Nature and Science of Sleep

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

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

This article was published in the following Dove Press journal:
Nature and Science of Sleep

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

[1]College of Education, Psychology and Social Work, Flinders University, Adelaide, SA, 5001, Australia; [2]Adelaide Institute for Sleep Health: AFlinders Centre of Research Excellence, College of Medicine and Public Health, Flinders University, Adelaide, SA, 5001, Australia

Introduction: 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 testing of its accuracy against polysomnography (PSG) with an independent sample of good and poor sleepers (Study 2). The accuracy of THIM was compared to two popular wearables, Fitbit and Actiwatch devices.

Methods: Twenty-five (Study 1) and twenty (Study 2) healthy individuals with good or poor sleep (defined by scores on the Insomnia Severity Index) 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, THIM, Fitbit and Actiwatch devices.

Results: In both studies, THIM had lower sensitivity (M = 0.89–0.91) compared to the Actiwatch (M = 0.95) and Fitbit devices (M = 0.96–0.98), yet higher specificity (M = 0.59 vs M = 0.32–0.59) in detecting sleep. 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, $p > 0.06$. Yet, there was high variability in the accuracy of all three actigraphy devices between individuals (evident on Bland–Altman plots) that was unexplained by sleep quality.

Discussion: Together, these studies suggest that THIM is capable of monitoring sleep and wake overnight in good and poor sleepers to a similar degree of accuracy as two of the most popular actigraphy devices available. 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

Introduction

There are many uses for a sleep tracker that can accurately measure 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.[1,2] 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

Correspondence: Hannah Scott
College of Education, Psychology and Social Work, Flinders University, GPO Box 2100, Adelaide, SA, 5001, Australia
Tel + 61 8 8201 2767
Email hannah.scott@flinders.edu.au

submit your manuscript | www.dovepress.com
DovePress
http://doi.org/10.2147/NSS.S287048

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,[3] 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.[4] 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.[5] It is therefore unsurprising that these devices tend to overestimate sleep and underestimate wakefulness in most individuals.[6] This can be particularly true for individuals that typically spend considerable durations of time in bed awake but inactive, such as those with insomnia,[7,8] as noted in the American Academy of Sleep Medicine Clinical Practice Guidelines [see[9] for a comprehensive review]. 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.

Whilst the actigraphy method has 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.[10] 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 a similar finger-worn actigraphy device have found promising results for estimating sleep and wake.[34] Across 41 healthy adolescents/young adults, the OURA ring achieved a high sensitivity of 0.96 and a specificity of 0.48, which is at least comparable, if not better, performance than similar wrist-based devices for the adolescent population.[11] We therefore predict that THIM may be more accurate than wrist-based actigraphy devices.

This article described the development and accuracy of the THIM device for tracking sleep and wakefulness overnight compared to PSG. Two studies will be presented. Study 1 aimed to 1) develop the algorithm that THIM uses to track sleep and wakefulness, 2) test whether it performs differently 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. Importantly, the validation was conducted on an entirely independent sample from that used to develop the algorithm in Study 1.

Study 1: Method

Participants

Ethics approval was obtained from the Flinders University Social and Behavioural Research Ethics Committee, South Australia. This study was conducted in accordance with the Declaration of Helsinki. All participants provided written, informed consent.

Participants were recruited via print and online advertisements and completed a battery of screening questionnaires to assess their eligibility. The screening questionnaires comprised the Insomnia Severity Index [ISI],[12] and the Pittsburgh Sleep Quality Index [PSQI],[13] 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 and 00:00 and wake up time between 06:00 and 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. See Table 1 for participant characteristic information.

Study Design

This was a within-groups observational study. All participants slept overnight in the sleep laboratory with 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.[14]

THIM

THIM (firmware v 1.0.3) is a ring-like device worn on the middle phalanx of the index finger on 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,

Table 1 Descriptive Characteristics for Participants in Study 1

Characteristics	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)

Abbreviations: BMI, body mass index; ISI, Insomnia Severity Index; N, sample size; PSQI, Pittsburgh Sleep Quality Index; SD, standard deviation.

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 (SE) – 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 upload the 30-second epoch data to the cloud-based servers. At present, the epoch data is not readily accessible for download. The manufacturers of THIM, Re-Time Pty. Ltd., retrieved and forwarded the data to us for the purpose of this study. The sleep periods were determined from the lights out/lights on times recorded on the laboratory night. 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.[15] The devices were worn on the wrist of the non-dominant hand 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 were 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 worn on the wrist of participants' non-dominant hands when they got in/out of bed. The times in and out of bed were manually entered from the lights out/lights on times recorded on the laboratory night. After the laboratory night, the "normal" Fitbit algorithm setting was applied to score the data into 60-second epochs. The sleep/wake epoch data were 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.[16] 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.[17,18] 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. This statistical approach minimised the likelihood of Type I errors and account for missing data.

The limit of agreement between PSG and each actigraphy device was also analysed using Bland–Altman plots.[19,20] 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 (ie, epochs defined as sleep). SOL was calculated from the sum of epochs that exceeded the sensitivity threshold (ie, 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. SE 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.[14] LMM analyses examined whether there were any significant differences between actigraphy devices (the fixed effect) for estimating each sleep parameter (SOL, TST, WASO, and SE). 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 2. As shown, all three actigraphy devices had high sensitivity. A LMM indicated that the sensitivities differed between devices, $F(2, 68) = 21.16$, $p < 0.001$. Post hoc tests showed that THIM had significantly lower sensitivity than the Actiwatch, $p = 0.001$, and the Fitbit Flex devices, $p < 0.001$. 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, $p = 0.07$.

Specificities also differed between devices, $F(2, 68) = 12.11$, $p < 0.001$. Post hoc tests indicated that THIM had a significantly higher specificity than the Actiwatch Spectrum device, $p = 0.001$, and the Fitbit Flex devices, $p < 0.001$. 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, $p = 0.99$. There were no significant differences in accuracy between devices, $F(2, 68) = 0.49$, $p = 0.61$.

Sleep Parameter Estimations

Table 2 also 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, $F(3, 92) = 6.39$, $p = 0.001$. Post hoc comparisons indicated there were large significant differences between PSG and the Actiwatch device, $p < 0.001$, $d = 1.31$, and Fitbit Flex devices, $p = 0.04$, $d = 0.77$. There was no significant difference between PSG and THIM estimations of SOL, $p = 0.99$.

There were significant differences for estimations of TST, $F(3, 92) = 3.84$, $p = 0.01$. However, post hoc

Table 2 Epoch-by-Epoch and Sleep Parameter Descriptive Statistics for PSG and the Actigraphy Devices

Variables	Actigraphy Device			
	PSG	THIM	Actiwatch	Fitbit Flex
Epoch-by-epoch analyses				
Sensitivity (SD)	–	0.91 (0.05)	0.95 (0.04)*	0.98 (0.03)*
Specificity (SD)	–	0.59 (0.21)	0.35 (0.18)*	0.32 (0.22)*
Accuracy (SD)	–	0.85 (0.07)	0.85 (0.10)	0.87 (0.09)
Sleep parameters				
SOL, mean min (SD)	24.22 (19.98)	20.40 (21.46)	5.36 (3.77)*	11.64 (11.54)*
TST, mean min (SD)	400.96 (67.02)	394.46 (48.72)	423.10 (51.00)	440.60 (45.06)
WASO, mean min (SD)	48.06 (36.34)	58.38 (22.55)	38.00 (26.53)	19.60 (18.37)*
SE, mean % (SD)	84.58 (11.10)	83.50 (7.56)	90.73 (6.04)	93.52 (5.10)*

Note: *p < 0.05 from LMM analyses between PSG and actigraphy device.
Abbreviations: PSG, polysomnography; SD, standard deviation; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset.

comparisons indicated no significant differences between PSG and any of the actigraphy devices, $p > 0.06$.

There were significant differences for WASO, $F(3, 92) = 9.48$, $p < 0.001$. Post hoc tests indicated a large significant difference between PSG and the Fitbit Flex device, $p = 0.002$, $d = 0.99$, but no significant differences between PSG and THIM, $p = 0.99$, or PSG and the Actiwatch device, $p = 0.99$.

There were significant differences for SE, $F(3,92) = 9.27$, $p < 0.001$. Post hoc comparisons indicated large significant differences between PSG and the Fitbit Flex, $p = 0.004$, $d = 1.04$. There were no significant differences between PSG and THIM, $p = 0.99$, or the Actiwatch for estimations of SE, $p = 0.06$.

Bland–Altman Plots

Figure 1 presents Bland–Altman plots for each actigraphy device on key sleep parameters. Significant proportional bias was observed for the Actiwatch and Fitbit Flex devices, such that increasing sleep onset latency resulted in greater underestimations. 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, with significant proportional bias potentially due to the presence of an outlier. 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 and significant proportional bias, again potentially due to the presence of an outlier. The 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 SE compared to PSG.

Good and Poor Sleeper Comparison

Table 3 contains the descriptive statistics for these secondary analyses. For sensitivity, there was a statistically significant interaction between device and the type of sleeper, $F(2,43.81) = 8.66$, $p = 0.001$. Post hoc analyses revealed that sensitivity was significantly higher for good sleepers compared to poor sleepers for the THIM device, $p < 0.001$. This was a large effect, $d = 1.50$. There were no significant differences between good and poor sleepers for the sensitivity of the Actiwatch, $p = 0.07$, or the Fitbit Flex devices, $p = 0.93$. A LMM examining the interaction between device and sleeper type on specificity was not significant, $p = 0.77$. There was a statistically significant interaction between device and the type of sleeper on accuracy, $F(2,42.55) = 6.44$, $p = 0.004$. However, post hoc comparisons between groups within devices were not significant, $p > 0.12$.

LMM analyses determined whether there were any significant differences between good and poor sleepers

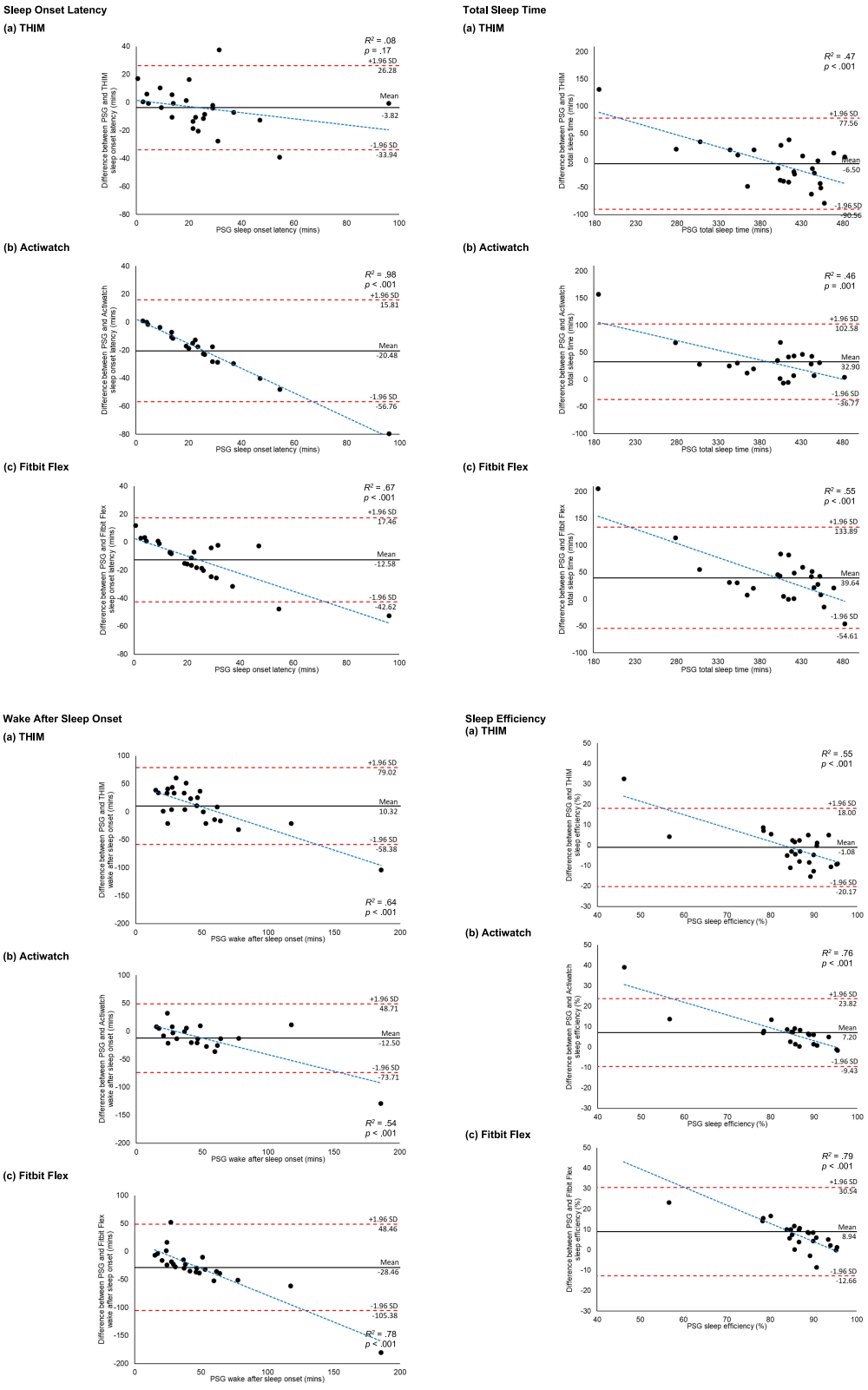

Figure 1 Bland–Altman plots showing the agreement between PSG and (**A**) THIM, (**B**) Actiwatch and (**C**) Fitbit Flex devices separately on sleep onset latency, total sleep time, wake after sleep onset and sleep efficiency. The solid black horizontal line indicates perfect agreement with PSG. The solid coloured horizontal line indicates the mean bias for the device, the dashed coloured horizontal lines indicates the upper and lower limits of agreement, and the dotted coloured lines are the lines of best fit.

Table 3 Epoch-by-Epoch and Sleep Parameter Descriptive Statistics for Each Actigraphy Device Comparing Good and Poor Sleepers

Variables	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 min (SD)			
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 min (SD)			
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 min (SD)			
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)
SE 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)

Note: *p < 0.05 between good and poor sleepers with this device.

on the mean discrepancies of the actigraphy devices for each sleep parameter. The descriptive statistics for these analyses are presented in Table 3. These LMM analyses found no significant interactions between device and the type of sleeper on SOL, $p = 0.66$, TST, $p = 0.06$, WASO, $p = 0.08$, or SE, $p = 0.08$.

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 SE 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,[6,21] although some studies have found no significant differences between good sleepers and those with insomnia.[22,23] 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 4.

Study 2: 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 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, $F(2, 54) = 14.52$, $p < 0.001$. Post hoc tests showed that THIM had a significantly lower sensitivity than the Actiwatch-2, $p < 0.001$, $d = 1.18$, and Fitbit Alta devices, $p < 0.001$, $d = 1.37$. There was no significant difference between the Actiwatch-2 and Fitbit Alta devices, $p = 0.99$.

Furthermore, a LMM indicated significant differences in the specificities between devices, $F(2, 54) = 7.72$, $p = 0.001$. Post hoc tests indicated that both THIM and the

Table 4 Descriptive Statistics for Participant Characteristics Collected at Screening in Study 2

Characteristics	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 hrs (SD)	8.06 (0.88)	7.09 (1.50)	7.53 (1.32)

Abbreviations: BMI, body mass index; ISI, Insomnia Severity Index; N, sample size; PSQI, Pittsburgh Sleep Quality Index; SD, standard deviation.

Table 5 Epoch-by-Epoch and Sleep Parameter Descriptive Statistics for PSG and the Actigraphy Devices from Study 2

Variables	Actigraphy Device			
	PSG	THIM	Actiwatch-2	Fitbit Alta
Epoch-by-epoch analyses				
Sensitivity (SD)	–	0.89 (0.06)	0.95 (0.04)*	0.96 (0.04)*
Specificity (SD)	–	0.59 (0.18)	0.59 (0.20)	0.39 (0.18)*
Accuracy (SD)	–	0.85 (0.06)	0.91 (0.05)*	0.89 (0.07)
Sleep parameters				
SOL, mean min (SD)	21.68 (16.65)	19.15 (17.82)	4.53 (4.25)*	14.03 (19.22)
TST, mean min (SD)	424.60 (48.73)	403.25 (46.15)	434.35 (46.66)	438.05 (43.54)
WASO, mean min (SD)	40.68 (34.50)	64.55 (30.36)	45.03 (24.84)	22.28 (21.88)
SE, mean % (SD)	87.36 (8.62)	82.93 (7.82)	89.88 (4.86)	92.33 (4.56)

Note: *p < 0.05 from LMM analyses between PSG and actigraphy device.
Abbreviations: PSG, polysomnography; SD, standard deviation; SOL, sleep onset latency; TST, total sleep time; WASO, wake after sleep onset.

Actiwatch-2 had significantly higher specificities than the Fitbit Alta, p < 0.005, d = 1.11 and 1.05, respectively. There was no difference between THIM and the Actiwatch-2, $p = 0.99$.

A significant main effect of device was found on accuracy, $F(2, 54) = 5.14$, $p = 0.009$. Post hoc tests indicated THIM had significantly lower accuracy than the Actiwatch-2, $p = 0.01$, but was not significantly different compared to the Fitbit Alta, $p = 0.09$. There was no significant difference between the Actiwatch-2 and Fitbit Alta devices, $p = 0.99$.

Sleep Parameter Estimations

Table 5 also presents the descriptive statistics for the sleep parameter estimations. A LMM determined significant differences on estimations of SOL, $F(3, 73) = 4.07$, $p = 0.01$. Post hoc comparisons indicated a large significant difference between PSG and the Actiwatch-2, $p = 0.01$, $d = 1.41$, and no significant differences between PSG and THIM, $p = 0.99$, or the Fitbit Alta, $p = 0.81$. There were no significant differences on estimations of TST, $F(3, 73) = 2.23$, $p = 0.75$. There were significant differences on WASO, $F(3, 73) = 7.44$, $p < 0.001$. However, post hoc comparisons indicated that these significant differences were not between PSG and any of the actigraphy devices, $p > 0.06$. Similarly, there were significant differences on SE, $F(3, 73) = 6.95$, $p < 0.001$, but post hoc comparisons indicated that there were no significant differences between PSG and any actigraphy device, $p > 0.14$.

Bland–Altman Plots

Figure 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. Significant proportional bias was observed for the Actiwatch and Fitbit Flex devices, such that increasing sleep onset latency resulted in greater underestimations. The TST plots indicate that THIM appears to have a mean bias towards underestimating TST compared to the Actiwatch-2 device with a mean bias close to zero and the Fitbit Alta device with 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 WASO plots illustrate lines of best fit with steep negative slopes (ie, significant proportional bias), though this is potentially driven by the presence of an outlier. The SE plots further illustrate that THIM had a small bias towards underestimating sleep efficiency, while the other devices overestimated sleep efficiency. Nonetheless, the findings of the LMM analyses above indicate that these biases do not produce estimation of SE significantly different to PSG.

Good and Poor Sleeper Comparison

Table 6 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

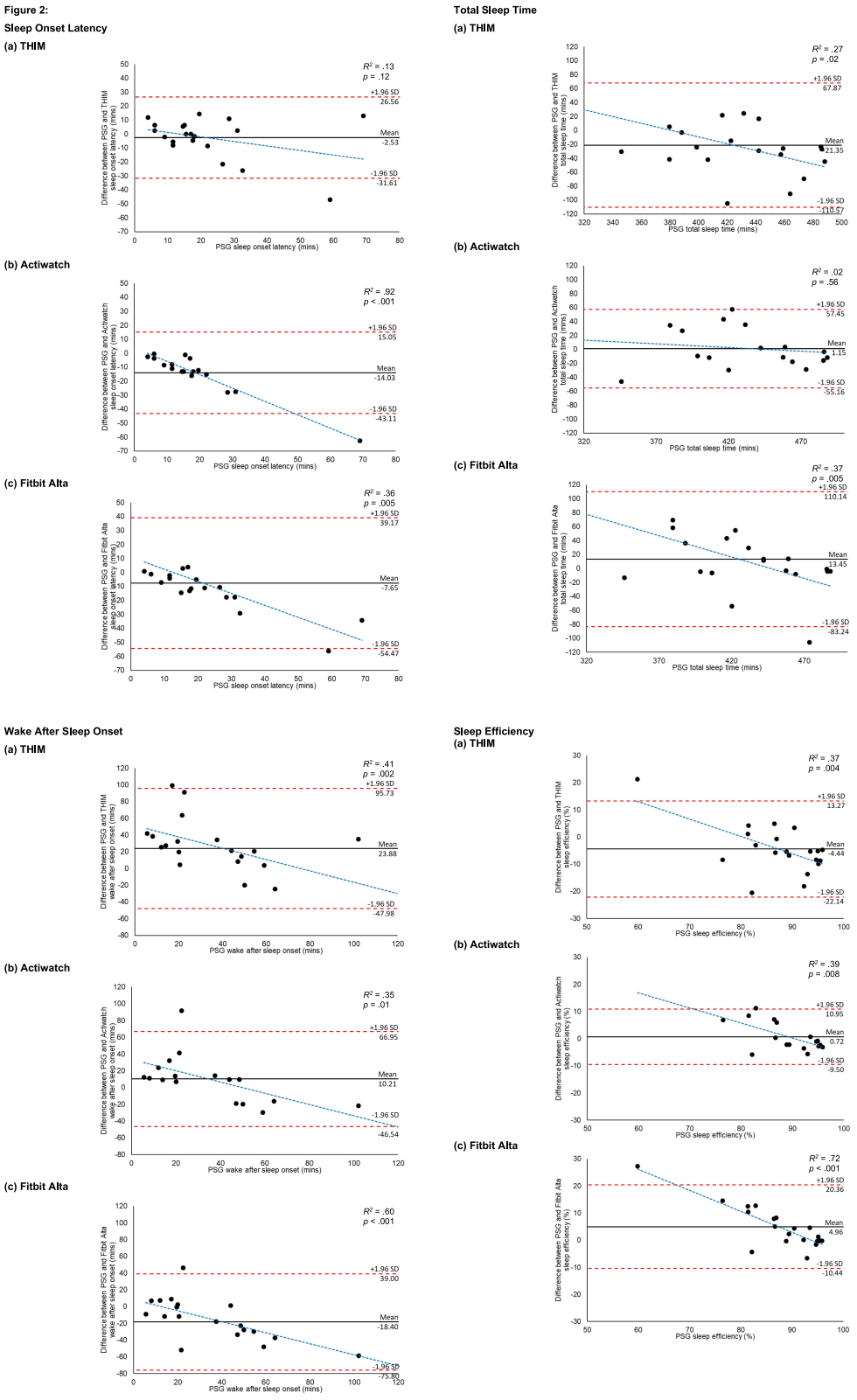

Figure 2 Bland–Altman plots showing the agreement between PSG and (**A**) THIM, (**B**) Actiwatch-2 and (**C**) Fitbit Alta devices separately on sleep onset latency (SOL), total sleep time (TST), wake after sleep onset (WASO), and sleep efficiency (SE). The solid black horizontal line indicates perfect agreement with PSG. The solid coloured horizontal line indicates the mean bias for the device, the dashed coloured horizontal lines indicates the upper and lower limits of agreement, and the dotted coloured lines are the lines of best fit.

Table 6 Epoch-by-Epoch and Sleep Parameter Descriptive Statistics Comparing Good and Poor Sleepers from Study 2

Variables	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)
SE 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)

Note: There were no significant differences between good and poor sleepers on any device.

interactions between the actigraphy devices and sleeper type were not statistically significant for sensitivity, $p = 0.78$, specificity, $p = 0.43$, or accuracy, $p = 0.37$.

The descriptive statistics for LMM analyses comparing mean discrepancies between PSG and each actigraphy device on the sleep parameters are also presented in Table 6. There were no significant interactions between device and sleeper type on SOL, $p = 0.55$, TST, $p = 0.75$, WASO, $p = 0.47$, or SE, $p = 0.95$.

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 sleep epochs as wake 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 high variability in the discrepancy between PSG and THIM shown across all sleep parameters (evident by the wide levels of agreement). 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 was slightly lower in accuracy than the Acitwatch-2 device. The Bland–Altman plots also indicated that THIM had a bias towards underestimating TST and SE, and overestimating WASO compared to the other two actigraphy devices that overestimated sleep. Notably, all three devices performed better on the sleep parameter

estimations than on the epoch-by-epoch statistics, demonstrating the need to assess device performance using agreement rather than correspondence statistics. These findings contrast previous research with Fitbit and Actiwatch devices that demonstrate a bias towards overestimating sleep and underestimating wake.[24–27] Further, differences in the findings from the current study compared to previous research may be due to differences between algorithms or device placement, or both, between THIM and other actigraphy devices. For instance, the presence of finger twitches during REM sleep may explain the trend towards overestimating WASO, as period of REM may have been miss-scored as wake.[10] Further research could investigate whether the presence of finger twitches during REM sleep correspond with wake periods identified by a finger-worn actigraphy device.

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,[21,28] particularly when assessed in people with insomnia.[22] Such differences between studies may be due to the algorithm used to score the actigraphy data, as noted in reviews of this area.[25] Further research is warranted to assess not only device accuracy but also algorithm accuracy across sleep disordered populations compared to good sleepers.

Nonetheless, as found in Study 1, there is still considerably high variability in 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[17] this degree of variability is presumably not acceptable to appropriately substitute for PSG, particularly when interpreting the data at the individual level.[29] 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.[17,29] 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.

Regarding the limitations of the current study, the PSG data were scored by one qualified sleep technician. The interrater reliability of PSG sleep scoring amongst qualified individuals can be low,[35] increasing the error of measurement of our gold standard measure. Additionally, the sample size was low compared to other validation studies. These limitations should be considered when interpreting differences between devices and when 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.[30] Additionally, considering that people with insomnia experience different sleep quality in the sleep laboratory compared to their home environments,[31,32] 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. The validity of THIM for estimating sleep and wake in individuals with insomnia is the next priority to further the long-term goal of improving the treatment of insomnia.

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.

Acknowledgments

The authors would like to acknowledge the contributions of study participants, Flinders University third-year Psychology placement students, and Adelaide Institute for Sleep Health staff that assisted with data collection.

Disclosure

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 Pty Ltd for this project. Hannah Scott reports grants from Re-Time Pty Ltd during the conduct of the study and has a patent pending: PCT/US2017/23240. Leon Lack reports grants from Re-Time Pty Ltd during the conduct of the study; is a shareholder of Re-Time Pty Ltd; and receives royalties for downloads of their book from the Re-Time website, outside the submitted work; in addition, Leon Lack has a US patent application pending to Re-Time Pty Ltd: 16/496360. The authors report no other potential conflicts of interest for this work.

References

1. Kuula L, Gradisar M, Martinmäki K, et al. Using big data to explore worldwide trends in objective sleep in the transition to adulthood. *Sleep Med*. 2019;62:69–76. doi:10.1016/j.sleep.2019.07.024
2. Ong JL, Tandi J, Patanaik A, Lo JC, Chee MWL. Large-scale data from wearables reveal regional disparities in sleep patterns that persist across age and sex. *Sci Rep*. 2019;9(1):3415. doi:10.1038/s41598-019-40156-x
3. Re-Time PL. *Thim - the First Wearable Device for Sleep, Which Will Improve Your Sleep*. 2016. Available from http://thim.io/. Accessed December 21, 2020.
4. Sadeh A, Acebo C. The role of actigraphy in sleep medicine. *Sleep Med Rev*. 2002;6(2):113–124. doi:10.1053/smrv.2001.0182
5. Pollak CP, Tryon WW, Nagaraja H, Dzwonczyk R. How accurately does wrist actigraphy identify the states of sleep and wakefulness? *Sleep*. 2001;24(8):957–965. doi:10.1093/sleep/24.8.957
6. Van den Water ATM, Holmes A, Hurley DA. Objective measurements of sleep for non-laboratory settings as alternative to polysomnography - a systematic review. *J Sleep Res*. 2011;20:183–200. doi:10.1111/j.1365-2869.2009.00814.x
7. Silvertsen B, Omvik S, Havik OE, et al. A comparison of actigraphy and polysomnography in older adults treated for chronic primary insomnia. *Sleep*. 2006;29(10):1353–1358. doi:10.1093/sleep/29.10.1353
8. Vallières A, Morin CM. Actigraphy in the assessment of insomnia. *Sleep*. 2003;26(7):902–906. doi:10.1093/sleep/26.7.902
9. Smith MT, McCrae CS, Cheung J, et al. 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. *J Clin Sleep Med*. 2018;14(7):1209–1230. doi:10.5664/jcsm.7228
10. Reiter AM, Roach GD, Sargent C, Lack L. Finger twitches are more frequent in REM sleep than in Non-REM sleep. *Nat Sci Sleep*. 2020;12:49–56. doi:10.2147/NSS.S233439
11. Meltzer LJ, Hiruma LS, Avis K, Montgomery-Downs H, Valentin J. Comparison of a commercial accelerometer with polysomnography and actigraphy in children and adolescents. *Sleep*. 2015;38(8):1323–1330. doi:10.5665/sleep.4918
12. Morin CM, Belleville G, Bélanger L, Ivers H. The Insomnia Severity Index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep*. 2011;34(5):601–608. doi:10.1093/sleep/34.5.601
13. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28:193–213. doi:10.1016/0165-1781(89)90047-4
14. American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events (Version 2.5)*. American Academy of Sleep Medicine; 2018.
15. Respironics P. *Equivalence of Activity Recordings and Derived Sleep Statistics: Actiwatch-64, Actiwatch 2 and Actiwatch Spectrum*. 2009.
16. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*. 2012;35(2):287–302. doi:10.5665/sleep.1642
17. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. *Med Sci Sports Exerc*. 2019;51(7):1538–1557. doi:10.1249/MSS.0000000000001947
18. Depner CM, Cheng PC, Devine JK, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. *Sleep*. 2019. doi:10.1093/sleep/zsz254
19. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986;327(8476):307–310. doi:10.1016/S0140-6736(86)90837-8
20. Giavarina D. Understanding Bland Altman analysis. *Biochem Med (Zagreb)*. 2015;25(2):141–151. doi:10.11613/BM.2015.015
21. Hedner J, Pillar G, Pittman SD, Zou D, Grote L, White DP. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. *Sleep*. 2004;27(8):1560–1566. doi:10.1093/sleep/27.8.1560
22. Kang SG, Kang JM, Ko KP, Park SC, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. *J Psychosom Res*. 2017;97:38–44. doi:10.1016/j.jpsychores.2017.03.009
23. Sanchez-Ortuno MM, Edinger JD, Means MK, Almirall D. Home is where sleep is: an ecological approach to test the validity of actigraphy for the assessment of insomnia. *J Clin Sleep Med*. 2010;6(1):21–29. doi:10.5664/jcsm.27706
24. Bianchi MT. Sleep Devices: wearables and nearables, informational and interventional, consumer and clinical. *Metabolism*. 2017. doi:10.1016/j.metabol.2017.10.008
25. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. *Int J Behav Nutr Phys Act*. 2015;12:159. doi:10.1186/s12966-015-0314-1
26. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of wristband fitbit models in assessing sleep: systematic review and meta-analysis. *J Med Internet Res*. 2019a;21(11):e16273. doi:10.2196/16273
27. Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Performance assessment of new-generation Fitbit technology in deriving sleep parameters and stages. *Chronobiol Int*. 2019b;37(1):47–59. doi:10.1080/07420528.2019.1682006
28. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. *J Affect Disord*. 2017;217:299–305. doi:10.1016/j.jad.2017.04.030

29. Danzig R, Wang M, Shah A, Trotti LM. 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. *J Sleep Res*. 2019;29(1):e12926. doi:10.1111/jsr.12926
30. Buysse DJ, Cheng Y, Germain A, et al. Night-to-night sleep variability in older adults with and without chronic insomnia. *Sleep Med*. 2010;11(1):56–64. doi:10.1016/j.sleep.2009.02.010
31. Edinger JD, Fins AI, Sullivan R Jr, et al. Sleep in the laboratory and sleep at home: comparisons of older insomniacs and normal sleepers. *Sleep*. 1997;20(12):1119–1126. doi:10.1093/sleep/20.12.1119
32. Edinger JD, Glenn DM, Bastian LA, et al. Sleep in the laboratory and sleep at home II: comparisons of middle-aged insomnia sufferers and normal sleepers. *Sleep*. 2001;24(7):761–770. doi:10.1093/sleep/24.7.761
33. Withrow D, Roth T, Koshorek G, Roehrs T. Relation between ambulatory actigraphy and laboratory polysomnography in insomnia practice and research. *J Sleep Res*. 2019;28(4):e12854. doi:10.1111/jsr.12854
34. de Zambotti M, Rosas L, Colrain IM, Baker FC. The Sleep of the Ring: Comparison of the ŌURA Sleep Tracker Against Polysomnography. *Behav Sleep Med*. 2019;17(2):124–136. doi:10.1080/15402002.2017.1300587
35. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. *J Clin Sleep Med*. 2013 Jan 15; 9(1):81–7. doi:10.5664/jcsm.2350