annual Review of Clinical Psychology, 7*, 435-458. doi:10.1146/annurev.clinpsy.3.022806.091516

Bootzin, R. R. (1972). Stimulus control treatment for insomnia. *Proceedings of the American Psychological Association, 7*, 395-396.

Brooks, A., & Lack, L. (2006). A brief afternoon nap following nocturnal sleep restriction:

Which nap duration is most recuperative? *Sleep, 29*(6), 831-840.

Burnett, K. F., Taylor, C. B., Thoresen, C. E., Rosekind, M. R., Miles, L. E., & DeBusk, R. F. (1985). Toward computerized scoring of sleep using ambulatory recordings of heart rate and physical activity. *Behavioral Assessment, 7*(3), 261-271.

Buysse, D. J., Cheng, Y., Germain, A., Moul, D. E., Franzen, P. L., Fletcher, M., & Monk, T. H. (2010). Night-to-night sleep variability in older adults with and without chronic insomnia. *Sleep Medicine, 11*(1), 56-64. doi:10.1016/j.sleep.2009.02.010

Buysse, D. J., Germain, A., Hall, M. L., Moul, D. E., Nofzinger, E. A., Begley, A., . . . Kupfer, D. J. (2008). EEG spectral analysis in primary insomnia: NREM period effects and sex differences. *Sleep, 31*(12), 1673-1682.

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*, 193-213. doi:10.1016/0165-1781(89)90047-4

Cantero, J. L., Atienza, M., Stickgold, R., & Hobson, J. A. (2002). Nightcap: A reliable system for determining sleep onset. *Sleep, 25*(2), 238-245. doi:10.1093/sleep/25.2.238

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

Carney, C. E., Lajos, L. E., & Waters, W. F. (2004). Wrist actigraph versus self-report in normal sleepers: Sleep schedule adherence and self-report validity. *Behavioral Sleep Medicine, 2*(3), 134-143.

Carskadon, M. A. (1986). Guidelines for the multiple sleep latency test (MSLT): A standard measure of sleepiness. *Sleep, 9*(4), 519-524. doi:10.1093/sleep/9.4.519

Carskadon, M. A., & Dement, W. C. (2017). Normal Human Sleep. In *Principles and practice of sleep medicine* (pp. 15-24.e13).

Casagrande, M., De Gennaro, L., Violani, C., Braibanti, P., & Bertini, M. (1997). A finger-tapping task and a reaction time task as behavioral measures of the transition from wakefulness to sleep: Which task interferes less with the sleep onset process? *Sleep, 20*(4), 301-312.

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

Cellini, N., McDevitt, E. A., Ricker, A. A., Rowe, K. M., & Mednick, S. C. (2015). Validation of an automated wireless system for sleep monitoring during daytime naps. *Behavioral Sleep Medicine, 13*, 157-168. doi:10.1080/15402002.2013.845782

Chae, K. Y., Kripke, D. F., Poceta, J. S., Shadan, F., Jamil, S. M., Cronin, J. W., & Kline, L. E. (2009). Evaluation of immobility time for sleep latency in actigraphy. *Sleep Medicine, 10*(6), 621-625. doi:0.1016/j.sleep.2008.07.009

Chakar, B., Senny, F., Poirrier, A. L., Cambron, L., Fanielle, J., & Poirrier, R. (2017). Validation of midsagittal jaw movements to measure sleep in healthy adults by comparison with actigraphy and polysomnography. *Sleep Science, 10*(3), 122-127. doi:10.5935/1984-0063.20170021

Cheng, P., Luik, A. I., Fellman-Couture, C., Peterson, E., Joseph, C. L. M., Tallent, G., . . . Drake, C. L. (2019). Efficacy of digital CBT for insomnia to reduce depression across demographic groups: A randomized trial. *Psychological medicine, 49*(3), 491-500. doi:10.1017/S0033291718001113

Choi, S. J., Kang, M., Sung, M. J., & Joo, E. Y. (2017). Discordant sleep parameters among actigraphy, polysomnography, and perceived sleep in patients with sleep-disordered breathing in comparison with patients with chronic insomnia disorder. *Sleep Breath, 21*(4), 837-843. doi:10.1007/s11325-017-1514-5

Chung, K. F., Lee, C. T., Yeung, W. F., Chan, M. S., Chung, E. W., & Lin, W. L. (2018). Sleep hygiene education as a treatment of insomnia: A systematic review and meta-analysis. *Family Practice, 35*(4), 365-375. doi:10.1093/fampra/cmx122

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*: Academic Press.

Cohen, Y. E., Bennur, S., Christison-Lagay, K., Gifford, A. M., & Tsunada, J. (2016). Functional organization of the ventral auditory pathway. *Advances in Experimental Medicine and Biology, 894*, 381-388. doi:10.1007/978-3-319-25474-6_40

Cole, R. J., Kripke, D. F., Gruen, W., Mullaney, D. J., & Gillin, J. C. (1992). Automatic sleep/wake identification from wrist activity. *Sleep, 15*(5), 461-469.

Colrain, I. M., Di Parsia, P., & Gora, J. (2000). The impact of prestimulus EEG frequency on auditory evoked potentials during sleep onset. *Canadian Journal of Experimental Psychology, 54*(4), 243.

Connelly, L. (2004). *Behavioural measurements of sleep onset: A comparison of two devices.* Honours thesis submitted to Flinders University, Adelaide, Australia. Retrieved from http://dspace.flinders.edu.au/xmlui/handle/2328/37268

Cook, J. D., Prairie, M. L., & Plante, D. T. (2017). Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. *Journal of Affective Disorders, 217*, 299-305. doi:10.1016/j.jad.2017.04.030

Cook, J. D., Prairie, M. L., & Plante, D. T. (2018). Ability of the multisensory Jawbone UP3 to quantify and classify sleep in patients with suspected central disorders of Hypersomnolence: A comparison against polysomnography and actigraphy. *Journal*

of Clinical Sleep Medicine, 14(5), 841-848. doi:10.5664/jcsm.7120

Cote, K. A., De Lugt, D. R., & Campbell, K. B. (2002). Changes in the scalp topography of event-related potentials and behavioral responses during the sleep onset period. *Psychophysiology, 39*(1), 29-37.

Cote, K. A., Etienne, L., & Campbell, K. B. (2001). Neurophysiological evidence for the detection of external stimuli during sleep. *Sleep, 24*(7), 1-13. doi:10.1093/sleep/24.7.1

Czisch, M., Wetter, T. C., Kaufmann, C., Pollmacher, T., Holsboer, F., & Auer, D. P. (2002). Altered processing of acoustic stimuli during sleep: Reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *Neuroimage, 16*(1), 251-258. doi:10.1006/nimg.2002.1071

D'Rozario, A. L., Dungan, G. C., Banks, S., Liu, P. Y., Wong, K. K., Killick, R., . . . Kim, J. W. (2015). An automated algorithm to identify and reject artefacts for quantitative EEG analysis during sleep in patients with sleep-disordered breathing. *Sleep and Breathing, 19*(2), 607-615.

Daley, M., Morin, C. M., LeBlanc, M., Grégoire, J.-P., & Savard, J. (2009a). The economic burden of insomnia: Direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers. *Sleep, 32*(1), 55-64.

Daley, M., Morin, C. M., LeBlanc, M., Gregoire, J. P., Savard, J., & Baillargeon, L. (2009b). Insomnia and its relationship to health-care utilization, work absenteeism, productivity and accidents. *Sleep Medicine, 10*(4), 427-438. doi:10.1016/j.sleep.2008.04.005

Danzig, R., Wang, M., Shah, A., & Trotti, L. M. (2019). The wrist is not the brain: Estimation of sleep by clinical and consumer wearable actigraphy devices is impacted by multiple patient- and device-specific factors. *Journal of Sleep Research*, e12926. doi:10.1111/jsr.12926

De Gennaro, L., Ferrara, M., & Bertini, M. (2001). The boundary between wakefulness and sleep: Quantitative electroencephalographic changes during the sleep onset period. *Neuroscience, 107*(1), 1-11.

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

de Zambotti, M., Claudatos, S., Inkelis, S., Colrain, I. M., & Baker, F. C. (2015a). Evaluation of a consumer fitness-tracking device to assess sleep in adults. *Chronobiology International, 32*(7), 1024-1028.

de Zambotti, M., Goldstone, A., Claudatos, S., Colrain, I. M., & Baker, F. C. (2018). A validation study of Fitbit Charge 2 compared with polysomnography in adults. *Chronobiol Int, 35*(4), 465-476. doi:10.1080/07420528.2017.1413578

de Zambotti, M., Baker, F. C., & Colrain, I. M. (2015b). Validation of sleep-tracking

technology compared with polysomnography in adolescents. *Sleep, 38*(9), 1461-1468. doi:10.5665/sleep.4990

de Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M., & Baker, F. C. (2019). Wearable sleep technology in clinical and research settings. *Medicine & Science in Sports & Exercise, 51*(7), 1538-1557. doi:10.1249/MSS.0000000000001947

de Zambotti, M., Rosas, L., Colrain, I. M., & Baker, F. C. (2017). The sleep of the ring: Comparison of the OURA sleep tracker against polysomnography. *Behavioral Sleep Medicine, 17*(2), 1-15. doi:10.1080/15402002.2017.1300587

Deloitte Access Economics. (2017). *Asleep on the job: Costs of inadequate sleep in Australia*: Sleep Health Foundation.

Dement, W., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *EEG Clinical Neurophysiology, 9*(4), 673-690. doi:10.1016/0013-4694(57)90088-3

Depner, C. M., Cheng, P. C., Devine, J. K., Khosla, S., de Zambotti, M., Robillard, R., . . . Drummond, S. P. A. (2019). Wearable technologies for developing sleep and circadian biomarkers: A summary of workshop discussions. *Sleep*. doi:10.1093/sleep/zsz254

Dick, R., Penzel, T., Fietze, I., Partinen, M., Hein, H., & Schulz, J. (2010a). AASM standards of practice compliant validation of actigraphic sleep analysis from SOMNOwatch versus polysomnographic sleep diagnostics shows high conformity also among subjects with sleep disordered breathing. *Physiological Measurement, 31*(12), 1623-1633. doi:10.1088/0967-3334/31/12/005

Dick, R., Penzel, T., Fietze, I., Partinen, M., Hein, H., & Schulz, J. (2010b). AASM standards of practice compliant validation of actigraphic sleep analysis from SOMNOwatch versus polysomnographic sleep diagnostics shows high conformity also among subjects with sleep disordered breathing. *Physiological Measurement, 31*(12), 1623-1633. doi:10.1088/0967-3334/31/12/005

Downey, R., & Bonnet, M. H. (1992). Training subjective insomniacs to accurately perceive sleep onset. *Sleep, 15*(1), 58-63.

Dubiel, J. (2019). *Initial Evaluation of a Portable Device for Phase Advancing the Circadian Rhythms in Those with Sleep-Onset Insomnia.* Honours thesis submitted to Flinders University, Adelaide, Australia.

Dunican, I. C., Murray, K., Slater, J. A., Maddison, K. J., Jones, M. J., Dawson, B., . . . Eastwood, P. R. (2018). Laboratory and home comparison of wrist-activity monitors and polysomnography in middle-aged adults. *Sleep Biological Rhythms, 16*(1), 85-97. doi:10.1007/s41105-017-0130-x

Edinger, J. D., & Carney, C. E. (2014). *Overcoming insomnia: A cognitive-behavioral therapy*

approach, therapist guide: Oxford University Press.

Edinger, J. D., Fins, A. I., Sullivan Jr, R. J., Marsh, G. R., Dailey, D. S., Hope, T. V., . . . Vasilas, D. (1997). Sleep in the laboratory and sleep at home: Comparisons of older insomniacs and normal sleepers. *Sleep, 20*(12), 1119-1126.

Edinger, J. D., Glenn, D. M., Bastian, L. A., Marsh, G. R., Daile, D., Hope, T. V., . . . Meeks, G. (2001). Sleep in the laboratory and sleep at home II: Comparisons of middle-aged insomnia sufferers and normal sleepers. *Sleep, 24*(7), 761-770.

Edinger, J. D., Means, M. K., Stechuchak, K. M., & Olsen, M. K. (2004). A pilot study of inexpensive sleep-assessment devices. *Behavioral Sleep Medicine, 2*(1), 41-49. doi:10.1207/s15402010bsm0201_4

Espie, C. A., Emsley, R., Kyle, S. D., Gordon, C., Drake, C. L., Siriwardena, A. N., . . . Luik, A. I. (2019). Effect of digital cognitive behavioral therapy for Insomnia on health, psychological well-being, and sleep-related quality of life: A randomized clinical trial. *JAMA Psychiatry, 76*(1), 21-30. doi:10.1001/jamapsychiatry.2018.2745

Espie, C. A., Kyle, S. D., Williams, C., Ong, J. C., Douglas, N. J., Hames, P., & Brown, J. S. (2012). A randomized, placebo-controlled trial of online cognitive behavioral therapy for chronic insomnia disorder delivered via an automated media-rich web application. *Sleep, 35*(6), 769-781. doi:10.5665/sleep.1872

Espie, C. A., Lindsay, W. R., Brooks, D. N., Hood, E. M., & Turvey, T. (1989). A controlled comparative investigation of psychological treatments for chronic sleep-onset insomnia. *Behaviour Research and Therapy, 27*(1), 79-88.

Evenson, K. R., Goto, M. M., & Furberg, R. D. (2015). Systematic review of the validity and reliability of consumer-wearable activity trackers. *International Journal of Behavioral Nutrition and Physical Activity, 12*, 159. doi:10.1186/s12966-015-0314-1

Farabi, S. S., Quinn, L., & Carley, D. W. (2017). Validity of actigraphy in measurement of sleep in young adults with Type 1 Diabetes. *Journal of Clinical Sleep Medicine, 13*(5), 669-674. doi:10.5664/jcsm.6580

Fernandez Guerrero, A., & Achermann, P. (2019). Brain dynamics during the sleep onset transition: An EEG source localization study. *Neurobiology of Sleep and Circadian Rhythms, 6*, 24-34. doi:10.1016/j.nbscr.2018.11.001

Ferrara, M., & de Gennaro, L. (2011). Going local: Insights from EEG and stereo-EEG studies of the human sleep-wake cycle. *Current Topics in Medicinal Chemistry, 11*, 2423-2437.

Fietze, I., Penzel, T., Partinen, M., Sauter, J., Kuchler, G., Suvoro, A., & Hein, H. (2015). Actigraphy combined with EEG compared to polysomnography in sleep apnea patients. *Physiological Measurement, 36*(3), 385-396. doi:10.1088/0967-3334/36/3/385

Finan, P. H., Richards, J. M., Gamaldo, C. E., Han, D. F., Leoutsakos, J. M., Salas, R., . . . Smith, M. T. (2016). Validation of a wireless, self-application, ambulatory electroencephalographic sleep monitoring device in healthy volunteers. *Journal of Clinical Sleep Medicine, 12*(11), 1443-1451. doi:10.5664/jcsm.6262

Fogel, R. B., Trinder, J., White, D. P., Malhotra, A., Raneri, J., Schory, K., . . . Pierce, R. J. (2005). The effect of sleep onset on upper airway muscle activity in patients with sleep apnoea versus controls. *The Journal of physiology, 564*(2), 549-562.

Fonseca, P., Weysen, T., Goelema, M. S., Most, E. I. S., Radha, M., Lunsingh Scheurleer, C., . . . Aarts, R. M. (2017). Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep, 40*(7). doi:10.1093/sleep/zsx097

Freedman, R. R., & Sattler, H. L. (1982). Physiological and psychological factors in sleep-onset insomnia. *Journal of Abnormal Psychology, 91*(5), 380.

Freeman, D., Sheaves, B., Goodwin, G. M., Yu, L.-M., Nickless, A., Harrison, P. J., . . . Espie, C. A. (2017). The effects of improving sleep on mental health (OASIS): A randomised controlled trial with mediation analysis. *The Lancet Psychiatry, 4*(10), 749-758. doi:10.1016/s2215-0366(17)30328-0

Fuller, K. L., Juliff, L., Gore, C. J., Peiffer, J. J., & Halson, S. L. (2017). Software thresholds alter the bias of actigraphy for monitoring sleep in team-sport athletes. *J Sci Med Sport, 20*(8), 756-760. doi:10.1016/j.jsams.2016.11.021

Gavriloff, D., Sheaves, B., Juss, A., Espie, C. A., Miller, C. B., & Kyle, S. D. (2018). Sham sleep feedback delivered via actigraphy biases daytime symptom reports in people with insomnia: Implications for insomnia disorder and wearable devices. *Journal of Sleep Research, 27*(6), e12726. doi:10.1111/jsr.12726

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica, 25*(2), 141-151. doi:10.11613/BM.2015.015

Goldstone, A., Baker, F. C., & de Zambotti, M. (2018). Actigraphy in the digital health revolution: still asleep? *Sleep, 41*(9). doi:10.1093/sleep/zsy120

Gorgoni, M., D'Atri, A., Scarpelli, S., Ferrara, M., & De Gennaro, L. (2019). Timing and Topography of Sleep Onset: Asynchronies and Regional Changes of Brain Activity. In *Handbook of Sleep Research* (pp. 19-31).

Gorney, S. (1997). *A parametric and sleep hysteresis approach to assessing sleep and wake from a wrist activity meter with enhanced frequency range.* Paper presented at the Paper Presented at the 11th Annual Meeting of the Associated Professional Sleep Societies, June 10-15, 1997, San Francisco, CA.

Gradisar, M., & Lack, L. (2004). Relationships between the circadian rhythms of finger temperature, core temperature, sleep latency, and subjective sleepiness. *Journal of*

biological rhythms, 19(2), 157-163. doi:10.1177/0748730403261560

Griessenberger, H., Heib, D. P. J., Kunz, A. B., Hoedlmoser, K., & Schabus, M. (2013). Assessment of a wireless headband for automatic sleep scoring. *Sleep Breath, 17*(2), 747-752. doi:10.1007/s11325-012-0757-4

Gruwez, A., Libert, W., Ameye, L., & Bruyneel, M. (2017). Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. *International Journal of Medical Informatics, 102*, 87-92. doi:10.1016/j.ijmedinf.2017.03.008

Guilleminault, C., Phillips, R., & Dement, W. C. (1975). A syndrome of hypersomnia with automatic behavior. *EEG Clinical Neurophysiology, 38*(4), 403-413.

Hamill, K., Jumabhoy, R., Kahawage, P., de Zambotti, M., Walters, E. M., & Drummond, S. P. A. (2020). Validity, potential clinical utility and comparison of a consumer activity tracker and a research-grade activity tracker in insomnia disorder II: Outside the laboratory. *Journal of Sleep Research, 29*(1), e12944. doi:10.1111/jsr.12944

Harris, J., Lack, L., Kemp, K., Wright, H., & Bootzin, R. (2012). A randomized controlled trial of intensive sleep retraining (ISR): A brief conditioning treatment for chronic insomnia. *Sleep, 35*(1), 49-60. doi:10.5665/sleep.1584

Harris, J., Lack, L., Wright, H., Gradisar, M., & Brooks, A. (2007). Intensive Sleep Retraining treatment for chronic primary insomnia: A preliminary investigation. *Journal of Sleep Research, 16*(3), 276-284. doi:10.1111/j.1365-2869.2007.00595.x

Harsh, J., Voss, U., Hull, J., Schrepfer, S., & Badia, P. (1994). ERP and behavioral changes during the wake/sleep transition. *Psychophysiology, 31*(3), 244-252.

Hauri, P. J. (1999). Evaluation of a sleep switch device. *Sleep, 22*(8), 1110-1117.

Hayashi, M., Motoyoshi, N., & Hori, T. (2005). Recuperative power of a short daytime nap with or without stage 2 sleep. *Sleep, 28*(7), 829-836.

Hedner, J., Pillar, G., Pittman, S. D., Zou, D., Grote, L., & White, D. P. (2004). A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep, 27*(8), 1560-1566.

Hedner, J., White, D. P., Malhotra, A., Herscovici, S., Pittman, S. D., Zou, D., . . . Pillar, G. (2011). Sleep staging based on autonomic signals: A multi-center validation study. *Journal of Clinical Sleep Medicine, 7*(3), 301-306. doi:10.5664/JCSM.1078

Hertig-Godeschalk, A., Skorucak, J., Malafeev, A., Achermann, P., Mathis, J., & Schreier, D. R. (2019). Microsleep episodes in the borderland between wakefulness and sleep. *Sleep*. doi:10.1093/sleep/zsz163

Hilditch, C. J., Centofanti, S. A., Dorrian, J., & Banks, S. (2016). A 30-minute, but not a 10-minute nighttime nap is associated with sleep inertia. *Sleep, 39*(3), 675-685. doi:10.5665/sleep.5550

Hilditch, C. J., Dorrian, J., & Banks, S. (2017). A review of short naps and sleep inertia: Do naps of 30 min or less really avoid sleep inertia and slow-wave sleep? *Sleep Medicine, 32*, 176-190. doi:10.1016/j.sleep.2016.12.016

Hillman, D. R., & Lack, L. C. (2013). Public health implications of sleep loss: The community burden. *Medical Journal of Australia, 199*(8), S7-S10. doi:10.5694/mja13.10620

Hirshkowitz, M. (2017). Polysomnography and beyond. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (6th ed., pp. 1564-1566): Elsevier.

Insana, S. P., Glowacki, S. S., & Montgomery-Downs, H. E. (2011). Assessing the efficacy to conduct the Multiple Sleep Latency Test with actigraphy. *Behavioral Sleep Medicine, 9*(4), 257-265. doi:10.1080/15402002.2011.607018

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.

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep, 14*(6), 540-545.

Kaas, J. H. (2012). Somatosensory System. In *The Human Nervous System* (pp. 1074-1109).

Kahawage, P., Jumabhoy, R., Hamill, K., de Zambotti, M., & Drummond, S. P. A. (2019). Validity, potential clinical utility, and comparison of consumer and research-grade activity trackers in Insomnia Disorder I: In-lab validation against polysomnography. *Journal of Sleep Research*, e12931. doi:10.1111/jsr.12931

Kanady, J. C., Drummond, S. P., & Mednick, S. C. (2011). Actigraphic assessment of a polysomnographic-recorded nap: A validation study. *Journal of Sleep Research, 20*(1 Pt 2), 214-222. doi:10.1111/j.1365-2869.2010.00858.x

Kang, S. G., Kang, J. M., Ko, K. P., Park, S. C., Mariani, S., & Weng, J. (2017). Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. *Journal of Psychosomatic Research, 97*, 38-44. doi:10.1016/j.jpsychores.2017.03.009

Kapella, M. C., Vispute, S., Zhu, B. Q., & Herdegen, J. J. (2017). Actigraphy scoring for sleep outcome measures in chronic obstructive pulmonary disease. *Sleep Medicine, 37*, 124-129. doi:10.1016/j.sleep.2017.06.012

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.

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. doi:10.2147/NSS.S71159

Khademi, A., El-Manzalawy, Y., Master, L., Buxton, O. M., & Honavar, V. G. (2019). Personalized sleep parameters estimation from actigraphy: A machine learning approach. *Nature and Science of Sleep, 11*, 387-399. doi:10.2147/NSS.S220716

Khosla, S., Deak, M. C., Gault, D., Goldstein, C. A., Hwang, D., Kwon, Y., . . . Kirsch, D. B. (2019). Consumer Sleep Technologies: How to balance the promises of new technology with evidence-based medicine and clinical guidelines. *Journal of Clinical Sleep Medicine, 15*(1), 163-165. doi:10.5664/jcsm.7598

Khosla, S., Deak, M. C., Gault, D., Goldstein, C. A., Hwang, D., Kwon, Y., . . . American Academy of Sleep Medicine Board of, D. (2018). Consumer sleep technology: An American Academy of Sleep Medicine position statement. *Journal of Clinical Sleep Medicine, 14*(5), 877-880. doi:10.5664/jcsm.7128

Kim, M. J., Lee, G. H., Kim, C. S., Kim, W. S., Chung, Y. S., Chung, S., & Lee, S. A. (2013). Comparison of three actigraphic algorithms used to evaluate sleep in patients with obstructive sleep apnea. *Sleep Breath, 17*(1), 297-304. doi:10.1007/s11325-012-0689-z

Kleitman, N. (1929). Sleep. *Physiological Reviews, 9*(4), 624-665.

Kosmadopoulos, A., Sargent, C., Darwent, D., Zhou, X., & Roach, G. D. (2014). Alternatives to polysomnography (PSG): A validation of wrist actigraphy and a partial-PSG system. *Behavior Research Methods, 46*(4), 1032-1041.

Kripke, D. F. (2016). Mortality risk of hypnotics: Strengths and limits of evidence. *Drug Safety, 39*(2), 93-107. doi:10.1007/s40264-015-0362-0

Kuderian, R. H., Ogilvie, R. D., McDonnell, G., & Simons, I. (1991). Behavioural response home monitoring of good and insomniac sleepers. *Canadian Journal of Psychology, 45*(2), 169-178.

Kuo, C. E., Liu, Y. C., Chang, D. W., Young, C. P., Shaw, F. Z., & Liang, S. F. (2017). Development and Evaluation of a Wearable Device for Sleep Quality Assessment. *IEEE Trans Biomed Eng, 64*(7), 1547-1557. doi:10.1109/tbme.2016.2612938

Kuula, L., Gradisar, M., Martinmäki, K., Richardson, C., Bonnar, D., Bartel, K., . . . Pesonen, A. K. (2019). Using big data to explore worldwide trends in objective sleep in the transition to adulthood. *Sleep Medicine, 62*, 69-76. doi:10.1016/j.sleep.2019.07.024

Kyle, S. D., Morgan, K., & Espie, C. A. (2010). Insomnia and health-related quality of life. *Sleep Medicine Reviews, 14*(1), 69-82. doi:10.1016/j.smrv.2009.07.004

Kyle, S. D., Morgan, K., Spiegelhalder, K., & Espie, C. A. (2011). No pain, no gain: An exploratory within-subjects mixed-methods evaluation of the patient experience of sleep restriction therapy (SRT) for insomnia. *Sleep Medicine, 12*(8), 735-747. doi:10.1016/j.sleep.2011.03.016

Laakso, M. L., Leinonen, L., Lindblom, N., Joutsiniemi, S. L., & Kaski, M. (2004). Wrist actigraphy in estimation of sleep and wake in intellectually disabled subjects with motor handicaps. *Sleep Medicine, 5*(6), 541-550. doi:10.1016/j.sleep.2004.05.002

Lack, L., & Baraniec, M. (2002). Intensive sleep onset training for sleep onset insomnia. *Sleep, 25*(Abstract Supplement), A478-A479.

Lack, L., & Mair, A. (1995). The relationship between EEG and a behavioural measure of sleep onset. http://hdl.handle.net/2328/36963

Lack, L., Scott, H., & Lovato, N. (2019). Intensive Sleep Retraining treatment of Insomnia. *Sleep Medicine Clinics, 14*(2), 245-252. doi:10.1016/j.jsmc.2019.01.005

Lack, L., Scott, H., Micic, G., & Lovato, N. (2017). Intensive Sleep Re-Training: From bench to bedside. *Brain Sciences, 7*(4). doi:10.3390/brainsci7040033

Lack, L., Wright, H., & Paynter, D. (2007). The treatment of sleep onset insomnia with bright morning light. *Sleep and Biological Rhythms, 5*(3), 173-179.

Lack, L. C., & Wright, H. R. (2007). Treating chronobiological components of chronic insomnia. *Sleep Medicine, 8*(6), 637-644.

Lauerma, H., Kaartinen, J., Polo, O., Sallinen, M., & Lyytinen, H. (1994). Asymmetry of instructed motor response to auditory stimuli during sleep. *Sleep, 17*(5), 444-448.

Lichstein, K. L., Stone, K. C., Donaldson, J., Nau, S. D., Soeffing, J. P., Murray, D., . . . Aguillard, N. (2006). Actigraphy validation with insomnia. *Sleep, 29*(2), 232-239.

Loomis, A. L., Harvey, E. N., & Hobart, G. (1935). Potential rhythms of the cerebral cortex during sleep. *Science, 81*(2111), 597-598. doi:10.1126/science.81.2111.597

Lovato, N., & Lack, L. (2010a). The effects of napping on cognitive functioning. *Progress in Brain Research, 185*, 155-166. doi:10.1016/B978-0-444-53702-7.00009-9

Lovato, N., & Lack, L. (2010b). The effects of napping on cognitive functioning. *Progress in Brain Research, 185*, 155-166. doi:10.1016/B978-0-444-53702-7.00009-9

Lovato, N., & Lack, L. (2018). Insomnia and mortality: A meta-analysis. *Sleep Medicine Reviews*. doi:10.1016/j.smrv.2018.10.004

Lovato, N., Lack, L., Wright, H., & Kennaway, D. J. (2014). Evaluation of a brief treatment program of cognitive behavior therapy for insomnia in older adults. *Sleep, 37*(1), 117-126. doi:10.5665/sleep.3320

Lucey, B. P., McLeland, J. S., Toedebusch, C. D., Boyd, J., Morris, J. C., Landsness, E. C., . . . Holtzma, D. M. (2016). Comparison of a single-channel EEG sleep study to polysomnography. *Journal of Sleep Research, 25*(6), 625-635. doi:10.1111/jsr.12417

MacLean, A. W., Arnedt, T., Biedermann, H., & Knowles, J. B. (1992). Behavioural responding as a measure of sleep quality. *Sleep Research, 21*, 105.

Maglione, J. E., Liu, L. Q., Neikrug, A. B., Poon, T., Natarajan, L., Calderon, J., . . . Ancoli-Israel, S. (2013). Actigraphy for the Assessment of Sleep Measures in Parkinson's

Disease. *Sleep, 36*(8), 1209-1217. doi:10.5665/sleep.2888

Mair, A. (1994). *The relation between EEG and behavioural sleep onset.* Honours thesis submitted to Flinders University, Adelaide, Australia.

Mair, A., Scott, H., & Lack, L. (2020). Intensive Sleep Retraining Treatment for Insomnia Administered by Smartphone in the Home. *Manuscript in preparation.*

Marino, M., Li, Y., Rueschman, M. N., Winkelman, J. W., Ellenbogen, J. M., Solet, J. M., . . . Buxton, O. M. (2013). Measuring sleep: Accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. *Sleep, 36*(11), 1747-1755. doi:10.5665/sleep.3142

Markwald, R. R., Bessman, S. C., Reini, S. A., & Drummond, S. P. (2016a). Performance of a portable sleep monitoring device in individuals with high versus low sleep efficiency. *Journal of Clinical Sleep Medicine, 12*(1), 95-103. doi:10.5664/jcsm.5404

Markwald, R. R., Bessman, S. C., Reini, S. A., & Drummond, S. P. A. (2016b). Performance of a portable sleep monitoring device in individuals with high versus low sleep efficiency. *Journal of Clinical Sleep Medicine, 12*(1), 95-103. doi:10.5664/jcsm.5404

Matsuo, M., Masuda, F., Sumi, Y., Takahashi, M., Yamada, N., Ohira, M. H., . . . Kadotani, H. (2016). Comparisons of portable sleep monitors of different modalities: Potential as naturalistic sleep recorders. *Frontiers in Neurology, 7*. doi:10.3389/fneur.2016.00110

Matthews, E. E., Arnedt, J. T., McCarthy, M. S., Cuddihy, L. J., & Aloia, M. S. (2013). Adherence to cognitive behavioral therapy for insomnia: A systematic review. *Sleep Medicine Reviews, 17*(6), 453-464. doi:10.1016/j.smrv.2013.01.001

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

McCurry, S. M., Logsdon, R. G., Teri, L., & Vitiello, M. V. (2007). Evidence-based psychological treatments for insomnia in older adults. *Psychology and Aging, 22*(1), 18-27. doi:10.1037/0882-7974.22.1.18

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist, 50*(9), 741.

Mezzanotte, W. S., Tangel, D. J., & White, D. P. (1996). Influence of sleep onset on upper-airway muscle activity in apnea patients versus normal controls. *American Journal of Respiratory and Critical Care Medicine, 153*(6), 1880-1887.

MicroSleep. (2015). Sleep On Cue-Research Application (Version 1.1).

Miller, C. B., Valenti, L., Harrison, C. M., Bartlett, D. J., Glozier, N., Cross, N. E., . . . Marshall, N. S. (2017). Time trends in the family physician management of Insomnia:

The Australian experience (2000-2015). *Journal of Clinical Sleep Medicine, 13*(6), 785-790. doi:10.5664/jcsm.6616

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The, P. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine, 6*(7), e1000097. doi:10.1371/journal.pmed.1000097

Morin, C. M., Belleville, G., Bélanger, L., & Ivers, H. (2011). The Insomnia Severity Index: Psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep, 34*(5), 601-608. doi:10.1093/sleep/34.5.601

Morin, C. M., & Espie, C. A. (2007). *Insomnia: A clinical guide to assessment and treatment*: Springer Science & Business Media.

Morin, C. M., Hauri, P. J., Espie, C. A., Spielman, A. J., Buysse, D. J., & Bootzin, R. R. (1999). Nonpharmacologic treatment of chronic insomnia. *Sleep, 22*(8), 1134-1156.

Morin, C. M., LeBlanc, M., Daley, M., Gregoire, J. P., & Merette, C. (2006). Epidemiology of insomnia: Prevalence, self-help treatments, consultations, and determinants of help-seeking behaviors. *Sleep Medicine, 7*(2), 123-130. doi:10.1016/j.sleep.2005.08.008

Morphy, H., Dunn, K. M., Lewis, M., Boardman, H. F., & Croft, P. R. (2007). Epidemiology of insomnia: A longitudinal study in a UK population. *Sleep, 30*(3), 274-280.

Morris, M., Lack, L., & Dawson, D. (1990). Sleep-onset insomniacs have delayed temperature rhythms. *Sleep, 13*(1), 1-14.

Mundt, J. M., Crew, E. C., Krietsch, K., Roth, A. J., Vatthauer, K., Robinson, M. E., . . . McCrae, C. S. (2016). Measuring treatment outcomes in comorbid insomnia and fibromyalgia: concordance of subjective and objective assessments. *Journal of Clinical Sleep Medicine, 12*(2), 215-223. doi:10.5664/jcsm.5488

Myllymaa, S., Muraja-Murro, A., Westeren-Punnonen, S., Hukkanen, T., Lappalainen, R., Mervaala, E., . . . Myllymaa, K. (2016). Assessment of the suitability of using a forehead EEG electrode set and chin EMG electrodes for sleep staging in polysomnography. *Journal of Sleep Research, 25*(6), 636-645. doi:10.1111/jsr.12425

Nakazaki, K., Kitamura, S., Motomura, Y., Hida, A., Kamei, Y., Miura, N., & Mishima, K. (2014). Validity of an algorithm for determining sleep/wake states using a new actigraph. *Journal of Physiological Anthropology, 33*(1), 31.

NHS North West London. (2018). Sleepio a free app for people in London who have trouble sleeping. Retrieved from https://www.healthiernorthwestlondon.nhs.uk/news/2018/10/19/sleepio-free-app-people-london-who-have-trouble-sleeping

O'Hare, E., Flanagan, D., Penzel, T., Garcia, C., Frohberg, D., & Heneghan, C. (2015). A comparison of radio-frequency biomotion sensors and actigraphy versus polysomnography for the assessment of sleep in normal subjects. *Sleep and*

Breathing, 19(1), 91-98. doi:10.1007/s11325-014-0967-z

Oakley, N. (1997). Validation with polysomnography of the Sleepwatch sleep/wake scoring algorithm used by the Actiwatch activity monitoring system. *Bend: Mini Mitter, Cambridge Neurotechnology*.

Ogilvie, R. D. (2001). The process of falling asleep. *Sleep Medicine Reviews, 5*(3), 247-270. doi:10.1053/smrv.2001.0145

Ogilvie, R. D., Simons, I. A., Kuderian, R. H., MacDonald, T., & Rustenburg, J. (1991). Behavioral, event-related potential, and EEG/FFT changes at sleep onset. *Psychophysiology, 28*(1), 54-64.

Ogilvie, R. D., & Wilkinson, R. T. (1988). Behavioral versus EEG-based monitoring of all-night sleep/wake patterns. *Sleep, 11*(2), 139-155.

Ogilvie, R. D., Wilkinson, R. T., & Allison, S. (1989). The detection of sleep onset: Behavioral, physiological, and subjective convergence. *Sleep, 12*(5), 458-474. doi:10.1093/sleep/12.5.458

Ohayon, M. M., & Caulet, M. (1995). Insomnia and psychotropic drug consumption. *Progress in neuro-psychopharmacology & biological psychiatry, 19*(3), 421-431.

Ohayon, M. M., & Sagales, T. (2010). Prevalence of insomnia and sleep characteristics in the general population of Spain. *Sleep Medicine, 11*(10), 1010-1018. doi:10.1016/j.sleep.2010.02.018

Ong, J. L., Tandi, J., Patanaik, A., Lo, J. C., & Chee, M. W. L. (2019). Large-scale data from wearables reveal regional disparities in sleep patterns that persist across age and sex. *Scientific Reports, 9*(1), 3415. doi:10.1038/s41598-019-40156-x

Oniz, A., Inanc, G., Guducu, C., & Ozgoren, M. (2016). Brain responsiveness to non-painful tactile stimuli prior and during sleep. *Sleep Biological Rhythms, 14*, 87-96. doi:10.1007/s41105-015-0026-6

Palotti, J., Mall, R., Aupetit, M., Rueschman, M., Singh, M., Sathyanarayana, A., . . . Fernandez-Luque, L. (2019). Benchmark on a large cohort for sleep-wake classification with machine learning techniques. *NPJ Digital Medicine, 2*, 50. doi:10.1038/s41746-019-0126-9

Paquet, J., Kawinska, A., & Carrier, J. (2007). Wake detection capacity of actigraphy during sleep. *Sleep, 30*(10), 1362-1369.

Perlis, M., Jungquist, C., Smith, M. T., & Posner, D. (2005). *The cognitive behavioral treatment of insomnia: A treatment manual*. In. Retrieved from https://www.amazon.com/Cognitive-Behavioral-Treatment-Insomnia-Session-ebook/dp/B000PC6BGA

Perlis, M., Shaw, P., Cano, G., & Espie, C. (2011). Models of Insomnia. In M. H. Kryger, T. Roth, & W. C. Dement (Eds.), *Principles and practice of sleep medicine* (5th edn ed.,

pp. 850-865). St Louis: Elsevier Saunders.

Perlis, M. L., Giles, D. E., Mendelson, W. B., Bootzin, R. R., & Wyatt, J. K. (1997). Psychophysiological insomnia: the behavioural model and a neurocognitive perspective. *Journal of Sleep Research, 6*(3), 179-188. doi:10.1046/j.1365-2869.1997.00045.x

Perlis, M. L., Merica, H., Smith, M. T., & Giles, D. E. (2001). Beta EEG activity and insomnia. *Sleep Medicine Reviews, 5*(5), 365-376.

Perlis, M. L., Smith, M. T., & Pigeon, W. R. (2005). Etiology and pathophysiology of insomnia. *Principles and practice of sleep medicine, 4*, 714-725.

Perlis, M. L., Vargas, I., Ellis, J. G., Grandner, M. A., Morales, K. H., Gencarelli, A., . . . Thase, M. E. (2019). The natural history of Insomnia: The incidence of acute Insomnia and subsequent progression to Chronic Insomnia or recovery in good sleeper subjects. *Sleep*. doi:10.1093/sleep/zsz299

Perrin, F., García-Larrea, L., Mauguière, F., & Bastuji, H. (1999). A differential brain response to the subject's own name persists during sleep. *Clinical Neurophysiology, 110*(12), 2153-2164.

Pigeon, W. R., Taylor, M., Bui, A., Oleynk, C., Walsh, P., & Bishop, T. M. (2018). Validation of the sleep-wake scoring of a new wrist-worn sleep monitoring device. *Journal of Clinical Sleep Medicine, 14*(6), 1057-1062.

Pollak, C. P., Tryon, W. W., Nagaraja, H., & Dzwonczyk, R. (2001). How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep, 24*(8), 957-965.

Prerau, M. J., Brown, R. E., Bianchi, M. T., Ellenbogen, J. M., & Purdon, P. L. (2017). Sleep neurophysiological dynamics through the lens of multitaper spectral analysis. *Physiology & Behavior, 32*(1), 60-92. doi:10.1152/physiol.00062.2015

Prerau, M. J., Hartnack, K. E., Obregon-Henao, G., Sampson, A., Merlino, M., Gannon, K., . . . Purdon, P. L. (2014). Tracking the sleep onset process: An empirical model of behavioral and physiological dynamics. *PLOS Computational Biology, 10*(10), e1003866. doi:10.1371/journal.pcbi.1003866

Quante, M., Kaplan, E. R., Cailler, M., Rueschman, M., Wang, R., Weng, J., . . . Redline, S. (2018). Actigraphy-based sleep estimation in adolescents and adults: A comparison with polysomnography using two scoring algorithms. *Nature and Science of Sleep, 10*, 13-20. doi:10.2147/nss.S151085

Razjouyan, J., Lee, H., Parthasarathy, S., Mohler, J., Sharafkhaneh, A., & Najafi, B. (2017). Improving sleep quality assessment using wearable sensors by including information from postural/sleep position changes and body acceleration: A comparison of chest-worn sensors, wrist actigraphy, and polysomnography. *Journal of Clinical Sleep Medicine, 13*(11), 1301-1310. doi:10.5664/jcsm.6802

Re-Time, P. L. (2016). Thim - the first wearable device for sleep, which will improve your sleep. Retrieved from http://thim.io/

Rechtschaffen, A., Hauri, P., & Zeitlin, M. (1966). Auditory awakening thresholds in REM and NREM sleep stages. *Perceptual and Motor Skills, 22*, 927-942.

Rechtschaffen, A., & Kales, A. (1968). A Manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.

Reid, K., & Dawson, D. (1999). Correlation between wrist activity monitor and electrophysiological measures of sleep in a simulated shiftwork environment for younger and older subjects. *Sleep, 22*(3), 378-385.

Reiter, A. M., Roach, G. D., Sargent, C., & Lack, L. (2020). Finger twitches are more frequent in REM sleep than in Non-REM sleep. *Nature and Science of Sleep, 12*, 49-56.

Respironics, P. (2009). *Equivalence of activity recordings and derived sleep statistics: Actiwatch-64, Actiwatch 2 and Actiwatch Spectrum*. Retrieved from http://www.actigraphy.com/assets/AWDataEquivalenceReport_0309_final-e69f789615144b7e1b8a3b3959e94556fee79f5ec67e2f02edea641c427524c3.pdf

Reynolds, A. C., Appleton, S. L., Gill, T. K., Taylor, A. W., McEvoy, R. D., Ferguson, S. A., & Adams, R. J. (2017). Sickness absenteeism is associated with sleep problems independent of sleep disorders: Results of the 2016 Sleep Health Foundation national survey. *Sleep Health, 3*(5), 357-361. doi:10.1016/j.sleh.2017.06.003

Riedel, B. W., Lichstein, K. L., Peterson, B. A., Epperson, M. T., Means, M. K., & Aguillard, R. (1998). A comparison of the efficacy of stimulus control for medicated and nonmedicated insomniacs. *Behavior Modification, 22*(1), 3-28.

Riemann, D., & Perlis, M. L. (2009). The treatments of chronic insomnia: A review of benzodiazepine receptor agonists and psychological and behavioral therapies. *Sleep Medicine Reviews, 13*(3), 205-214. doi:10.1016/j.smrv.2008.06.001

Ritterband, L. M., Thorndike, F. P., Ingersoll, K. S., Lord, H. R., Gonder-Frederick, L., Frederick, C., . . . Morin, C. M. (2017). Effect of a web-based Cognitive Behavior Therapy for Insomnia intervention with 1-Year follow-up: A randomized clinical trial. *JAMA Psychiatry, 74*(1), 68-75. doi:10.1001/jamapsychiatry.2016.3249

Robbins, R., Krebs, P., Rapoport, D. M., Jean-Louis, G., & Duncan, D. T. (2018). Examining use of mobile phones for sleep tracking among a national sample in the USA. *Health Communication*, 1-7. doi:10.1080/10410236.2017.1422104

Rosenberg, R. S., & Van Hout, S. (2013). The American Academy of Sleep Medicine inter-scorer reliability program: Sleep stage scoring. *Journal of Clinical Sleep Medicine, 9*(1), 81-87. doi:10.5664/jcsm.2350

Roth, T. (2007). Insomnia: Definition, prevalence, etiology, and consequences. *Journal of*

Clinical Sleep Medicine, 3(5 Suppl), S7-S10.

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

Russo, K., & Bianchi, M. T. (2017). Consumer sleep device monitoring as therapeutic feedback for adults with insomnia.

Rybarczyk, B., Stepanski, E., Fogg, L., Lopez, M., Barry, P., & Davis, A. (2005). A placebo-controlled test of cognitive-behavioral therapy for comorbid insomnia in older adults. *Journal of Consulting and Clinical Psychology, 73*(6), 1164-1174. doi:10.1037/0022-006X.73.6.1164

Sadeh, A., & Acebo, C. (2002). The role of actigraphy in sleep medicine. *Sleep Medicine Reviews, 6*(2), 113-124. doi:10.1053/smrv.2001.0182

Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: An empirical test of methodological issues. *Sleep, 17*(3), 201-207.

Sanchez-Ortuno, 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), 21-29.

Sangal, R. B., Mitler, M. M., & Sangal, J. M. (1999). Subjective sleepiness ratings (Epworth sleepiness scale) do not reflect the same parameter of sleepiness as objective sleepiness (maintenance of wakefulness test) in patients with narcolepsy. *Clinical Neurophysiology, 110*(12), 2131-2135.

Sargent, C., Lastella, M., Halson, S. L., & Roach, G. D. (2016). The validity of activity monitors for measuring sleep in elite athletes. *J Sci Med Sport, 19*(10), 848-853. doi:10.1016/j.jsams.2015.12.007

Sawh, M. (2018). Fitbit's fitness tracker sales take a dip, as smartwatches step up. Retrieved from https://www.wareable.com/fitbit/fitbit-fitness-tracker-sales-decline-forecast-6048

Scatena, M., Dittoni, S., Maviglia, R., Frusciante, R., Testani, E., Vollono, C., . . . Della Marca, G. (2012). An integrated video-analysis software system designed for movement detection and sleep analysis. Validation of a tool for the behavioural study of sleep. *Clinical Neurophysiology, 123*(2), 318-323. doi:10.1016/j.clinph.2011.07.026

Scott, H. (2016). *Validation of a novel smartphone application for the detection of sleep and wakefulness.* Honours thesis submitted to Flinders University, Adelaide, Australia.

Scott, H., Lack, L., & Lovato, N. (2018). A pilot study of a novel smartphone application for the estimation of sleep onset. *Journal of Sleep Research, 27*(1), 90-97. doi:10.1111/jsr.12575

Scott, H., Lack, L., & Lovato, N. (2019). A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Medicine Reviews, 49*, 101227.

doi:10.1016/j.smrv.2019.101227

Senny, F., Maury, G., Cambron, L., Leroux, A., Destine, J., & Poirrier, R. (2012). The sleep/wake state scoring from mandible movement signal. *Sleep Breath, 16*(2), 535-542. doi:10.1007/s11325-011-0539-4

Shambroom, J. R., Fabregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research, 21*(2), 221-230. doi:10.1111/j.1365-2869.2011.00944.x

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.

Silvertsen, B., Omvik, S., Havik, O. E., Pallesen, S., Bjorvatn, B., Nielsen, G. H., . . . Nordhus, I. H. (2006). A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. *Sleep, 29*(10), 1353-1358.

Sivertsen, B., Omvik, S., Havik, O. E., Pallesen, S., Bjorvatn, B., Nielsen, G. H., . . . Nordhus, I. H. (2006). A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. *Sleep, 29*(10), 1353-1358.

Slater, J. A., Botsis, T., Walsh, J., King, S., Straker, L. M., & Eastwood, P. R. (2015). Assessing sleep using hip and wrist actigraphy. *Sleep and Biological Rhythms, 13*(2), 172-180.

Sleepio. (2019). Sleepio. Retrieved from https://www.sleepio.com/

Smith, M. T., McCrae, C. S., Cheung, J., Martin, J. L., Harrod, C. G., Heald, J. L., & Carden, K. A. (2018). Use of actigraphy for the evaluation of sleep disorders and circadian rhythm sleep-wake disorders: An American Academy of Sleep Medicine Systematic Review, meta-analysis, and GRADE assessment. *Journal of Clinical Sleep Medicine, 14*(7), 1209-1230. doi:10.5664/jcsm.7228

Smith, M. T., Perlis, M. L., Park, A., Smith, M. S., Pennington, J., Giles, D. E., & Buysse, D. J. (2002). Comparative meta-analysis of pharmacotherapy and behavior therapy for persistent insomnia. *American Journal of Psychiatry, 159*(1), 5-11.

Spiegelhalder, K., Regen, W., Feige, B., Holz, J., Piosczyk, H., Baglioni, C., . . . Nissen, C. (2012). Increased EEG sigma and beta power during NREM sleep in primary insomnia. *Biological Psychology, 91*(3), 329-333.

St-Jean, G., Turcotte, I., Pérusse, A. D., & Bastien, C. H. (2013). REM and NREM power spectral analysis on two consecutive nights in psychophysiological and paradoxical insomnia sufferers. *International Journal of Psychophysiology, 89*(2), 181-194.

Statista. (2020). Number of Fitbit devices sold worldwide from 2010 to 2019. Retrieved from https://www.statista.com/statistics/472591/fitbit-devices-sold/

Stone, K. L., Ensrud, K. E., & Ancoli-Israel, S. (2008). Sleep, insomnia and falls in elderly

patients. *Sleep Medicine, 9*, S18-S22. doi:10.1016/s1389-9457(08)70012-1

Streiner, D. L., & Norman, G. R. (2006). "Precision" and "accuracy": two terms that are neither. *Journal of Clinical Epidemiology, 59*(4), 327-330. doi:10.1016/j.jclinepi.2005.09.005

Svensson, T., Chung, U. I., Tokuno, S., Nakamura, M., & Svensson, A. K. (2019). A validation study of a consumer wearable sleep tracker compared to a portable EEG system in naturalistic conditions. *J Psychosom Res, 126*, 109822. doi:10.1016/j.jpsychores.2019.109822

Swan, M. (2012). Sensor mania! The Internet of Things, wearable computing, objective metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks, 1*(3), 217-253. doi:10.3390/jsan1030217

Sweetman, A., Lack, L., Lambert, S., Gradisar, M., & Harris, J. (2017). Does comorbid obstructive sleep apnea impair the effectiveness of cognitive and behavioral therapy for insomnia? *Sleep Medicine, 39*, 38-46. doi:10.1016/j.sleep.2017.09.003

Taibi, D. M., Landis, C. A., & Vitiello, M. V. (2013). Concordance of polysomnographic and actigraphic measurement of sleep and wake in older women with insomnia. *Journal of Clinical Sleep Medicine, 9*(3), 217-225. doi:10.5664/jcsm.2482

Takahashi, M., Fukuda, H., & Arito, H. (1998). Brief naps during post-lunch rest: effects on alertness, performance, and autonomic balance. *European Journal of Applied Physiology and Occupational Physiology, 78*(2), 93-98. doi:10.1007/s004210050392

Tang, N. K., & Harvey, A. G. (2006). Altering misperception of sleep in insomnia: Behavioral experiment versus verbal feedback. *Journal of Consulting and Clinical Psychology, 74*(4), 767.

Taylor, D. J., & Pruiksma, K. E. (2014). Cognitive and behavioural therapy for insomnia (CBT-I) in psychiatric populations: a systematic review. *International Review of Psychiatry, 26*(2), 205-213. doi:10.3109/09540261.2014.902808

Thomas, A., Grandner, M., Nowakowski, S., Nesom, G., Corbitt, C., & Perlis, M. L. (2016). Where are the behavioral sleep medicine providers and where are they needed? A geographic assessment. *Behavioral Sleep Medicine, 14*(6), 687-698.

Thorndike, F. P., Saylor, D. K., Bailey, E. T., Gonder-Frederick, L., Morin, C. M., & Ritterband, L. M. (2008). Development and perceived utility and impact of an internet intervention for insomnia. *E-journal of applied psychology: clinical and social issues, 4*(2), 32.

Tietzel, A. J., & Lack, L. C. (2002). The recuperative value of brief and ultra-brief naps on alertness and cognitive performance. *Journal of Sleep Research, 11*(3), 213-218.

Tonetti, L., Cellini, N., de Zambotti, M., Fabbri, M., Martoni, M., Fabregas, S. E., . . . Natale, V. (2013). Polysomnographic validation of a wireless dry headband technology for

sleep monitoring in healthy young adults. *Physiology & Behavior, 118*, 185-188. doi:10.1016/j.physbeh.2013.05.036

Tonetti, L., Pasquini, F., Fabbri, M., Belluzzi, M., & Natale, V. (2008). Comparison of two different actigraphs with polysomnography in healthy young subjects. *Chronobiology International, 25*(1), 145-153. doi:10.1080/07420520801897228

Tryon, W. W. (2004). Issues of validity in actigraphic sleep assessment. *Sleep, 27*(1), 158-165.

Vallieres, A., & Morin, C. M. (2003). Actigraphy in the assessment of insomnia. *Sleep, 26*(7), 902-906.

Vallières, A., & Morin, C. M. (2003). Actigraphy in the assessment of insomnia. *Sleep, 26*(7), 902-906.

Van den Water, A. T. M., Holmes, A., & Hurley, D. A. (2011). Objective measurements of sleep for non-laboratory settings as alternative to polysomnography - A systematic review. *Journal of Sleep Research, 20*, 183-200. doi:0.1111/j.1365-2869.2009.00814.x

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

van Straten, A., van der Zweerde, T., Kleiboer, A., Cuijpers, P., Morin, C. M., & Lancee, J. (2018). Cognitive and behavioral therapies in the treatment of insomnia: A meta-analysis. *Sleep Medicine Reviews, 38*, 3-16. doi:10.1016/j.smrv.2017.02.001

Wade, A. G. (2010). The societal costs of insomnia. *Neuropsychiatric Disease and Treatment, 7*(1), 1-18. doi:10.2147/NDT.S15123

Walch, O., Huang, Y., Forger, D., & Goldstein, C. (2019). Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep*. doi:10.1093/sleep/zsz180

Wang, D., Wong, K. K., Dungan, G. C., Buchanan, P. R., Yee, B. J., & Grunstein, R. R. (2008). The validity of wrist actimetry assessment of sleep with and without sleep apnea. *Journal of Clinical Sleep Medicine, 4*(5), 450-455.

Watson, N. F., Lawlor, C., & Raymann, R. (2019). Will Consumer Sleep Technologies Change the Way We Practice Sleep Medicine? *J Clin Sleep Med, 15*(1), 159-161. doi:10.5664/jcsm.7596

White, D. P., Gibb, T. J., Wall, J. M., & Westbrook, P. R. (1995). Assessment of accuracy and analysis time of a novel device to monitor sleep and breathing in the home. *Sleep, 18*(2), 115-126. doi:10.1093/sleep/18.2.115

Wilt, T. J., MacDonald, R., Brasure, M., Olson, C. M., Carlyle, M., Fuchs, E., . . . Kane, R. L. (2016). Pharmacologic treatment of Insomnia Disorder: An evidence report for a clinical practice guideline by the American College of Physicians. *Annals of Internal*

Medicine, 165(2), 103-112. doi:10.7326/M15-1781

Withrow, D., Roth, T., Koshorek, G., & Roehrs, T. (2019). Relation between ambulatory actigraphy and laboratory polysomnography in insomnia practice and research. *J Sleep Res, 28*(4), e12854. doi:10.1111/jsr.12854

Yang, C. M., Han, H. Y., Yang, M. H., Su, W. C., & Lane, T. (2010). What subjective experiences determine the perception of falling asleep during sleep onset period? *Consciousness and Cognition, 19*(4), 1084-1092. doi:10.1016/j.concog.2009.12.017

Zhang, Z., Guan, C. T., Chan, T. E., Yu, J. H., Ng, A. K., Zhang, H. H., . . . Ieee. (2014). Automatic sleep onset detection using single EEG sensor. In *2014 36th Annual International Conference of the Ieee Engineering in Medicine and Biology Society* (pp. 2265-2268).

Zinkhan, M., Berger, K., Hense, S., Nagel, M., Obst, A., Koch, B., . . . Stang, A. (2014). Agreement of different methods for assessing sleep characteristics: A comparison of two actigraphs, wrist and hip placement, and self-report with polysomnography. *Sleep Medicine, 15*(9), 1107-1114. doi:10.1016/j.sleep.2014.04.015

Appendices

Appendix 1.

Database Search Strategies

PubMed Search:

(("Sleep Latency/physiology"[Mesh] OR "Sleep Stages/physiology"[Mesh]) OR ("sleep onset"[tw] OR "sleep onsets"[tw] OR "sleep latency"[tw] OR "sleep latencies"[tw] OR "sleep onset latency"[tw] OR "sleep onset latencies"[tw] OR "sleep onset period"[tw] OR "sleep onset periods"[tw] OR "sleep onset process"[tw] OR "sleep wake transition"[tw] OR "sleep wake transitions"[tw])) AND (("Wearable Electronic Devices"[Mesh] OR "Mobile Applications"[Mesh] OR "Actigraphy"[Mesh] OR "Monitoring, Ambulatory"[Mesh]) OR ("device"[tw] OR "devices"[tw] OR "wearable"[tw] OR "wearables"[tw] OR "actigraph"[tw] OR "actigraphs"[tw] OR "actigraphy"[tw] OR "accelerometer"[tw] OR "accelerometers"[tw] OR "mobile application"[tw] OR "mobile applications"[tw] OR "smartphone application"[tw] OR "smartphone applications"[tw] OR "mobile app"[tw] OR "mobile apps"[tw] OR "smartphone app"[tw] OR "smartphone apps"[tw] OR "ambulatory"[tw] OR "portable"[tw])) AND ("Data Accuracy"[Mesh] OR ("measure"[tw] OR "measures"[tw] OR "measurement"[tw] OR "measuring"[tw] OR "monitor"[tw] OR "monitors"[tw] OR "monitoring"[tw] OR "estimate"[tw] OR "estimates"[tw] OR "estimation"[tw] OR "estimating"[tw] OR "detect"[tw] OR "detects"[tw] OR "detection"[tw] OR "detecting"[tw] OR "accuracy"[tw] OR "accurate"[tw] OR "accurately"[tw] OR "valid"[tw] OR "validation"[tw] OR "validity"[tw] OR "reliable"[tw] OR "reliability"[tw] OR "performance"[tw] OR "perform"[tw] OR "performs"[tw] OR "assess"[tw] OR "assesses"[tw] OR "assessment"[tw]))

Returned – 1157 results; Restricted to English articles – 1115 results

Web of Science Search:

1. TS=("sleep onset" OR "sleep onsets" OR "sleep latency" OR "sleep latencies" OR "sleep onset latency" OR "sleep onset latencies" OR "sleep onset period" OR "sleep onset periods" OR "sleep onset process" OR "sleep wake transition" OR "sleep wake transitions")

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years

2. TS=("device" OR "devices" OR "wearable" OR "wearables" OR "actigraph" OR "actigraphs" OR "actigraphy" OR "accelerometer" OR "accelerometers" OR "mobile application" OR "mobile applications" OR "smartphone application" OR "smartphone applications" OR "mobile app" OR "mobile apps" OR "smartphone app" OR "smartphone apps" OR "ambulatory" OR "portable")

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years

3. TS=("measure" OR "measures" OR "measurement" OR "measuring" OR "monitor" OR "monitors" OR "monitoring" OR "estimate" OR "estimates" OR "estimation" OR "estimating" OR "detect" OR "detects" OR "detection" OR "detecting" OR "accuracy" OR "accurate" OR "accurately" OR "valid" OR "validation" OR "validity" OR "reliable" OR "reliability" OR "performance" OR "perform" OR "performs" OR "assess" OR "assesses" OR "assessment")

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI, CCR-EXPANDED, IC Timespan=All years

(#3 AND #2 AND #1) AND LANGUAGE: (English)

Returned – 866 results; Restricted to English articles – 857 results

SCOPUS Search:

1. INDEXTERMS ("Sleep Latency" OR "Sleep Stages")

2. TITLE-ABS-KEY (sleep onset OR sleep onsets OR sleep latency OR sleep latencies OR sleep onset latency OR sleep onset latencies OR sleep onset period OR sleep onset periods OR sleep onset process OR sleep wake transition OR sleep wake transitions)

3. INDEXTERMS ("Wearable Electronic Devices" OR "Mobile Applications" OR "Actigraphy" OR "Monitoring, Ambulatory")

4. TITLE-ABS-KEY (device OR devices OR wearable OR wearables OR actigraph OR actigraphs OR actigraphy OR accelerometer OR accelerometers OR mobile application OR mobile applications OR smartphone application OR smartphone applications OR mobile app OR mobile apps OR smartphone app OR smartphone apps OR ambulatory OR portable)

5. INDEXTERMS ("Data Accuracy")

6. TITLE-ABS-KEY (measure OR measures OR measurement OR measuring OR monitor OR monitors OR monitoring OR estimate OR estimates OR estimation OR estimating OR detect OR detects OR detection OR detecting OR accuracy OR accurate OR accurately OR valid OR validation OR validity OR reliable OR reliability OR performance OR perform OR performs OR assess OR assesses OR assessment)

(#1 OR #2) AND (#3 OR #4) AND (#5 OR #6)

Returned: 1,549 results; Restricted to English articles: 1,477 results

PsycINFO Search:

Searches	Results
1	sleep onset/
2	mobile devices/
3	actigraphy/
4	monitoring/
5	2 or 3 or 4
6	("sleep onset" or "sleep onsets" or "sleep latency" or "sleep latencies" or "sleep onset latency" or "sleep onset latencies" or "sleep onset period" or "sleep onset periods" or "sleep onset process" or "sleep wake transition" or "sleep wake transitions").tw.
7	("device" or "devices" or "wearable" or "wearables" or "actigraph" or "actigraphs" or "actigraphy" or "accelerometer" or "accelerometers" or "mobile application" or "mobile applications" or "smartphone application" or "smartphone applications" or "mobile app" or "mobile apps" or "smartphone app" or "smartphone apps" or "ambulatory" or "portable").tw.
8	("measure" or "measures" or "measurement" or "measuring" or "monitor" or "monitors" or "monitoring" or "estimate" or "estimates" or "estimation" or "estimating" or "detect" or "detects" or "detection" or "detecting" or "accuracy" or "accurate" or "accurately" or "valid" or "validation" or "validity" or "reliable" or "reliability" or "performance" or "perform" or "performs" or "assess" or "assesses" or "assessment").tw.
9	1 or 6
10	5 or 7
11	8 and 9 and 10
12	limit 11 to English language

Returned – 526 results; Restricted to English articles – 511 results

CINAHL (EBSCOhost) Search:

Search Terms	Search Options
S9	S5 AND S6 AND S7
S8	S5 AND S6 AND S7
S7	S2 OR S3
S6	S1 OR S4
S5	TX ("measure" OR "measures" OR "measurement" OR "measuring" OR "monitor" OR "monitors" OR "monitoring" OR "estimate" OR "estimates" OR "estimation" OR "estimating" OR "detect" OR "detects" OR "detection" OR "detecting" OR "accuracy" OR "accurate" OR "accurately" OR "valid" OR "validation" OR "validity" OR "reliable" OR "reliability" OR "performance" OR "perform" OR "performs" OR "assess" OR "assesses" OR "assessment")
S4	TX ("sleep onset" OR "sleep onsets" OR "sleep latency" OR "sleep latencies" OR "sleep onset latency" OR "sleep onset latencies" OR "sleep onset period" OR "sleep onset periods" OR "sleep onset process" OR "sleep wake transition" OR "sleep wake transitions")
S3	TX ("device" OR "devices" OR "wearable" OR "wearables" OR "actigraph" OR "actigraphs" OR "actigraphy" OR "accelerometer" OR "accelerometers" OR "mobile application" OR "mobile applications" OR "smartphone application" OR "smartphone applications" OR "mobile app" OR "mobile apps" OR "smartphone app" OR "smartphone apps" OR "ambulatory" OR "portable")
S2	(MH "Monitoring, Physiologic") OR (MH "Actigraphy") OR (MH "Mobile Applications") OR (MH "Wearable Sensors")
S1	(MH "Sleep Stages")

Returned – 240 results; Restricted to English articles – 239 results

Appendix 2.

Table A2. *Study Characteristics including sample information, setting, and PSG specifications.*

Study	Sample Characteristics			Testing Characteristics		Gold-standard measure	
	N	Age (years)	Sample Description (N in brackets)	Sleep Type	Setting	PSG Specifications	PSG Scoring Criteria
Ajilore et al. (1995)	10	M = 23.7 (SD = 7.0), range: 19-42	Average sleepers	nocturnal sleep	Laboratory	Grass Model 8-10 polygraph	R&K
Alsaadi et al. (2014)	50	M = 42.7 (SD = 15.15)	All experienced non-specific lower back pain	nocturnal sleep	Laboratory	Sandman system	AASM
Baandrup and Jennum (2015)	42	M = 46.1 (SD = 9.5)	Schizophrenia (37) or bipolar (5)	nocturnal sleep	Homes	Trackit ambulatory PSG	AASM
Blackwell et al. (2011)	889	M = 76.28 (SD = 5.47)	Elderly men with osteoporotic fractures	nocturnal sleep	Homes	Compumedics Safiro Unit	R&K
Burnett et al. (1985)	10	M = 53 (SD = 17.8), range: 25-83	No sleep complaints (2), disturbed sleep ≥3 nights per week (8)	nocturnal sleep	Homes	Grass Model 7D polygraph	R&K, N2-sleep

							onset as 10 mins of N2
Cantero, Atienza, Stickgold, and Hobson (2002)	10	range: 20-25	healthy with no sleep disturbances	daytime nap	Laboratory	Grass Model 8-10 polygraph	R&K
Cellini et al. (2013)	30	M = 20.77 (SD = 3.14)	healthy with no sleep disorders	daytime nap	Laboratory	Astro-Med Grass Heritage model 15 amplifiers	AASM
Cellini et al. (2015)	30	M = 20.3 (SD = 2.76)	healthy with no sleep disorders	daytime nap	Laboratory	Astro-Med Grass Heritage model 15	AASM
Chae et al. (2009)	33	M = 54 (SD = 8.7)	OSA (20), OSA and PLMS (13)	nocturnal sleep	Laboratory	Compumedics E Series	AASM
Chakar et al. (2017)	38	M = 23.5 (SD = 1.5)	healthy with no sleep disturbances	nocturnal sleep	Laboratory	Embla N7000 system	AASM

Choi et al. (2017)	66	SDB group: M = 49.76 (SD = 14.6) CI group: M = 58.11 (SD = 9.73)	sleep disordered breathing (SDB, 36), chronic insomnia (CI, 30)	nocturnal sleep	Laboratory	Remlogic Embla Systems	AASM
Cole et al. (1992)	21	M = 47.8 (SD = 15.7)	Controls (7), elderly (2), psychiatric disorder (6), OSA (2), DIMS (2), widows (2)	nocturnal sleep	Laboratory	Two-channel unspecified PSG system	R&K First epoch of 20mins of sleep
Cook et al. (2017)	21	M = 26.5 (SD = 4.6)	Mild-moderate unipolar MDD	nocturnal sleep	Laboratory	Alice Sleepware	AASM
Cook et al. (2018)	43	M = 33.3 (SD = 11)	narcolepsy (3), idiopathic hypersomnia (13), organic/unspecified hypersomnia (18), mild OSA (6), hypersomnolence from other condition (4)	nocturnal sleep	Laboratory	Alice Sleepware	AASM
De Souza et al. (2003)	21	range: 18-33	healthy with no sleep disturbances	nocturnal sleep	Laboratory	sleep analyzer computer, v9.2	R&K

de Zambotti et al. (2015a)	28	M = 50.1 (SD = 3.9)	Healthy (12), chronic insomnia (12), PLMI > 10 (2), PLMI > 10 and AHI > 5 (2)	nocturnal sleep	Laboratory	Compumedics amplifiers	AASM
de Zambotti et al. (2018)	44	range: 19-61	no health conditions (35), PLMS (9)	nocturnal sleep	Laboratory	Compumedics Grael-PSG system	AASM
Dick et al. (2010a)	28	M = 56 (SD = 10)	All tentative OSA diagnosis, PLMS (12), arrhythmia (3)	overnight, diagnostic or treatment	Laboratory	SOMNOscreen system (SSC)	R&K
Dunican et al. (2018)	50	M = 57 (SD = 5), range: 46-73	Community sample of middle-aged adults	nocturnal sleep	Laboratory	Compumedics Grael-PSG system	AASM
Edinger et al. (2004)	33	M = 58.6 (SD = 13.5)	All psychiatrically stable veterans, OSA (10), poor sleep hygiene (12), PLMS (3), psychophysiological insomnia (3), insomnia and anxiety (3), insomnia and mood disorder	nocturnal sleep	medical centre	Compumedics PS2 or Safiro systems	R&K

			(1), hypnotic-dependent sleep disorder (1)				
Farabi, Quinn, and Carley (2017)	27	M = 23.8 (SD = 4.1), range: 18–30	All diagnosed with T1 diabetes, no sleep disorders	overnight, TIB >7hrs	Laboratory	Alice 5 system	AASM
Fietze et al. (2015)	A: 30 B: 20	A: M = 57 (SD = 14), range: 18-80 B: M = 60 (SD = 11), range: 41-74	A: all OSA, PLMS (6), arrhythmia (1) B: all OSA, PLMS (12), Cheyne-Stokes respiration (1)	overnight, diagnostic or treatment night	Laboratory	SOMNOscreen system (SSC)	R&K
Finan et al. (2016)	14	M = 26.43 (SD = 3.74) range: 22–34	healthy, good sleepers	overnight, TIB 8hrs	Laboratory	Embla N7000 system	AASM
Fonseca et al. (2017)	1: 16 2: 35	1: M = 51.2 (SD = 8.4) 2: M = 52.0 (SD = 6.9)	All healthy	nocturnal sleep	hotel	Alice PDx system	AASM
Fuller et al. (2017)	21	M = 22.5 (SD = 2.7)	Elite athletes	nocturnal sleep	apartment	Compumedics Siesta 802 system	AASM

Griessenberger et al. (2013)	10	M = 32.5 (SD = 7.63), range: 23-45	Insomnia (7), no sleep disorders (3)	nocturnal sleep	Not provided	Synamps EEG amplifiers	semiautomatic scoring (AASM)
Gruwez et al. (2017)	20	M = 30 (SD = 5)	All healthy	nocturnal sleep	Homes	Dream ambulatory PSG	AASM
Hauri (1999)	25	Insomnia: M = 44.5, range: 22-65 Controls: M = 25.5, range: 19-40	Insomnia (19), controls (6)	nocturnal sleep	Laboratory	Not provided	First epoch of S1 First epoch of S2 First epoch of 10min of sleep
Hedner et al. (2004)	228	M = 48.8 (SD = 14.0)	Controls (38), mild OSA (54), moderate OSA (83), severe OSA (53)	nocturnal sleep	Laboratory or homes	Embla system Alice III system	R&K
Hedner et al. (2011)	227	M = 49 (SD = 14)	Controls (38), mild OSA (54), moderate OSA (82), severe OSA (53)	nocturnal sleep	Laboratory or homes	Embla system Alice III system	R&K

Insana et al. (2011)	41	M = 27.65 (SD = 4.72), range: 18.44-38.44	first-time parents, participated during M = 6.93 (SD = 1:26) postpartum week	MSLTs	Laboratory	Embla N7000 system	R&K
Kanady et al. (2011)	19	M = 19.7 (SD = 1.5)	Healthy with no sleep disturbances	daytime nap	Laboratory	Astro-Med Grass Heritage Model 15 amplifiers	R&K
Kang et al. (2017)	50	Insomnia: M = 38.4 (SD = 11.2) Good sleeper: M = 32.1 (SD = 7.4)	Insomnia, good sleepers	nocturnal sleep	Homes	Embletta X100 ambulatory PSG	AASM
Kapella, Vispute, Zhu, and Herdegen (2017)	50	M = 63.2 (SD = 8.4)	Mild to severe COPD, no other sleep disorders or conditions	nocturnal sleep	Laboratory	Alice 3 system	AASM

Kaplan, Talbot, Gruber, and Harvey (2012)	54	Bipolar group: M = 33.1 (SD = 10.3) Matched controls: M = 38.1 (SD = 13.0)	Bipolar type I or type II but in-between mood episodes (27) Matched controls with no history of mental illness or sleep disturbance (27)	nocturnal sleep	Laboratory	Compumedics Siesta802 system	R&K
Kaplan et al. (2014)	99	Mdn = 32.7, range: 18-60	Average sleepers (49), and those with chronic insomnia symptoms, suspected restless leg syndrome, PLMS, or suspected sleep apnea (50)	nocturnal sleep	Laboratory	PSG system (Consolidated Research)	R&K
Kosmadopoulos et al. (2014)	22	M = 23.9 (SD = 3.8)	healthy with no sleep disturbances	nocturnal sleep	Laboratory	Compumedics Siesta system	R&K
Kuo et al. (2017)	59	information not available	good sleepers (43), poor sleepers (16)	nocturnal sleep	Laboratory	Compumedics Siesta802 system	R&K
Laakso et al. (2004)	39	able-bodied:	Able-bodied with normal sleep (10), sleep-disordered without motor	nocturnal sleep	Laboratory Intellectuall	Embla system	R&K

		M = 28 (SD = 10) sleep-disordered: M = 38 (SD = 14) sleep-disordered with motor disabilities: M = 36 (SD = 13)	disabilities (13), sleep-disordered with motor disabilities (16)		y Disabled Centre or homes		
Lichstein et al. (2006)	57	range: 21-87	All had insomnia	nocturnal sleep	Laboratory	Alice 3 system	R&K
Lucey et al. (2016)	29	M = 54 (SD = 15.7), range: 25-80	Either no OSA or mild OSA (18), PLMS (5)	overnight, diagnostic or treatment	Laboratory	Polysmith system	AASM
Maglione et al. (2013)	61	M = 67.74 (SD = 9.26)	Mild-moderate Parkinson's Disease	nocturnal sleep	Laboratory	Compumedics somtè system	AASM
Markwald et al. (2016a)	29	M = 24.0 (SD = 5.3)	Healthy with no sleep disorders	nocturnal sleep	Laboratory	Grass Comet Plus system	AASM
Matsuo et al. (2016)	20	M = 20.70 (SD = 0.39), range: 19-24	Healthy with no sleep disorders	Not provided	Not provided	Alice 5 system	Not provided

McCall and McCall (2012)	54	M = 41.3 (SD = 12.9)	Diagnosed with insomnia and major depressive episodes	nocturnal sleep	Laboratory	VIASYS SomnoStar system	AASM First epoch of 10min of sleep
Mundt et al. (2016)	113	M = 52.68 (SD = 10.91)	Diagnosed with insomnia and fibromyalgia	nocturnal sleep	Homes	AURA Portable System	R&K
Myllymaa et al. (2016)	31	M = 31.3 (SD = 11.8)	Sleep bruxism and healthy controls	nocturnal sleep	Laboratory	Embla N7000 system	AASM
Nakazaki et al. (2014)	17	M = 21.9 (SD = 1.7)	Healthy with no sleep disorders	nocturnal sleep	Laboratory	Neurofax EEG-1200 system	First epoch of sleep (120s epochs)
O'Hare et al. (2015)	20	M = 30 (SD = 6)	No sleep disorders	overnight, TIB 8 hrs	Laboratory	Embla N7000 system	First epoch of 2mins of sleep
Paquet et al. (2007)	15	M = 39.3 (SD = 15.1), range: 20-60	healthy with no sleep disturbances	Nocturnal sleep and daytime	Laboratory	Grass Model 15 Neurodata system	First epoch of 10mins of sleep

				recovery sleeps			
Pigeon et al. (2018)	20	M = 30.1 (SD = 13.1)	Healthy, good sleepers	nocturnal sleep	Laboratory	Embla N7000 system	AASM
Razjouyan et al. (2017)	21	M = 50.8 (SD = 12.8)	Self-reported sleep problems	nocturnal sleep	Laboratory	Grass Comet PLUS XL Embla S4500 system	
Reid and Dawson (1999)	32	Group 1: M = 21.2 (SD = 2.7), range: 18-30 Group 2: M = 43.9 (SD = 6.8), range: 35-56	Both groups healthy with no sleep disorders	Overnight simulated shift work protocol	Laboratory	Sleep analyser computer	R&K
Rupp and Balkin (2011)	29	M = 24.3 (SD = 5.4)	Healthy with no sleep disturbances	nocturnal sleep	Laboratory	Not provided	First epoch of sleep
Sanchez-Ortuno et al. (2010)	62	Insomnia: M = 28.3 (SD = 4.9) Controls: M = 28.4 (SD = 6.0)	Insomnia group (31) and matched normal sleepers (31)	nocturnal sleep	Laboratory and homes	Oxford Medilog 9000 or 9200 systems	AASM

Sargent et al. (2016)	16	M = 19.3 (SD = 1.5)	Elite cyclists, no sleep disorders	nocturnal sleep	training camp	Compumedics system	AASM
Scatena et al. (2012)	25	M = 44.3 (SD = 18.4), range: 25-63	All healthy	nocturnal sleep	Laboratory	Not provided	AASM
Scott et al. (2018)	12	M = 21.67 (SD = 1.23)	healthy with no sleep disturbances	overnight, sleep onset trials	Laboratory	Compumedics Somtè system	AASM
Senny et al. (2012)	124	M = 50.8 (SD = 12.4)	OSA (68), insomnia/depression (27), Other including PLMS and circadian rhythm disorders (29)	nocturnal sleep	Laboratory	Embla S7000 or N7000 systems	First epoch of sleep First epoch of 15min of sleep
Shambroom et al. (2012)	26	M = 38 (SD = 13), range: 19-60	Healthy with no sleep disorders	nocturnal sleep	Laboratory	Cadwell Easy III EEG	R&K First epoch of 10min of sleep

Signal, Gale, and Gander (2005)	21	M = 41.8 (SD = 9.1)	Flight crew (11 Captains, 10 First Officers)	in-flight rest and layover sleeps	Airplane and hotel	Embla system	R&K
Sivertsen et al. (2006)	34	M = 60.5 (SD = 4.5)	Insomnia, mostly sleep maintenance insomnia	nocturnal sleep	Laboratory	Embla A10	R&K
Slater et al. (2015)	108	M = 22.7 (SD = 0.2)	All healthy	nocturnal sleep	Laboratory	Compumedics Grael-PSG system	AASM
Taibi et al. (2013)	16	M = 69.4 (SD = 8.1)	All had insomnia	nocturnal sleep	Laboratory	Model not provided	R&K
Tonetti et al. (2013)	11	M = 24.75 (SD = 3.62)	Healthy with no sleep disorders	overnight, TIB 8 hrs	Laboratory	Compumedics Siesta802 system	R&K
Tonetti et al. (2008)	12	M = 22.97 (SD = 2.62)	Healthy with no sleep disorders	nocturnal sleep	Laboratory	Grass Heritage PSG System	R&K
Vallieres and Morin (2003)	17	M = 41.6 (SD = 5.7), range: 34-50	Sleep onset insomnia (1), sleep maintenance insomnia (9), or mixed (7)	nocturnal sleep	Laboratory	Not provided	R&K

Wang et al. (2008)	21	M = 38.9 (SD = 13.0)	Non-OSA (10) and OSA (11)	nocturnal sleep	Laboratory	Alice 5 system	R&K
White et al. (1995)	30	M = 51.1 (SD = 2.9)	Suspected OSA diagnosis	nocturnal sleep	Laboratory	Grass Model 78E polygraph	R&K
Zhang et al. (2014)	20	M = 29.7 (SD = 7.5)	All healthy	daytime nap	Laboratory	Not provided	AASM
Zinkhan et al. (2014)	100	M = 51.3 (SD = 13.0), range: 19-73	Community sample	nocturnal sleep	Laboratory	Not provided	AASM

AASM = American Academy of Sleep Medicine scoring criteria, AHI = Apnea–Hypopnea Index, DIMS = disorders of initiating and maintaining sleep, M = mean, Mdn = median, N2 = Non-rapid eye movement Stage 2, OSA = obstructive sleep apnea, PLMI = periodic limb movement index, PLMS = Periodic limb movements of sleep, R&K = Rechtschaffen and Kales criteria for sleep scoring, S1 = Stage 1 sleep, S2 = Stage 2 sleep, SD = standard deviation.

Appendix 3.

Table A3. *Tested Device Specifications.*

Study	Wearable Device Type	Wearable Device Description	
		Device Name	Scoring Software and Algorithm Description
Ajilore et al 1995	EOG and actigraphy	Nightcap	Computer software not described
Alsaadi et al 2014	actigraphy	SenseWear Pro 3 Armband Actiwatch-2	Armband: SenseWear Professional software v 6.1, 60s epochs Actiwatch: ActiWare software v 5.52.0003 (Philips Respironics), 30-s epochs, medium threshold, 10 immobile minutes criterion
Baandrup & Jennum	actigraphy	Actiwatch Spectrum	Actiware software v 6.0.0 (Respironics), 30s epochs, medium sensitivity
Blackwell et al 2011	actigraphy	Sleepwatch-O	Action W-2 software, Cole et al. (1992) algorithm with ZCM, University of California San Diego (Jean-Louis, Kripke, Mason, Elliott, & Youngstedt, 2001) algorithm for PIM and TAT
Burnett et al 1985	actigraphy and ECG	Vitalog PMS-8 to store the data Actigraphy: switches mounted on a cube ECG: three chest electrodes	Two independent scorers and their scores were averaged. Sleep onset as 10mins of inactivity and/or reduced heart rate VSTAT program also automatically scored the data

Cantero et al 2002	EOG	Nightcap	Scored on device
Cellini et al 2013	actigraphy	Actiwatch-64 GT3X+	Actiwatch: Actiware 5.52.0003 (Philips Respironics) at low, medium and high thresholds, 60s epochs GT3X+: ActiLife 6.4.3 (Actigraph) using Sadeh et al. (1994) algorithm, 60s epochs
Cellini et al 2015	EEG	Zeo wireless system	bedside base station, 30s epochs, algorithm follows R&K criteria
Chae et al 2009	actigraphy	Actiwatch-L	Actiware v 5.0 (Minimitter-Respironics Inc), 15s epochs, all thresholds and sensitivities
Chakar et al 2017	actigraphy	Actiwatch-2	Actiware v 6.0.1, 30s epochs with default settings
Choi et al 2017	actigraphy	Actiwatch-2	Actiware v 5.70, 30s epochs
Cole et al 1992	actigraphy	Motionlogger Actigraph	Algorithm developed in this study, 2s epoch, ZCM
Cook et al 2017	actigraphy	Actiwatch-2 Fitbit Flex	Actiwatch: medium threshold, 10 immobile minutes Fitbit: website (version not provided), normal and sensitive settings
Cook et al 2018	actigraphy	Actiwatch-2 Jawbone UP3	Actiwatch: 30s epochs, medium sensitivity, 5 immobile minutes UP3: UP Android app v 4.24, 60s epochs

de Souza et al 2003	actigraphy	Mini Motionlogger Basic Actigraph	ZCM, 60s epochs, 5 immobile minutes, Cole et al. (1992) (Action 3 v 3.15 software) and Sadeh et al. (1994) (Action for Windows v 1.05) algorithms
de Zambotti et al 2015	actigraphy	Jawbone UP	UP mobile app (version not provided)
de Zambotti et al 2018	actigraphy	Fitbit Charge 2	Fitbit Inc. provided sleep stage data, 30s epochs
Dick et al 2010	actigraphy	SOMNOwatch	DOMINO light software, Gorney (1997) algorithm, activity threshold set to 28 units, 'number of epochs before/after' set to 4 epochs, 30s epochs
Dunican et al 2018	actigraphy	GT3X+ Readiband version 3	GT3X+: ActiLife 6 software, 60s epochs, Cole et al. (1992) algorithm Readiband: proprietary algorithm, Readiband Sync software
Edinger et al 2004	EOG and actigraphy	REMview Mini-Mitter Actiwatch device	REMview: RV software using proprietary algorithm Actiwatch: proprietary algorithm, 60s epochs
Farabi et al 2017	actigraphy	Actiwatch-2	Actiware 6, 30s epochs, 10 immobile minutes, all thresholds
Fietze et al 2015	actigraphy and EEG	SOMNOwatch Study A: F4-M1 EEG channel	Manual scoring, R&K criteria

		Study B: EEG, EMG and EOG	
Finan et al 2016	actigraphy and EEG	X4 Sleep Profiler headband	Sleep Profiler Portal software, sleep onset as four consecutive sleep epochs in the first 5 minutes of recording
Fonseca et al 2017	actigraphy and PPG	Device containing PPG and accelerometer Actiwatch Spectrum	PPG Device: Authors' algorithm, 30s epochs Actiwatch: Actiware software with the default settings
Fuller et al 2017	actigraphy	Actical Z-series	Actiware v 5.61, all thresholds
Griessenberger et al 2013	EEG	Zeo wireless system	Zeo device
Gruwez et al 2017	actigraphy	Jawbone UP MOVE Withings Pulse 02 SenseWear Pro Armband	Accompanying smartphone apps (versions not provided)
Hauri 1999	Response to a stimulus	Sleep switch device	Data read off display on the device (acts like a stopwatch)
Hedner et al 2004	PAT	Watch_PAT100 system with actigraph (actigraphy-measured sleep)	Authors' algorithm (see paper for details), 30s epochs

Hedner et al 2011	actigraphy	Watch_PAT100 system (PAT-measured sleep)	Authors' algorithm (see paper for details), 30s epochs
Insana et al 2011	actigraphy	Actiwatch-64	Actiware v 5.5, 15s epochs, all thresholds and three immobile minute criteria (2, 5, 10mins)
Kanady et al 2011	actigraphy	Actiwatch-64	Automatic rest interval detection from Actiware 5.52.0003 software, 60s epochs, 3 AMRI sensitivities (high, medium, low) and 2 interval durations (15, 40mins). Once rest intervals were established, used the default settings to score sleep.
Kang et al 2017	actigraphy	Fitbit Flex Actiwatch-2	Fitbit: website, 30s epochs Actiwatch: Actiware v 6.0.8, default settings
Kapella et al 2017	actigraphy	Actiwatch-2	Actiware v 6.0.8, 30s epochs, 5 thresholds (0, 5, 10, 20 and 40) and 3 immobile minute (5, 10, 15mins) criteria
Kaplan et al 2012	actigraphy	Actiwatch-64	Actiware v 5.57, all three thresholds (low, med, high), 30s epochs, immobile minutes criteria (number of minutes not reported)
Kaplan et al 2014	EEG	Zmachine	Automated sleep–wake detection algorithm (Z-ALG), 30s epochs, algorithm described in paper

Kosmadopoulos et al 2014	actigraphy device and EEG device	Actiwatch-64 Actical Z-series Zeo wireless system	Actiwatch and Actical: Actiware v 3.4, 30s epochs, four thresholds (low, medium, high and custom 10), algorithm described in paper, 10 immobile minutes Zeo: Zeo automatic software (no version given)
Kuo et al 2017	actigraphy	A device the authors created	See paper for algorithm details
Laakso et al 2004	actigraphy	Actiwatch	Actiwatch Sleep Analysis software v 4.15, 60s epochs, medium and high thresholds, 10 immobile minutes
Lichstein et al 2006	actigraphy	Actiwatch-64	Actiware Sleep v. 3.3, 30s epochs, high threshold
Lucey et al 2016	EEG	Sleep Profiler	Manual scoring using the Sleep Profiler Manual
Maglione et al 2013	actigraphy	Actiwatch-L	Actiware v 5.0, 30s epochs, 7 thresholds (0, 5, 10, 20, 40, 60, 80) and 3 immobile minute (0, 5, 10mins) criteria
Markwald et al 2016	actigraphy	Actiwatch-64 Zeo wireless system	Actiwatch: Actiware v 3.3, 30s epochs, medium sensitivity Zeo: proprietary algorithm, 30s epochs
Matsuo et al 2016	actigraphy	Actiwatch-2 MTN-210	Actiwatch: Actiware v 6.0.1, all three thresholds, 2min epochs MTN: SleepSign Act software, default algorithm
McCall & McCall	actigraphy	Actiwatch-64	Actiware v 5.0, 30s epochs, medium threshold

Mundt et al 2016	actigraphy	Actiwatch-2	Actiware v.5.3.2, 30s epochs, high sensitivity, Oakley (1997) algorithm
Myllymaa et al 2016	EEG	Forehead EEG, zygomatic (Sp1, Sp2) and mastoid (T9, T10), right-EOG, ECG and pulse oximetry	Manual scoring using AASM criteria, RemLogic software
Nakazaki et al 2014	actigraphy	FS-750 actigraph worn on waist	2-min epochs, algorithm described in paper
O'Hare et al 2015	actigraphy	Actiwatch-2	Information not provided
Paquet et al 2007	actigraphy	Actiwatch-L	Actiware v 5.0., 60s epochs, 2 thresholds (20 and 40) using algorithms by Actiwatch-L manufacturers. Two additional algorithms described in paper. All algorithms defined sleep onset as first epoch of 10mins of sleep
Pigeon et al 2018	actigraphy	myCadian watch Actiwatch-2	myCadian: CURA System, 30s epochs Actiwatch: Actiware v 6.0.2, medium sensitivity, 30s epochs
Razjouyan et al 2017	actigraphy	Actiwatch-L	Actiwatch-L: Action4 v 1.16, 60s epochs, Cole et al. (1992) algorithm
Reid & Dawson	actigraphy	Z80-32k V1 activity monitor	30s epochs, sleep onset as first epoch of 5mins of inactivity

Rupp & Balkin	actigraphy	Motionlogger Watch Actiwatch-64	Motionlogger: Action-W v 2, Cole et al. (1992) algorithm, 30s epochs, ZCM and default settings Actiwatch: Actiware v 3.4, 30s epochs
Sanchez-Ortuno et al 2010	actigraphy	Mini-Mitter Actiwatch devices	Actiwatch software, medium threshold, 60s epochs, 10 immobile minutes
Sargent et al 2016	actigraphy	Philips Respironics activity monitor	Actiware v 3.1, 60s epochs split into 30s epochs for analysis, default algorithm using all 3 thresholds (low, medium, high)
Scatena et al 2012	actigraphy	Actiwatch-64	Actiwatch software v 7.31, Sadeh et al. (1994) algorithm
Scott et al 2018	Response to a stimulus	Sleep On Cue smartphone application	Scored on device
Senny et al 2012	jaw movement, oxygen saturation, nasal airflow	Somnolter device records nasal airflow, SpO2, body position and jaw movements	Somnolter analysis software
Shambroom et al 2012	actigraphy device and EEG device	Zeo wireless system Actiwatch-64	Zeo: scored on device Actiwatch: Actiware v 5.0, 30s epochs, medium threshold

Signal et al 2005	Actigraphy	Actiwatch	Actiware v 3.14, 60s epochs, all three thresholds, sleep onset as start of 20 epochs where 19 were scored as sleep
Sivertsen et al 2006	actigraphy	Actiwatch Plus	Actiwatch software v 1.19, 30s epochs, medium sensitivity
Slater et al 2015	actigraphy	GT3X+	ActiLife 6.8 software, 60s epochs, Sadeh et al. (1994) algorithm
Taibi et al 2013	actigraphy	Actiwatch-64	Actiware v 5.57, 30s epochs, 10 immobile minutes
Tonetti et al 2013	EEG	Zeo wireless system	Scored on device
Tonetti et al 2008	actigraphy	Basic Mini-Motionlogger Actiwatch	Motionlogger: Action W-2 v 3.23, 60s epochs, Cole et al. (1992) algorithm Actiwatch: Actiwatch software v 5.32, 60s epochs, Oakley (1997) algorithm
Vallieres & Morin	actigraphy	IM Systems actigraphy device	IM Systems software and algorithm (version 3.15a)
Wang et al 2008	actigraphy	Actiwatch-64	Actiware v 5.0, 30s epochs, medium threshold, 10 immobile minutes
White et al 1995	Various physiological measures	NightWatch System: EOG, leg movement, oxygen saturation, nasal airflow, chest and abdomen	Algorithm described in paper

		motion, body position, movement and heart rate	
Zhang et al 2014	EEG	A wireless headband	Algorithm described in paper
Zinkhan et al 2014	actigraphy	SOMNOwatch plus (one on the hip, one on the wrist) GT3X+ (placed on hip)	SOMNOwatch: Domino Light software, Dick et al. (2010a) algorithm GT3X+: ActiLife v 5, Cole et al. (1992) algorithm

EEG = electroencephalography, EOG = electrooculography, ECG = electrocardiography, PIM = proportional integration mode, TAT =time above threshold, ZCM = zero-crossing mode

Appendix 4.

Table A4. *Overview of the Accuracy of the Wearable Devices, separated by Type of Device.*

Actigraphy Devices					
Study	**Sample Type**	**Device Tested**	**Results**		
			Sleep Onset Latency Means (SD)	**Discrepancy (SD) between PSG and Device, or Bland-Altman Plot statistics**	**Correlations between PSG and Device [95% CIs]**
Alsaadi et al 2014	Other health conditions (lower back pain)	SenseWear Pro 3 Armband Actiwatch-2	PSG: M = 15.19 (SD = 14.23, range: 2-73) Armband: M = 15.23 (SD = 24.98, range: 0-124) Actiwatch: M = 4.46 (SD = 8.80, range: 0-51)		Armband: ICC = 0.13, [-0.15, 0.39] Actiwatch: ICC = 0.33, [-0.05, 0.63]
Baandrup & Jennum	Other health conditions (mental health conditions)	Actiwatch Spectrum	PSG: M = 21 (SD = 31)	M = -2 (SD = 45), n.s.	ICC = 0.00, [0.00, 0.19]

Blackwell et al 2011	Other health conditions (elderly men with osteoporotic fractures)	Sleepwatch -O	PSG: M = 12.53 (SD = 20.53) PIM: M = 9.76 (SD = 17.27) TAT: M = 10.11 (SD = 22.58) ZCM: M = 29.88 (SD = 52.78)	PIM: M = -2.77 (SD = 22.03) TAT: M = -2.43 (SD = 27.82) ZCM: M = 17.56 (SD = 51.37)	PIM: $r_{(s)}$ = 0.44, ICC = 0.32, [0.26, 0.38] TAT: $r_{(s)}$ = 0.39, ICC = 0.17, [0.10, 0.23] ZCM: $r_{(s)}$ = 0.36, ICC = 0.12, [0.06, 0.19]
Cellini et al 2013	Healthy	Actiwatch-64 GT3X+	PSG: M = 10.60 (SD = 5.88) Actiwatch-64: M = 14.37 (SD = 15.36), n.s. GT3X+: ACT M = 7.67 (SD = 5.44), n.s. LFE M = 7.8 (SD = 5.40), n.s.	Actiwatch-64: M = 3.77 (SD = 13.76), n.s. GT3X+: ACT M = -2.93 (SD = 4.98)* LFE M = -2.8 (SD = 4.94) *	Actiwatch-64: ICC = 0.29 GT3X+: ACT ICC = 0.55, LFE ICC = 0.56
Chae et al 2009	Sleep disorders (OSA and OSA+PLMS)	Actiwatch-L	PSG: M = 5.73 (SD = 5.43) ACT: 4min ACT: M = 3.59 (SD = 4.05) 5min: M = 5.76 (SD = 5.40) 6min: M = 16.99 (SD = 34.53) 10min: M = 25.86 (SD = 40.72)		4min ACT: r = 0.44, [0.12, 0.68] 5min: r = 0.65, [0.40, 0.82] 6min: r = 0.22, [-0.13, 0.52] 10min: r = 0.29, [-0.06, 0.58] 15min: r = 0.39, [0.05, 0.65]

			15min: M = 44.61 (SD = 66.67)	
Chakar et al 2017	Healthy	Actiwatch-2	PSG: M = 25.2 (SD = 21.8) ACT: M = 3.0 (SD = 2.1)	r = .01*
Choi et al 2017 (Choi et al., 2017)	Sleep disorders (sleep disordered breathing, insomnia)	Actiwatch-2	Sleep disordered breathing group, PSG: M = 11.7 (SD = 13.98) ACT: M = 8.7 (SD = 14.6), n.s. Insomnia group, PSG: M = 21.7 (SD = 18.87) ACT: M = 6.6 (SD = 14.67)***	Sleep disordered breathing group, ICC = 0.24, [-0.50, 0.62], n.s. Insomnia group, ICC = 0.70***, [0.36, 0.86]
Cole et al 1992	Controls and sleep disorders, mental health conditions	Motionlogger Actigraph	PSG: M = 59.2 (SD = 46.1) ACT: M = 50.1 (SD = 50.7), n.s.	r = .90***

Cook et al 2017	Other health conditions (MDD)	Actiwatch-2 Fitbit Flex	PSG: M = 19.2 (SD = 22.7) Actiwatch-2: M = 5.8 (SD = 7.7) Fitbit-Normal: M = 17.2 (SD = 14.2) Fitbit-Sensitive: M = 30.7 (SD = 28.6)	Actiwatch-2: M = −13.5* Fitbit-Normal: M = −2.0, n.s. Fitbit-Sensitive: M = 11.5*	
Cook et al 2018	Sleep disorders (narcolepsy, hypersomnia)	Actiwatch-2 Jawbone UP3	PSG: M = 16.8 (SD = 23.3) UP3: M = 11.7 (SD = 16.6) Actiwatch-2: M = 3.98 (SD = 8.61)	UP3-PSG: M = -5.13, n.s. Actiwatch-2-PSG: M = -12.9***	
de Souza et al 2003	Healthy	Mini Motion-logger Basic Actigraph	PSG: M = 6.9 (SD = 4.5) Cole: M = 8.3 (SD = 4.7) Sadeh: M = 9.4 (SD = 5.1)	Cole: bias = 1.3 (SD = 3.6) Sadeh: bias = 2.4 (SD = 4.0)	Cole: r = 0.69 Sadeh: r = 0.64
de Zambotti et al 2015	Healthy and sleep disorders	Jawbone UP	PSG: M = 9.1 (SD = 6.9) UP: M = 14.3 (SD = 10.1)**	M = 5.2 (SD = 9.6), lower limit: -24.1, upper limit: 13.7	

de Zambotti et al 2018	Healthy and sleep disorders (PLMS)	Fitbit Charge 2	Main group, PSG: M = 14 (SD = 11), [10, 17] Charge: M = 9 (SD =6), [7, 11], t = −2.70* PLMS group, PSG: M = 15 (SD = 13), [4, 25] Charge: M =8 (SD = 5), [5, 12], t = –1.91, n.s.	B-A plot, Main group, Bias = 4 (SD = 9), [1, 8] lower limit: −14, upper limit: 23 PLMS group, bias = 7 (SD = 10), [1, 15] lower limit: –14, upper limit: 27	
Dick et al 2010	Sleep disorders (OSA)	SOMNOwatch	PSG: M = 19 ACT: M = 14, t(28)= −3.249***		r = 0.89***
Dunican et al 2018	Representative community samples	GT3X+ Readiband v3	PSG: M = 18 (SD = 18) ACT: M = 4 (SD = 4)* Readiband: M = 40 (SD = 32)*	ACT: M = -14 (SD = 35)* Readiband: M = 22 (SD = 74)*	

Edinger et al 2004	Sleep disorders	Mini-Mitter Actiwatch device	PSG: M = 28.7 (SD = 41.9) ACT: M = 31.9 (SD = 67.1), n.s.		ACT: r = .87**
Farabi et al 2017	Other health conditions (T1 diabetes)	Actiwatch-2	PSG: M = 36.2 (SD = 18.4) ACT: M = 18.1 (SD = 24.3)*	B-A plot: bias = −18.1 (SD = 25.3) lower limit: -28.1, upper: -8.1	
Fonseca et al 2017	Healthy	Actiwatch Spectrum	PSG: M = 15.53 (SD = 8.23)	Subset of Set 2, ACT-PSG: M = −8.59 (SD = 9.05)	
Fuller et al 2017	Healthy	Actical Z-series	PSG: M = 16.0 (SD = 15.5) Actical: M = 6.3 (SD = 8.3)*	B-A plot, bias = -9.5 (SEE = 15.2), [−13.4, 5.7]	r = 0.24
Gruwez et al 2017	Healthy	Jawbone UP Withings Pulse 02	PSG: M = 14 (SD = 13) Withings: M = 13 (SD = 4), n.s. UP: M = 20 (SD = 10), n.s. Armband: M = 11(SD = 9), n.s.		Withings: n.s. UP: n.s. Armband: r = 0.5*

		SenseWear Pro Armband	
Hedner et al 2004	Healthy and sleep disorders (OSA)	Watch_PAT 100 with actigraph (actigraphy-measured sleep)	In 30-s epochs, Normal group, PSG: M = 51.2 (SD = 52.6) Watch_PAT: M = 62.2 (SD = 33.2), n.s. Mild group, PSG: M = 37.8 (SD = 38.8) Watch_PAT: M = 54.4 (SD = 27.1)* Moderate group, PSG: M = 39.9 (SD = 36.7) Watch_PAT: M = 54.4 (SD = 30.8)* Severe group, PSG: M = 48.6 (SD = 57.0) Watch_PAT: M = 59.1 (SD = 35.2), n.s.

			All groups, PSG: M = 43.3 (SD = 45.4) Watch_PAT: M = 56.8 (SD = 31.4)*		
Insana et al 2011	Disrupted sleep, otherwise healthy	Actiwatch-64	PSG: M = 9.96 (SD = 4.64), range: 3.13–20.00 2-min ACT: M = 0.67 (SD = 0.69), range: 0.00–2.69, t(40) = 12.80*** 5-min: M = 2.18 (SD = 2.46), range: 0.06–10.31, t(40) = 9.50*** 10-min: M = 6.46 (SD = 5.60), range: 0.13–20.00, t(40) = 3.29**	2-min: M = -9.29 (SD = 4.65) 5-min: M = -7.82 (SD = 5.23) 10-min: M = -3.49 (SD = 6.79)	2-min: r = .07, n.s. 5-min: r = .01, n.s. 10-min: r = .13, n.s.
Kanady et al 2011	Healthy	Actiwatch-64	PSG: M = 11.9 (SD = 9.6) high-15: M = 7.6 (SD = 12.1) high-40: M = 7.7 (SD = 12.3) med-15: M = 2.5 (SD = 4.6) med-40: M = 2.0 (SD = 4.8) low-15: M = 0.4 (SD = 1.2) low-40: M = 0.5 (SD = 1.4)	B-A plot, high-15: bias = -3.13 (SD = 12.64) upper limit: 1.39, lower limit: -7.65 high-40:	high-15: r = 0.726*** high-40: r = 0.319, n.s. med-15: r = 0.319, n.s. med-40: r = 0.185, n.s. low-15: r = 0.185, n.s. low-40: r = 0.396, n.s.

				bias = -3.38 (SD = 12.79), upper limit: 1.27, lower limit: -8.03	
Kang et al 2017	Healthy and sleep disorders (insomnia)	Fitbit Flex Actiwatch-2	In insomnia group, PSG: M = 15.6 (SD = 13.6) ACT: M = 20.4 ± 25.6, Z =−0.51, n.s. Flex-N: M = 13.2 (SD = 9.8), t=1.22, n.s. In good sleeper group, PSG: M = 10.2 (SD = 10.2) ACT: M = 13.8 (SD = 16.4), Z =−0.03, n.s. Flex-N: M = 10.9 (SD = 11.2), Z =−0.47, n.s.	Fitbit-Normal setting: in good sleepers, M = 0.7 in insomnia group, M = -2.4	In insomnia group, Flex-N: ICC = 0.673, [0.35, 0.84] Flex-S: ICC = 0.403, [−0.14, 0.70] ACT: ICC = 0.737, [0.48, 0.87] In good sleeper group, Flex-N: ICC = 0.865, [0.62, 0.95] Flex-S: ICC = 0.581, [−0.08, 0.85] ACT: ICC = 0.586, [−0.12, 0.85]

Kapella et al 2017	Other health conditions (COPD)	Actiwatch-2	PSG: M = 34.5 (SD = 31.4) ACT: M = 29.0 (SD = 29.1), n.s.	
Kaplan et al 2012	Controls and other health conditions (bipolar)	Actiwatch-64	Night 1 in bipolar group, PSG: M= 12.6 (SD = 13.7) ACT: M = 11.8 (SD = 14.4), n.s. Night 2 in bipolar group, PSG: M = 15.9 (SD = 17.9) ACT: M = 12.8 (SD = 21.0), n.s. Night 1 in control group, PSG: M = 10.8 (SD = 11.7) ACT: M = 18.2 (SD = 28.0), n.s. Night 2 in control group, PSG: M = 11.6 (SD = 13.4) ACT: M = 10.0 (SD = 15.9), n.s.	bipolar group, r = .33* control group, r = .41**
Kosmadopoulos et al 2014	Healthy	Actiwatch-64	PSG: M = 27.1 (SD = 28.5) Actiwatch-64: M = 11.1 (SD = 12.3)* Actical: M = 5.1 (SD = 7.6)*	

		Actical Z-series			
Kuo et al 2017	Healthy and poor sleepers	Author-developed actigraphy device	PSG: M = 15.4 (SD = 22.2) ACT: M = 16.1 (SD = 16.7)	B-A plot, bias = 0.74 (SD = 19.16)	ICC = 0.53
Laakso et al 2004	Controls and other health conditions (sleep disorders, motor disabilities)	Actiwatch	able-bodied group, PSG: M = 14, range: 1-73 sleep-disordered, PSG: M = 48, range: 8-258 sleep-disordered with motor disabilities, PSG: M = 121, range: 7-606	able-bodied group, M = -6 (SD = 7) sleep-disordered, M = -48 (SD = 62) sleep-disordered with motor disabilities, M = -152 (SD = 194)	able-bodied group, r = 0.82** sleep-disordered, r = 0.73** sleep-disordered with motor disabilities, r = 0.17, n.s
Lichstein et al 2006	Sleep disorders (insomnia)	Actiwatch-64	PSG: M = 17.5 (SD = 12.6) ACT: M = 17.3 (SD = 26.8), n.s.		r = .30, n.s.
Maglione et al 2013	Other health conditions	Actiwatch-L	PSG: M= 14.45 (SD = 19.24) ACT-0 min: M = 0 (SD = 0)	0min: M = 14.45 (SD = 19.24)***	

	(Parkinson's Disease)		ACT-5 min: M = 9.70 (SD = 32.49) ACT-10 min: M = 13.80 (SD = 28.90)	5min: M = 4.75 (SD = 31.07), n.s. 10min: M = 0.65 (SD = 27.44), n.s.	
Markwald et al 2016	Healthy	Actiwatch-64	PSG: M = 14.7 (SD = 17.3) ACT: M = 2.4 (SD = 1.8)* Sleep Efficiency < 85%, PSG: M = 21.0 (SD = 23.4) ACT: M = 2.2 (SD = 1.4)* Sleep Efficiency > 85%, PSG: M = 9.6 (SD = 7.6) ACT: M = 2.6 (SD = 2.1), n.s.		
Matsuo et al 2016	Healthy	Actiwatch-2 MTN-210	PSG: M = 6.20 (SD = 0.43) ACT80: M = 8.50 (SD = 3.09) ACT40: M = 12.60 (SD = 3.54) ACT20: M = 15.00 (SD = 3.67) MTN-Body: M = 21.50 (SD = 4.86)* MTN-Wrist: M = 29.40 (SD = 5.77)**	ACT80: bias = 2.3, [−24.33, 28.93] ACT40: bias = 8.5, [−63.3, 80.3] ACT20: bias = 8.8, [−22.62, 40.22]	ACT80: r = 0.1851, n.s. ACT40: r = 0.2434, n.s. ACT20: r = 0.2518, n.s. MTN-B: r = 0.1626, n.s. MTN-W: r = 0.06577, n.s.

				MTN-B: bias = 15.3, [−26.87, 57.47] MTN-W: bias = 23.2, [−27.25, 73.65]	
McCall & McCall	Sleep disorders (insomnia + MDD)	Actiwatch-64	PSG: M = 28.4 (SD = 36.9) PSG-Latency to persistent sleep: M = 35.2 (SD = 37.6) ACT: M = 24.2 (SD = 28.0)	PSG-ACT: M = -4.19 (SE = 5.3), n.s Latency to persistent sleep, ACT: M = -11.0 (SE = 4.8)*	PSG-ACT: r = 0.31* Latency to persistent sleep, ACT: r = 0.44***
Mundt et al 2016	Sleep disorders (insomnia + fibromyalgia)	Actiwatch-2	PSG: M = 25.73 (SD = 41.63) ACT: M = 46.93 (SD = 52.26)*		r = 0.08, n.s.
Nakazaki et al 2014	Healthy	FS-750 actigraph	PSG: M = 11.8 (SD = 2.9) ACT: M = 12.7 (SD = 2.7)		ICC = 0.403**
O'Hare et al 2015	Healthy	Actiwatch-2	PSG: M = 20 (SD = 13) ACT: M = 3 (SD = 2)*		r = 0.214, n.s.
Paquet et al 2007	Healthy	Actiwatch-L	Nocturnal sleep PSG: M = 21.2 (SD = 33.6)		

	ACT-40 threshold: M = 7.3 (SD = 7.6) ACT-20 threshold: M = 12.3 (SD = 9.4) Lot-reg algorithm: M = 8.0 (SD = 9.7) Lot-coeff algorithm: M = 9.6 (SD = 10.6) Day recovery sleep, PSG: M = 5.5 (SD = 6.5) ACT-40 threshold: M = 4.1 (SD = 8.2) ACT-20 threshold: M = 4.5 (SD = 8.4) Lot-reg algorithm: M = 3.1 (SD = 6.3) Lot-coeff algorithm: M = 3.4 (SD = 6.4) Caffeine recovery sleep, PSG: M = 11.7 (SD = 17.1) ACT-40 threshold: M = 3.5 (SD = 4.0) ACT-20 threshold: M = 4.3 (SD = 4.0) Lot-reg algorithm: M = 3.4 (SD = 4.3)

			Lot-coeff algorithm: M = 3.7 (SD = 4.5)		
Pigeon et al 2018	Healthy	myCadian watch Actiwatch-2	PSG: M = 42.9 (SD = 27.2) myCadian: M = 55.7 (SD = 55.2), n.s. Actiwatch: M = 3.0 (SD = 2.5)** Latency to persistent sleep, PSG: M = 47.8 (SD = 28.3) myCadian: M = 55.9 (SD = 55.2), n.s. Actiwatch: M = 35 (SD = 29.7)*		
Razjouyan et al 2017	Sleep disorders (self-reported)	Actiwatch-L	PSG: M = 22.1 (SD = 19.2)		ACT: r = .04, n.s.
Reid & Dawson	Healthy	Z80-32k V1 activity monitor		Overall, Young group: M = 6.9-20.9 Older group: M = 6.8-11.4	
Rupp & Balkin	Healthy	Actiwatch-64	Baseline night, PSG: M = 13.7 (SD = 10.6)	Baseline night,	

		Motionlogger Watch	Motionlogger: M = 14.8 (SD = 9.6), n.s. Actiwatch: M = 6.9 (SD = 5.6)*** Recovery night, PSG: M = 5.0 (SD = 5.5) Motionlogger: M = 9.1 (SD = 7.0)*** Actiwatch: M = 4.9 (SD = 3.9), n.s.	Actiwatch: M = −6.85 (SD = 1.71) Recovery night, Motionlogger: M = 4.05 (SD = 0.87)	
Sanchez-Ortuno et al 2010	Healthy and sleep disorders (insomnia)	Mini-Mitter Actiwatch devices	Night 1: In insomnia group, PSG: M = 28.92 (SD = 28.20) ACT: M = 17.80 (SD = 15.60) in normal sleepers, PSG: M = 10.75 (SD = 7.14) ACT: M = 10.43 (SD = 10.10)	No significant difference in accuracy of ACT between insomnia group and normal sleeper group.	insomnia group, between-subjects: r = 0.57*** within-subjects: r = 0.43** normal sleepers, between-subjects: r = 0.80*** within-subjects: r = 0.41**
Sargent et al 2016	Healthy	Philips Respironics	PSG: M = 18.3 (SD = 12.6) ACT: M = 17.3 (SD = 14.2)	M = -0.9 (SD = 14.0), n.s.	

		activity monitor			
Scatena et al 2012	Healthy	Actiwatch-64	PSG: M = 14.1 (SD = 17.5) ACT: M = 13.9 (SD = 18.6)		kendall W coefficient = 0.724
Shambroom et al 2012	Healthy	Actiwatch-64	PSG1: M = 12.7 (SD = 3.1) PSG2: M = 9.7 (SD = 2.0) ACT: M = 2.4 (SD = 0.6)* PSG1-LPS: M = 18.4 (SD = 4.2) PSG2-LPS: M = 22.4 (SD = 4.6) ACT: M = 9.5 (SD = 2.5)*		Normal PSG-SOL criteria, ACT-PSG1: ICC = -0.07 ACT-PSG2: ICC = 0.13 Latency to persistent sleep, ACT-PSG1: ICC = 0.40 ACT-PSG2: ICC = 0.22
Signal et al 2005	Disrupted sleep	Actiwatch	In Flight, PSG: M = 9 (SD = 7) ACT: M = 9 (SD = 12) Layover, PSG: M = 6 (SD = 4) ACT: M = 8 (SD = 6)	In flight: M = 2.2, [-13.2, 17.6] Layover: M = 0.1, [-22.5, 22.7]	In flight: r = .06 Layover: r = .40

Sivertsen et al 2006	Sleep disorders (insomnia)	Actiwatch Plus	Before treatment, PSG: M = 23.8 (SEM = 24.7) ACT: M = 11.6 (SEM = 19.6) After treatment, PSG: M = 15.7 (SEM = 20.9) ACT: M = 10.8 (SEM = 18.1)		
Slater et al 2015	Healthy	GT3X+	PSG: M = 18.8 (SD = 18.0) Wrist-ACT: M = 11.5 (SD = 13.1)* Hip-ACT: M = 3.7 (SD = 9.3)*	B-A plot, Wrist-ACT: bias = −7.3 (SD = 18.0), upper limit: 3.9, lower limit: -10.7 Hip-ACT: bias = −15.1 (SD = 19.4), upper limit: -11.5, lower limit: -18.8	Wrist-ACT: ICC = 0.32 Hip-ACT: ICC = 0.05
Taibi et al 2013	Sleep disorders (insomnia)	Actiwatch-64	PSG: M = 18.36 (SD = 12.31) ACT: M = 4.98 (SD = 4.80)	M = -13.39	

Tonetti et al 2008	Healthy	Basic Mini-Motionlogger Actiwatch	PSG: M = 12.96 (SD = 11.95) Motionlogger: M = 5.61 (SD = 6.81)*** Actiwatch: M = 8.51 (SD = 12.01)***	Motionlogger: M = -7.35 Actiwatch: M = -4.45	
Vallieres & Morin	Sleep disorders (insomnia)	IM Systems actigraphy device	Across all nights, PSG: M = 12.67 (SD = 2.21) ACT: M = 4.57 (SD = 0.73)*	night 1: M = -14.16 (SD = 21.59), cohen's d = -1.79 night 2: M = -7.01 (SD = 11.77), cohen's d = -2.50	
Wang et al 2008	Healthy and sleep disorders (OSA)	Actiwatch-64	Overall, PSG: M = 24.74 (SD = 29.91) ACT: M = 27.60 (SD = 53.77), n.s. Non-OSA, PSG: M = 39.60 (SD = 37.58) ACT: M = 41.70 (SD = 75.20), n.s. OSA, PSG: M = 11.23 (SD = 9.79) ACT: M = 14.77 (SD = 17.70), n.s.		Overall: ICC = 0.35 Non-OSA: ICC = 0.20 OSA: ICC = 0.31

Zinkhan et al 2014	Community sample	SOMNOwatch plus GT3X+	PSG: M = 18.6 (SD = 18.3) SOMNO-wrist: M = 10.2 (SD = 18.7) SOMNO-hip: M = 0.4 (SD = 2.7) GT3X+: M = 12.5 (SD = 46.1)	SOMNO-wrist: M = 7.7 (SD = 25.9) SOMNO-hip: M = 16.7 (SD = 18.2) GT3X+: M = 6.4 (SD = 46.8)	

Actigraphy plus a Physiological Signal

Study	**Sample Type**	**Type of Device**	**Device Tested**	**Results**		
				Sleep Onset Latency Means	**Discrepancy between PSG and device, or Bland-Altman Plot statistics**	**Correlations**
Ajilore et al 1995	Healthy	actigraphy and EOG	Nightcap	Nightcap: M = 21 (SD = 24) PSG: M = 14 (SD = 12) t = 1.37, n.s.		
Burnett et al 1985	Sleep disorders	actigraphy and ECG	Vitalog PMS-8	Night 1, PSG: M = 23 (SD = 19) PSG-N2: M = 33 (SD = 20)		

				Device Manual scoring: M = 26 (SD = 21) Device Automatic scoring: M = 34 (SD = 42) Night 2, PSG: M = 17 (SD = 12) PSG-N2: M = 24 (SD = 17) Device Manual scoring: M = 22 (SD = 19) Device Automatic scoring: M = 22 (SD = 19)		
Edinger et al 2004	Sleep disorders	actigraphy and EOG	REMview	PSG: M = 28.7 (SD = 41.9) REMview: M = 59.4 (SD = 59.8)*		REMview: r = .82**
Fietze et al 2015	Sleep disorders	actigraphy and EEG	SOMNOwatch Study A: F4-M1 EEG channel	Study A, PSG: M = 26 SOMNOwatch: M = 29 Study B, PSG: M = 17	B-A plot, Study A, bias = 3 (SD = 6), lower limit: -10, upper: 16 Study B,	Study A, r = 0.98*** Study B, r = 0.87***

			Study B: EEG, chin EMG and EOG	SOMNOwatch: M = 21	bias = 4, lower limit: -12, upper: 20	
Finan et al 2016	Healthy	actigraphy and EEG	X4 Sleep Profiler headband	PSG: M = 8.04 (SD = 7.84) Sleep Profiler Automatic algorithm: M = 5.00 (SD = 5.52)* Sleep Profiler Manual scoring: M = 4.82 (SD = 6.18)*		Sleep Profiler Automatic algorithm: r = .59* Sleep Profiler Manual scoring: r = .45, n.s.
Fonseca et al 2017	Healthy	actigraphy and PPG	Device containing PPG and accelerometer	PSG: M = 15.53 (SD = 8.23)	M = −6.80 (SD = 7.69), lower limit: −21.88, upper limit: 8.28 Subset of Set 2, M = −7.48 (SD = 6.64)	
Hedner et al 2011	Controls and sleep disorders	PAT	Watch_PAT100 system (PAT-measured sleep)	In epochs, PSG: M = 57 (SD = 31) Watch-PAT: M = 43 (SD = 45)*		ICC = 0.57**

Physiology-based Devices						
Study	**Sample Type**	**Type of Device**	**Device Tested**	**Results**		
				Sleep Onset Latency Means	**Discrepancy between PSG and device, or Bland-Altman Plot statistics**	**Correlations**
Cantero et al 2002	Healthy	EOG	Nightcap			r = 0.98*
Cellini et al 2015	Healthy	EEG	Zeo wireless system	PSG: M = 9.85 (SD = 6.11) Zeo: M = 7.82 (SD = 5.68)*	B-A plot, Bias: -2.03 (SD = 4.95) upper limit: -0.26, lower limit: -3.80	ICC = 0.53***
Griessenberger et al 2013	Controls and sleep disorders	EEG	Zeo wireless system	PSG: M = 10.79 (SD = 11.18)		r = .072, n.s.
Kaplan et al 2014	Controls and sleep disorders	EEG	Zmachine		B-A plot, bias = −5.64, (SD = 12.12)	r = 0.962
Kosmadopoulos et al 2014	Healthy	EEG	Zeo wireless system	PSG: M = 27.1 (SD = 28.5) Zeo: M = 11.0 (SD = 11.0)*		

Lucey et al 2016	Sleep disorders	EEG	Sleep Profiler	PSG: M = 16.3 (SEM = 3.9) Sleep Profiler: M = 15.1 (SEM = 2.8)		ICC = 0.67
Markwald et al 2016	Healthy	EEG	Zeo wireless system	PSG: M = 14.7 (SD = 17.3) Zeo: M = 7.3 (SD = 9.5)* Sleep efficiency < 85%, PSG: M = 21.0 (SD = 23.4) Zeo: M = 7.3 (SD = 10.9)* Sleep efficiency > 85%, PSG: M = 9.6 (SD = 7.6) Zeo: M = 7.4 (SD = 8.6), n.s.		
Myllymaa et al 2016	Controls and sleep disorders	EEG	Frontal EEG, right-EOG, ECG and pulse oximetry	PSG-N1: M = 16.2 (SD = 14.3) Device: M = 15.4 (SD = 12.7), n.s. PSG-N2: M = 27.9 (SD = 18.6)	N1-Device: M = -0.8 (SD = 12.8), n.s. N2 Device: M = 0.4 (SD = 13.3), n.s.	N1-Device: ICC = 0.548 N2-Device: ICC = 0.771

				Device: M = 28.3 (SD = 20.5), n.s.
Senny et al 2012	Sleep disorders	jaw movements, oxygen saturation, nasal airflow	Somnolter device records nasal airflow, SpO2, body position and jaw movements	Overall, PSG-30s epoch: M = 33.1 (SD = 77.4), [14.7, 51.7]* PSG-15min epoch: M = 4.2 (SD = 59.6), [-20.4, 28.8] OSA group, PSG-30s epoch: M = 36.1 (SD = 88.7), [6.6, 65.4]* PSG-15min epoch: M = -1.4 (SD = 59.6), [-32.4, 34.8] Insomnia group, PSG-30s epoch: M = 29.7 (SD = 68.2), [-4.7, 64.1]* PSG-15min epoch: M = 8.8 (SD = 65.7), [-44.5, 62.1] Other group,

					PSG-30s epoch: M = 33.3 (SD = 59.5), [-6.0, 72.0]* PSG-15min epoch: M = 11 (SD = 33.9), [-33.3, 55.1]	
Shambroom et al 2012	Healthy	EEG	Zeo wireless system	PSG1: M = 12.7 (SD = 3.1) PSG2: M = 9.7 (SD = 2.0) Zeo: M = 7.8 (SD = 2.4), n.s. PSG1-LPS: M = 18.4 (SD = 4.2) PSG2-LPS: M = 22.4 (SD = 4.6) Zeo: M = 17.4 (SD = 4.0), n.s.		Normal PSG-SOL criteria, Zeo-PSG1: ICC = 0.42 Zeo-PSG2: ICC = 0.50
Tonetti et al 2013	Healthy	EEG	Zeo wireless system	PSG: M = 22.68 (SD = 13.74) Zeo: M = 15.86 (SD = 13.05), n.s.		

White et al 1995	Sleep disorders	Various physiological measures	NightWatch System	PSG: M = 15 (SE = 3) NightWatch: M = 14 (SE = 2)	r = 0.54*, [.22, .76]
Zhang et al 2014	Healthy	EEG	Single frontal electrode in headband	35/40 (87.5%) of sleep onsets detected within transition from N1-N2 sleep	

Behaviour-based Devices

Study	**Sample Type**	**Type of Device**	**Device Tested**	**Results**		
				Sleep Onset Latency Means	**Discrepancy between PSG and device, or Bland-Altman Plot statistics**	**Correlations**
Hauri 1999	Controls and sleep disorders (insomnia)	Depression of a switch	Sleep switch device	Switch: M = 32.4 (SD = 30.7) PSG-N1: M = 17.3 (SD = 13.0)** PSG-N2: M = 22.9 (SD = 16.8), n.s.	N1 sleep: M = 15.1 N2 sleep: M = 9.5 solid sleep: M = -1.8	N1 sleep: r = .60 N2 sleep: r = .55 solid sleep: r = .98

				PSG-solid sleep: M = 34.2 (SD = 33.3), n.s.	
Scott et al 2017	Healthy	Response to a stimulus	Sleep On Cue smartphone application	PSG-N1: M = 3.17 (SD = 3.04) PSG-6 epochs of sleep: M = 2.75 (SD = 3.11) PSG-N2: M = 0.81 (SD = 1.96)	PSG-N1: $r_{(s)}$ = 0.79*** PSG-6 epochs of sleep: $r_{(s)}$ = 0.81*** PSG-N2 sleep: $r_{(s)}$ = 0.92***

[] contain 95% confidence intervals, * p < .05, ** p < .01, *** p < .001, n.s. no significant difference.

ACT = actigraphy, COPD = Chronic obstructive pulmonary disease, ICC = intra-class correlation, M = mean, PAT = peripheral arterial tone, PSG = polysomnography, LFE = low-frequency extension algorithm option, LPS = latency to persistent sleep, N1 = Non-rapid eye movement Stage 1, N2 = Non-rapid eye movement Stage 2, n.s. = not significant, OSA = obstructive sleep apnea, PIM = proportional integration mode, PLMS = Periodic limb movements of sleep, SD = standard deviation, SE = standard error, SOL = sleep onset latency, TAT =time above threshold, ZCM = zero-crossing mode

Appendix 5.

Instructions to participants about operating THIM

To ensure that participants understood how to correctly use the device, they practiced undergoing a sleep onset trial whilst awake Research assistants read the following instructions to participants, demonstrating each action:

"To begin a trial, double-tap THIM by moving your index finger to meet your thumb twice. THIM will glow blue to signal that it has started the trial. THIM will start emitting low intensity vibrations. When you feel these vibrations, respond to THIM by giving a big finger twitch response, similar to the double-tap movement, but you only twitch your finger once. After a while, THIM will emit a very long, high intensity alarm vibration to signal the end of the trial. To turn this alarm off, double-tap THIM. THIM will glow a different colour other than blue, typically green, to signal that it has ended the trial. During Night 2 and Night 3, we will ask that you complete a very short questionnaire at this point at the end of the trial. Once you are ready to begin the next trial, double-tap THIM again and it will glow blue to signal that it has started. This process will be repeated for four hours of testing."

After hearing these instructions, participants were required to respond to five vibratory stimuli successfully with the appropriate finger taps. If participants did not successfully respond to two consecutive vibrations – typically by not tapping the device with enough force, or by responding to the stimulus too late – the device would administer the high intensity alarm vibration. If this occurred, participants were required to commence another trial and respond to five consecutive vibrations until they were able to do so without activating the high intensity alarm vibration. Participants were then prompted to not respond to the next two vibratory stimuli so that they could experience the high intensity alarm vibration and practice ending the

trial with a double-tap response.

Once the research assistant was certain that the participant understood the instructions and could operate THIM successfully, they stopped practicing. Participants repeated any aspects above and research assistants answered any questions, if necessary.

Appendix 6.

Table A6. *Good versus poor sleeper comparison.*

Interaction Terms	Inferential Statistics		
	F	*dfs*	*p*
Five-way interaction (night, sleeper status, sleep stage, time and behavioural response)			
Delta, %	0.73	12, 963.95	.72
Theta, %	0.39	12, 1011.68	.97
Alpha, %	0.47	12, 1016.81	.93
Sigma, %	0.83	12, 1040.12	.62
Beta, %	0.53	12, 1026.23	.90
Four-way interaction (sleeper status, sleep stage, time and behavioural response)			
Delta, %	0.58	12, 965.76	.86
Theta, %	0.82	12, 1016.65	.63
Alpha, %	1.03	12, 1022.25	.42
Sigma, %	1.11	12, 1048.06	.35
Beta, %	0.92	12, 1032.58	.52

Note: as a reminder, the fixed effects were Night (Night 2, Night 3), Sleeper Status (good, poor sleeper), sleep stage (wake, N1 or N2), time (the five-second epochs) and behavioural response (response/no response).