

SCIENTIFIC INVESTIGATIONS

The accuracy of the THIM wearable device for estimating sleep onset latency

Hannah Scott, PhD^{1,2}; Ashwin Whitelaw²; Alex Canty¹; Nicole Lovato, PhD²; Leon Lack, PhD^{1,2}

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

Study Objectives: THIM is a wearable device designed to accurately estimate sleep onset. This article presents 2 studies that tested the original (study 1) and a refined (study 2) THIM algorithm against polysomnography (PSG) for estimating sleep onset latency.

Methods: Twelve (study 1) and 20 (study 2) individuals slept in the laboratory on 2 nights where participants underwent THIM-administered sleep onset trials with simultaneous PSG recording. Participants attempted to fall asleep while using THIM, which woke them once it determined sleep onset.

Results: In study 1, there was no significant difference between PSG (mean = 1.94 minutes, SD = 1.32) and THIM sleep onset latency (mean = 2.05 minutes, SD = 1.38) on the first or second night ($P > .07$). There were moderate correlations between PSG and THIM on both nights [$r_{(s)} > .57, P < .001$]. In 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 (mean = 3.41 minutes, SD = 2.21) and THIM sleep onset latency (mean = 3.65 minutes, SD = 2.18) ($P = .25$). There was strong correspondence between the two devices [$r_{(s)} > .73, P < .001$], narrow levels of agreement on Bland-Altman plots, and significantly fewer trials where PSG sleep onset had not occurred (10.24%), $P = .04$.

Conclusions: THIM showed a high degree of correspondence and agreement with PSG for estimating sleep onset latency. Future research will investigate whether THIM is accurate with an insomnia sample for clinical purposes.

Keywords: sleep onset latency, intensive sleep retraining, wearable device, consumer sleep technology, polysomnography, actigraphy

Citation: Scott H, Whitelaw A, Canty A, Lovato N, Lack L. The accuracy of the THIM wearable device for estimating sleep onset latency. *J Clin Sleep Med*. 2021; 17(5):973–981.

BRIEF SUMMARY

Current Knowledge/Study Rationale: Monitoring the onset of sleep outside of the laboratory setting is required for many purposes, yet there are few simple objective methods available. Here, we discuss the accuracy of a new wearable device called THIM.

Study Impact: The revised version of the THIM algorithm showed high agreement with the gold-standard measure of sleep, polysomnography, on a number of indices. Further research is required to examine the accuracy of THIM with individuals with insomnia to inform its clinical utility for administering a brief (24-hour) but effective behavioral treatment for insomnia, once restricted to the sleep laboratory, in the home environment.

INTRODUCTION

Accurate assessment of sleep onset latency (SOL) is required for a variety of research and clinical purposes. For instance, Intensive Sleep Retraining is a behavioral treatment for chronic insomnia that involves repeatedly falling asleep and waking up shortly thereafter over the course of one overnight session.^{1,2} Additionally, brief daytime sleep episodes such as power naps or sleep diagnostic tests such as the Multiple Sleep Latency Test involve achieving a precise amount of sleep.^{3,4} These 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.⁵ 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 developed by Re-Time Pty Ltd Adelaide, South Australia, Australia, that is worn like a ring.⁶ 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 prespecified duration of sleep. THIM was designed to administer Intensive Sleep Retraining (ISR) and may be capable of administering power naps and daytime diagnostic tests (eg, the Multiple Sleep Latency Test) outside of the laboratory setting, without the need for expensive equipment or trained individuals to set up, administer, or score the data. However, the accuracy of THIM for estimating sleep onset is currently unknown and must be tested to ensure that the device 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 behavioral responses to external stimuli.^{7,8} Hence, this behavioral 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.^{9,10}

While similar devices using the stimulus-response method are accurate for estimating SOL,^{11,12} 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,¹³ whereas vibratory stimuli emitted from THIM are perceived through the somatosensory system.^{14,15} Whether these pathways show similar inhibition across the sleep onset period is currently unknown. MacLean and colleagues¹⁶ tested the discrepancy between PSG sleep onset and behavioral 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 5 SDs above the participant's waking threshold. Therefore, responsiveness to minimal intensity tactile stimuli—as utilized by THIM—during the sleep onset period is yet to be tested.

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.

The current article summarizes 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 in 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 in a larger independent sample. 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 in sleep patterns found in the general population.

STUDY 1: METHODS

Participants

Ethical approval was obtained from the Flinders University Social and Behavioral Research Ethics Committee, South

Australia. Potential participants were recruited via advertisements on community noticeboards and social media. Eligibility criteria were as follows: self-reported average habitual bedtime between 22:00 and 00:00 and wake-up time between 06:00 and 08:00; fluent in English; no diagnosis of a physical or mental health condition; no active nicotine or illicit substance use or alcohol (>10 standard drinks/week) or caffeine (>250 mg/day) dependence; no consumption of medications known to interfere with sleep; no overnight shift work or trans-meridian travel within the last 2 months; and not pregnant or lactating. Screening questionnaires comprised the Insomnia Severity Index (ISI)¹⁷ and the Pittsburgh Sleep Quality Index¹⁸ 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 the eligibility criteria, but one participant withdrew after participating in night 1. The final sample comprised 12 individuals (see **Table 1** for participant characteristics information). Scores on the ISI indicated that 5 participants had subthreshold levels of insomnia and were categorized as poor sleepers (ISI score ≥ 7), and 7 were good sleepers (ISI score <7).

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, chin electromyography, and electrocardiography sites were sampled at 256 Hz. PSG data were scored using Profusion Compumedics software (version 4; Charlotte, NC) by a qualified, independent sleep technician. In accordance with American Academy of Sleep Medicine scoring criteria,¹⁹ 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 (most commonly, the beginning of N1 sleep).

THIM

THIM (firmware version 1.0.3) is a small, ring-like device worn on the index finger of the dominant hand. THIM comes with 4 different-sized ring bands so that the device can fit securely onto fingers of almost all sizes. To set up THIM, the device was connected via Bluetooth to the accompanying smartphone application (version 1.0.1) 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 1**). During the trials, the device emitted low-intensity, short-duration vibratory stimuli at nonregular intervals (averaging 30 seconds apart). The intensity of the vibrations was individually calibrated to the minimum level that the participant could consistently respond to while 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

Table 1—Descriptive characteristics for participants in studies 1 and 2.

	Study 1 (n = 12)		Study 2 (n = 20)		Study Comparison
Characteristics					
Age, mean (SD), y	24.9 (6.1)		23.6 (4.9)		$t(30) = 0.68, P = .50$
Sex, n (%)					
Men	3 (25)		7 (35)		$\chi(1) = 1.66, P = .20$
Women	9 (75)		13 (65)		
Weekly alcohol consumption, n servings (SD)	0.75 (0.97)		1.60 (1.79)		$t(29.80) = -1.51, P = .14$
Daily caffeine consumption, n servings (SD)	1.29 (1.05)		1.89 (1.47)		$t(30) = -1.20, P = .24$
	Good Sleeper (n = 7)	Poor Sleeper (n = 5)	Good Sleeper (n = 10)	Poor Sleeper (n = 10)	
Sleep characteristics					
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-up 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), h	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, PSQI = Pittsburgh Sleep Quality Index, SD = standard deviation, TST = total sleep time.

detected by the device’s accelerometer. If participants failed to respond to 2 consecutive vibratory stimuli, the device inferred that sleep onset had occurred and it emitted a high-intensity alarm vibration to wake them up, signaling 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 2 consecutively missed vibratory stimuli.

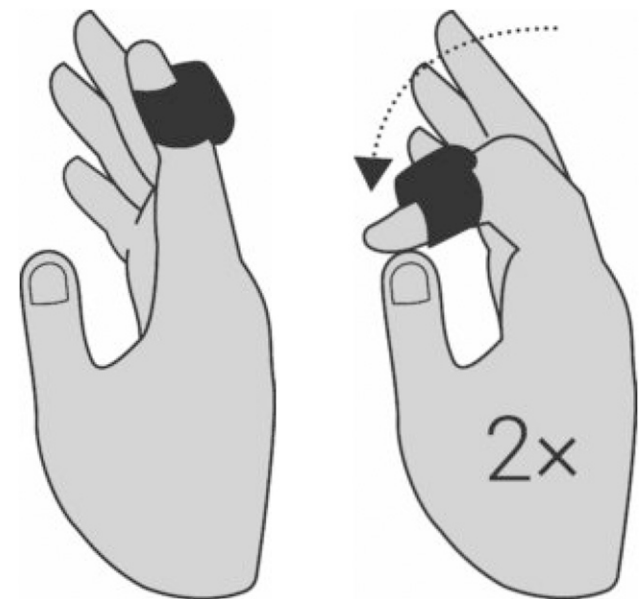
To monitor THIM, we mounted a small piezo-electric sensor to the side of the THIM device using adhesive tape. This sensor was inputted into a channel on the PSG device. From this sensor, we observed 4 events of interest: vibrations emitted from THIM, finger taps as responses to the vibrations, as well as the beginning (the double-tap motion) and end (the high-intensity alarm vibration) of each trial. These 4 events were scored manually on the Profusion Compumedics software by 2 scorers (H.S. and A.W.) and used to determine sleep onset based on the rules of the proprietary THIM algorithm. 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 2 scorers.

Procedure

Home testing

Participants completed a sleep diary based on the Consensus Sleep Diary²⁰ and wore an actigraphy device (Actiwatch-2; Philips Respironics Murrysville, Pennsylvania, United States) every day for 1 week to monitor their sleep pattern prior to the first laboratory night. Participants’ average bedtimes

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



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 while monitored by PSG and THIM. They were woken 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 set up for overnight PSG recording. The THIM device was placed on the participants' 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 the supplemental material for this procedure).

THIM-administered sleep onset trials began 1 hour prior to a participant's bedtime and were maintained continuously for 4 hours in total (3 hours past 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 for 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 were compared separately for nights 2 and 3. Cohen's d was calculated as the mean difference in PSG and THIM estimations of SOL divided by the pooled SD. 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 with PSG. Paired-samples t tests were then conducted to test whether THIM significantly underestimated or overestimated SOL compared with PSG, separately for both laboratory nights. Additionally, the degree of correspondence between PSG and THIM was calculated across all sleep onset trials using Spearman's rank correlation coefficients separately for nights 2 and 3.

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-SOL (x axis) across all trials on each night.²¹ This involved calculating the mean difference (bias) and the limits of agreement (± 1.96 SD of the mean difference) between these measures. Upper and lower limits of

agreement within ± 5 minutes of PSG were considered acceptable, as previously defined as an acceptable criterion for the administration of ISR with a wearable device.⁵ The R^2 value for the linear regression line and coefficient P value are reported in the Bland-Altman plot figures, as an indicator of the degree of proportional bias.²¹ 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 vs night 3. Additionally, paired-samples t tests were conducted to compare differences in the discrepancy between PSG and THIM-SOL on night 2 vs night 3 for each trial (eg, 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 (mean = 1.94, SD = 1.32 minutes) and mean THIM-SOL [mean = 2.05, SD = 1.38 minutes; $t(11) = -0.88, P = .40, d = .08$]. The mean discrepancy between PSG and THIM-SOL on this night was low (mean = 0.08, SD = 0.49 minutes). There was also a significant moderate correlation between PSG and THIM-SOL across all sleep onset trials [$r_{(s)} = .67, P < .001$].

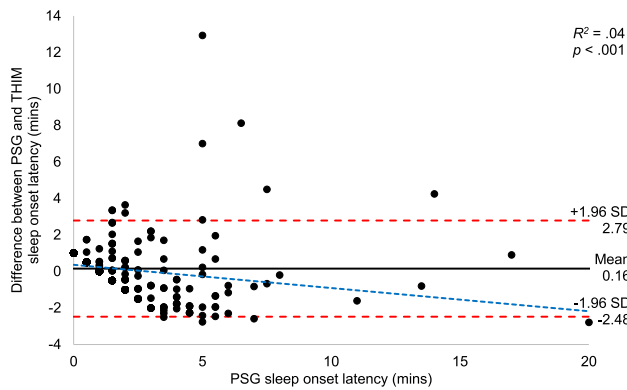
The level of agreement between PSG and THIM-SOL on night 2 is illustrated in [Figure 2](#). As shown by the narrow levels of agreement, there is little variability in the discrepancy between PSG and THIM-SOL across the 411 trials. Furthermore, the discrepancy between PSG and THIM is consistent across 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+ minutes 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.

Second sleep onset trial night

There was no significant difference between mean PSG-SOL (mean = 1.40, SD = 0.64 minutes) and mean THIM-SOL (mean = 2.12, SD = 1.71 minutes) on laboratory night 3 [$t(11) = -2.02, P = .07$]. Despite a medium effect size ($d = .56$), the mean discrepancy between PSG and THIM-SOL on this night was still relatively low (mean = 0.57, SD = 1.10 minutes). Additionally, there was a significant moderate correlation between PSG and THIM-SOL across all sleep onset trials [$r_{(s)} = .57, P < .001$].

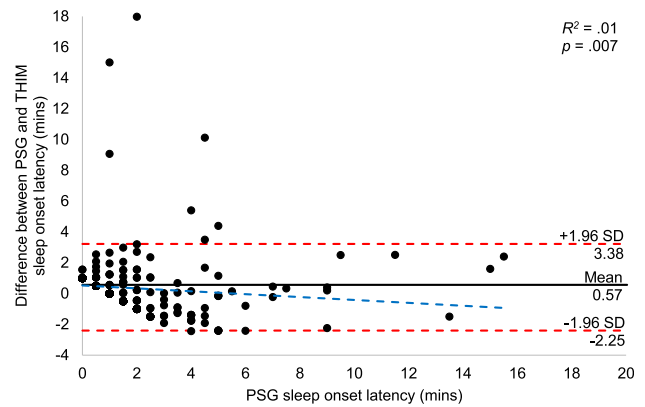
[Figure 3](#) is a Bland-Altman plot illustrating the level of agreement between PSG and THIM-SOL across all night 3

Figure 2—Bland-Altman plot indicating agreement between PSG and THIM-SOL on night 2 for study 1 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. The R^2 value and P value represent the linear regression line as indicators of the degree of proportional bias. Some datapoints represent many overlapping values. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency.

Figure 3—Bland-Altman plot indicating agreement between PSG and THIM-SOL on night 3 for study 1 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. The R^2 value and P value represent the linear regression line as indicators of the degree of proportional bias. Some datapoints represent many overlapping values. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency.

trials. Similar to **Figure 2**, the variability in the discrepancy between PSG and THIM-SOL across 527 trials is low. **Figure 3** also shows trials where participants were responding to THIM’s vibratory stimuli while fluctuating between wake and N1 sleep (points above the upper limit of agreement).

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 ($d = .57$). Paired-samples t tests revealed no significant differences in the discrepancy between PSG and THIM on night 2 vs night 3 for any trial (eg, on the first, second, third trial, etc) ($P > .10$). The accuracy of THIM compared with 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 (mean = 0.06, SD = 0.44 minutes) compared with poor sleepers (mean = 0.09, SD = 0.60 minutes) [$t(10) = -0.11, P = .92, d = .08$]. Similarly, there was no significant difference in the mean discrepancy on night 3 between good sleepers (mean = 0.34, SD = 0.21 minutes) and poor sleepers (mean = 0.88, SD = 1.75 minutes) [$t(4.08) = -0.68, P = .53$], although there was a medium effect size ($d = .48$). 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 **Figure 2** and **Figure 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, which 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) trials per participant on night 2 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 in the number of false-positive trials [$t(11) = 1.47, P = .17, d = .49$].

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 2 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 1 or both of the last 2 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 moderate agreement between THIM and PSG, regardless of sleeper type (good or poor sleeper status) and repeated use (night 2 vs night 3). Having said this, THIM had estimated sleep onset and prematurely ended the trial before PSG sleep onset criteria were met in a considerable number of trials. This is an issue for 2 reasons. First, 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. Second, 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.

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: METHODS

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 version 1.0.4) with a larger, independent sample.

Participants

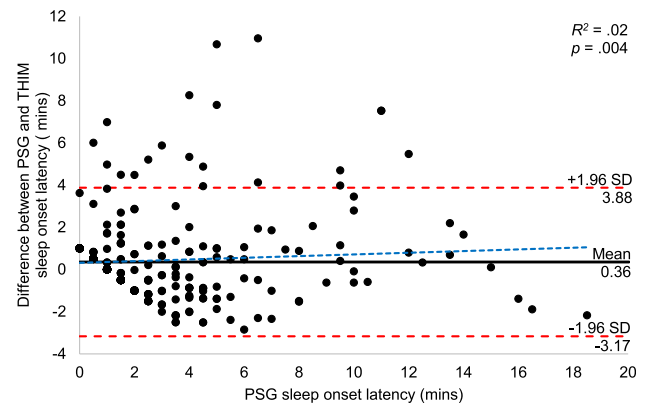
Participants of the second study were required to meet the same eligibility criteria as participants in the first study. Twenty healthy individuals met eligibility criteria and consented to participate. The ISI scores at screening indicated that 10 participants had subthreshold levels of insomnia and were categorized as poor sleepers (ISI score ≥ 7) and 10 were good sleepers (ISI score < 7). See [Table 1](#) for participant characteristics information and a comparison between the study 1 and study 2 samples. There were no significant differences in the participant characteristics between the 2 samples.

STUDY 2: RESULTS

First sleep onset trial night

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

Figure 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. The R^2 value and P value represent the linear regression line as indicators of the degree of proportional bias. Some datapoints represent many overlapping values. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency.

between PSG (mean = 3.41, SD = 2.21 minutes) and THIM estimations of mean SOL (mean = 3.65, SD = 2.18 minutes) on laboratory night 2 [$t(18) = -1.18, P = .25, d = .11$]. There was a small mean discrepancy between the 2 measures (mean = 0.24, SD = 0.90 minutes). There was also a significant strong correlation between PSG and THIM-SOL across all sleep onset trials [$r_{(s)} = .77, P < .001$]. As shown in [Figure 4](#), there was strong agreement between PSG and THIM-SOL across 535 trials.

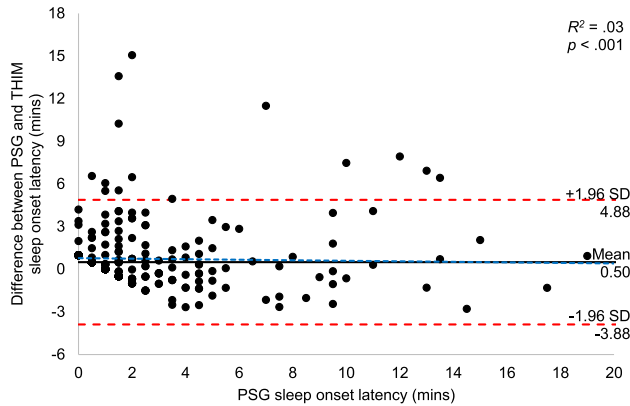
Second sleep onset trial night

Unlike night 2, on night 3 there was a significant difference between PSG (mean = 3.93, SD = 3.32 minutes) and THIM-SOL (mean = 4.75, SD = 3.85 minutes). THIM significantly overestimated SOL compared with PSG [$t(19) = -2.78, P = .01, d = .23$]. However, the effect size and mean discrepancy between PSG and THIM were still low (mean = 0.82, SD = 1.31 minutes). Additionally, there was a significant strong correlation between PSG and THIM-SOL across all sleep onset trials [$r_{(s)} = .73, P < .001$]. [Figure 5](#) shows continued strong agreement between PSG and THIM across 578 trials, as evidenced by the narrow levels of agreement.

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 independent-samples t tests to determine whether the issue occurred in fewer trials with the revised THIM algorithm compared with the original algorithm. There was a significantly lower number of false-positive trials with the revised algorithm compared with 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

Figure 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. The R^2 value and P value represent the linear regression line as indicators of the degree of proportional bias. Some datapoints represent many overlapping values. PSG = polysomnography, SD = standard deviation, SOL = sleep onset latency.

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 improved this issue without substantially increasing the mean discrepancy between THIM and PSG, although the issue was not entirely resolved.

We also conducted independent-samples t tests to determine whether the revised THIM algorithm (study 2) had a lower mean discrepancy for estimating SOL than the original algorithm (study 1) on night 2 and night 3. There were no significant differences in the mean discrepancy between the original algorithm or the revised algorithm on night 2 [$t(30) = -1.73$, $P = .10$] or night 3 [$t(30) = -0.68$, $P = .50$].

Learning effects

As in study 1, there was no significant difference between the mean discrepancy of PSG and THIM-SOL on night 2 compared with night 3 [$t(18) = -1.84$, $P = .08$], although there was a medium effect size ($d = .51$). Additional paired-samples t tests revealed no significant differences in the discrepancy between PSG and THIM on night 2 vs night 3 in any trial ($P > .13$). Therefore, the accuracy of THIM does not appear to be significantly reduced 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 (mean = 0.45, SD = 0.88 minutes) compared with poor sleepers (mean = 0.55, SD = 0.68 minutes) [$t(17) = -0.28$, $P = .78$, $d = .13$]. Similarly, there was no significant difference in the mean discrepancy on night 3 between good sleepers (mean = 0.89, SD = 1.65 minutes) and poor sleepers (mean = 0.87, SD = 1.06 minutes) [$t(18) = 0.03$, $P = .98$,

$d = .01$]. 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 with 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. In study 2, THIM-SOL showed strong correspondence and agreement with PSG sleep onset evidenced by the correlations, mean discrepancy tests, and Bland-Altman plots, for both good and poor sleepers and even after repeated use (night 2 compared with night 3). The revised THIM algorithm also improved 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. While this issue still occurred in approximately 10% of sleep onset trials with the revised algorithm, this is not thought to be a substantial issue that would impact the use of THIM for the device's main purpose of administering ISR. The ISR procedure involves 30–40 sleep onsets over the course of treatment, and if only 3–4 trials are unsuccessful, as anticipated from the findings of this study, then individuals should still experience many successful sleep onset trials across the treatment session. Additionally, the low degree of overestimation of sleep onset latency means that, when THIM wakens the individual, they are unlikely to have reduced homeostatic sleep drive enough to impact the subsequent sleep onset attempt and the efficacy of the treatment. Therefore, the revised algorithm was an improvement upon the original algorithm, and the device appears to be accurate enough at estimating SOL for the purpose of administering ISR. Future research would need to test whether the device is accurate enough for reliably administering power naps and the Multiple Sleep Latency Test, noting that the number of false-positive trails may be an issue for these purposes.

THIM had considerably closer agreement with PSG sleep onset compared with other wearable devices.^{22,23} The next generation of actigraphy devices that incorporate information from additional sensors, such as heart rate variability, appears to have greater accuracy compared with standard actigraphy devices.^{24,25} However, THIM shows greater agreement with PSG for estimating SOL than these multisensor devices: an underestimation of 7.48 minutes (SD = 6.64) was found in Fonseca et al²⁵ and a mean bias of 4 minutes (SD = 9) was found in de Zambotti et al²⁴ (see Scott et al⁵ for a review). In fact, THIM produced comparable accuracy to simplified EEG-based devices.^{26–28}

THIM also showed closer agreement with PSG sleep onset than similar devices that also use the stimulus-response method of sleep onset estimation.^{10,11} This may be due to differences in the stimulus type. THIM uses vibratory stimuli, which are perceived via a different sensory processing pathway compared with the auditory stimuli utilized by similar devices.^{13,15} It was evident from the piezo-electric sensor data collected during study 2 that once participants entered PSG-defined sleep, they ceased responding to the vibratory stimuli. This suggests that participants either (1) did not perceive the vibratory stimulus and remained totally asleep or (2) 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 PSG-defined sleep, which may also elucidate how disruptive the stimulus and finger-tap responses are on the process of falling asleep. Regardless, 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.

While this study evaluated the accuracy of THIM for estimating sleep onset latency across more than 2,000 sleep onset trials, the relatively low sample size has limited statistical power for detecting between-participant effects. Another limitation to consider is that the PSG data were scored by only 1 qualified sleep technician in the current study. The interrater reliability of N1 sleep onset, in particular, is low at approximately 68% and 74% for the epochs before and at sleep onset, respectively.²⁹ This adds to the error of measurement in the gold-standard measure that should be considered when interpreting the findings of the current study. Furthermore, although investigating THIM over 2 nights is a strength of this investigation, it is possible that observation over additional nights is necessary to detect learning effects. This would be important if individuals use the device frequently, such as for power naps. The use of THIM over more than 2 occasions should therefore be investigated in future research to explore its utility for frequent power napping.

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,³⁰ 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 with 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 2 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 theorized to develop over time,³¹ whereas poor sleep in general may be episodic in nature.³² Additionally, “poor sleepers” in this study could have also included those with sleep maintenance and early morning awakening nocturnal symptoms and not necessarily individuals with sleep-initiation difficulties, as are germane to the clinical utility of THIM for administering ISR. Therefore, the accuracy of THIM should be investigated with individuals with sleep onset insomnia specifically in future research.

Additional future research should be conducted to explore the clinical utility of THIM in other sleep-disordered populations.

Whether individuals will comply adequately with the instructions of tapping in response to THIM’s vibrations, a necessity for the device to estimate sleep onset appropriately, will need to be investigated in future research (see Lack et al³³ for further discussion). Whether THIM can also reliably wake people from sleep with the high-intensity alarm vibration is also a topic for further investigation. This may be particularly problematic for clinical uses with excessively sleepy patients who may be difficult to wake, such as for administering Multiple Sleep Latency Tests.

This article showcased the development of the THIM algorithm for estimating SOL in comparison to PSG. The revised algorithm demonstrated strong correspondence and agreement with PSG, 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. More research is needed to investigate whether other individual characteristics affect the accuracy of THIM, particularly a diagnosis of insomnia.

ABBREVIATIONS

EEG, electroencephalography
 ISI, Insomnia Severity Index
 ISR, Intensive Sleep Retraining
 PSG, polysomnography
 SOL, sleep onset latency

REFERENCES

- Harris J, Lack L, Wright H, Gradisar M, Brooks A. Intensive Sleep Retraining treatment for chronic primary insomnia: a preliminary investigation. *J Sleep Res*. 2007;16(3):276–284.
- Harris J, Lack L, Kemp K, Wright H, Bootzin R. A randomized controlled trial of intensive sleep retraining (ISR): a brief conditioning treatment for chronic insomnia. *Sleep*. 2012;35(1):49–60.
- Carskadon MA, Dement WC, Mitler MM, Roth T, Westbrook PR, Keenan S. Guidelines for the Multiple Sleep Latency Test (MSLT): a standard measure of sleepiness. *Sleep*. 1986;9(4):519–524.
- Lovato N, Lack L. The effects of napping on cognitive functioning. *Prog Brain Res*. 2010;185:155–166.
- Scott H, Lack L, Lovato N. A systematic review of the accuracy of sleep wearable devices for estimating sleep onset. *Sleep Med Rev*. 2020;49:101227.
- Re-Time Pty Ltd. THIM—the first wearable device for sleep, which will improve your sleep. Adelaide, Australia: Re-Time; 2016. Available from: <https://thim.io/>. Accessed August 22, 2017.
- Dement W, Kleitman N. Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalogr Clin Neurophysiol*. 1957;9(4):673–690.
- Loomis AL, Harvey EN, Hobart G. Potential rhythms of the cerebral cortex during sleep. *Science*. 1935;81(2111):597–598.
- Ogilvie RD. The process of falling asleep. *Sleep Med Rev*. 2001;5(3):247–270.
- Ogilvie RD, Wilkinson RT, Allison S. The detection of sleep onset: behavioral, physiological, and subjective convergence. *Sleep*. 1989;12(5):458–474.
- Mair A. The Relation between EEG and Behavioural Sleep Onset [honors thesis]. Flinders University, Adelaide, Australia; 1994.
- Scott H, Lack L, Lovato N. A pilot study of a novel smartphone application for the estimation of sleep onset. *J Sleep Res*. 2018;27(1):90–97.
- Cohen YE, Bennur S, Christison-Lagay K, Gifford AM, Tsunada J. Functional organization of the ventral auditory pathway. *Adv Exp Med Biol*. 2016;894:381–388.

14. Abraira VE, Ginty DD. The sensory neurons of touch. *Neuron*. 2013;79(4):618–639.
15. Kaas JH. Somatosensory system. In: Mai JK, Paxinos G, eds. *The Human Nervous System*. Amsterdam, Netherlands: Elsevier: Academic Press Books; 2012:1074–1109.
16. MacLean AW, Arnett T, Biedermann H, Knowles JB. Behavioural responding as a measure of sleep quality. *Sleep Res*. 1992;21:105.
17. 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.
18. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. 1989;28(2):193–213.
19. Berry RB, Albertario CL, Harding SM, et al; for the American Academy of Sleep Medicine. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Version 2.5. Darien, IL: American Academy of Sleep Medicine; 2018.
20. Carney CE, Buysse DJ, Ancoli-Israel S, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*. 2012;35(2):287–302.
21. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;327(8476):307–310.
22. Cellini N, Buman MP, McDevitt EA, Ricker AA, Mednick SC. Direct comparison of two actigraphy devices with polysomnographically recorded naps in healthy young adults. *Chronobiol Int*. 2013;30(5):691–698.
23. Chae KY, Kripke DF, Poceta JS, et al. Evaluation of immobility time for sleep latency in actigraphy. *Sleep Med*. 2009;10(6):621–625.
24. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. *Chronobiol Int*. 2018;35(4):465–476.
25. Fonseca P, Weysen T, Goelema MS, et al. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. *Sleep*. 2017;40(7):zsx097.
26. Cellini N, McDevitt EA, Ricker AA, Rowe KM, Mednick SC. Validation of an automated wireless system for sleep monitoring during daytime naps. *Behav Sleep Med*. 2015;13(2):157–168.
27. Kaplan RF, Wang Y, Loparo KA, Kelly MR, Bootzin RR. Performance evaluation of an automated single-channel sleep-wake detection algorithm. *Nat Sci Sleep*. 2014;6:113–122.
28. Markwald RR, Bessman SC, Reini SA, Drummond SP. Performance of a portable sleep monitoring device in individuals with high vs low sleep efficiency. *J Clin Sleep Med*. 2016;12(1):95–103.
29. Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scoring reliability program: sleep stage scoring. *J Clin Sleep Med*. 2013;9(1):81–87.
30. Perlis ML, Giles DE, Mendelson WB, Bootzin RR, Wyatt JK. Psychophysiological insomnia: the behavioural model and a neurocognitive perspective. *J Sleep Res*. 1997;6(3):179–188.
31. Perlis M, Jungquist C, Smith MT, Posner D. *The Cognitive Behavioral Treatment of Insomnia: A Treatment Manual*. New York, NY: Springer; 2005.
32. Perlis ML, Vargas I, Ellis JG, et al. The natural history of insomnia: the incidence of acute insomnia and subsequent progression to chronic insomnia or recovery in good sleeper subjects. *Sleep*. 2019;43(6):zsz299.
33. Lack L, Scott H, Micic G, Lovato N. Intensive sleep re-training: from bench to bedside. *Brain Sci*. 2017;7(4):E33.

ACKNOWLEDGMENTS

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

SUBMISSION & CORRESPONDENCE INFORMATION

Submitted for publication September 28, 2020

Submitted in final revised form December 23, 2020

Accepted for publication December 24, 2020

Address correspondence to: Hannah Scott, PhD, Adelaide Institute for Sleep Health, Flinders University, Box 6, 5 Laffer Drive, Bedford Park, SA 5042, Australia; Email: hannah.scott@flinders.edu.au

DISCLOSURE STATEMENT

All authors have seen and approved the manuscript. Work for this study was performed at the authors' respective institutions. Research costs were partially funded by Re-Time Pty Ltd, the company that sells THIM. None of the study authors were financially supported by Re-Time for this project. L.L. is a shareholder of Re-Time Pty Ltd. L.L. and H.S. have a patent pending regarding the THIM device. The other authors report no conflicts of interest.